ONTAP e applicazioni aziendali

Enterprise applications

NetApp
July 31, 2025

This PDF was generated from https://docs.netapp.com/it-it/ontap-apps-dbs/index.html on July 31, 2025.
Always check docs.netapp.com for the latest.

Sommario

ONTAP e applicazioni aziendali . . . 1
EPIC . . . 2
EPIC su ONTAP . . . 2
Scopo . . . 2
Scopo . . . 2
Pubblico . . . 3
EPIC su ONTAP . . . 3
EPIC su ONTAP . . . 3
Disponibilità EPIC su ONTAP . . . 4
EPIC su ONTAP Consolidation . . . 4
EPIC su ONTAP Efficiency . . . 4
Prestazioni epiche su ONTAP . . . 5
EPIC su scalabilità ONTAP . . . 6
Configurazione straordinaria dell’efficienza dello storage . . . 7
EPIC su ONTAP Security . . . 7
Architettura e design epici . . . 8
Architettura epica . . . 8
Dimensioni epiche . . . 11
Requisiti di storage EPIC . . . 12
Architettura EPIC a quattro nodi . . . 13
Architettura EPIC a sei nodi . . . 14
Architettura EPIC a otto nodi . . . 16
Configurazione e Best practice . . . 17
EPIC su ONTAP - Utilità host . . . 17
Configurazione EPIC di LUN e volume . . . 17
Protocolli EPIC e file . . . 19
Gestione epica delle performance . . . 19
EPIC su ONTAP - protocolli . . . 20
Configurazione straordinaria dell’efficienza dello storage . . . 20
Configurazione straordinaria dell’efficienza dello storage . . . 21
Dimensionamento dello storage per Epic . . . 22
Informazioni aggiuntive per Epic su ONTAP . . . 22
Documenti di orientamento epici per i clienti . . . 22
Hyper-V. . . . 24
Linee guida per l’implementazione e Best practice per lo storage . . . 24
Panoramica . . . 24
Storage NetApp e ambiente Windows Server . . . 25
Provisioning negli ambienti SAN . . . 29
Provisioning negli ambienti SMB . . . 37
Infrastruttura storage Hyper-V su NetApp . . . 41
Efficienza dello storage . . . 52
Sicurezza . . . 54
Distribuire il server Nano . . . 54

Implementa il cluster Hyper-V. 58
Distribuire Hyper-V Live Migration in un ambiente in cluster 59
Implementazione di Hyper-V Live Migration all'esterno di un ambiente in cluster 60
Distribuire Hyper-V Storage Live Migration 61
Implementazione di replica Hyper-V all'esterno di un ambiente in cluster 62
Distribuire la replica Hyper-V in un ambiente in cluster 63
Dove trovare ulteriori informazioni 64
Microsoft SQL Server 66
Panoramica 66
Workload di Microsoft SQL Server 66
Configurazione del database 67
Configurazione CPU 67
Configurazione della memoria 70
Istanza condivisa contro istanza dedicata 73
File tempdb 74
Configurazione dello storage su sistemi AFF/FAS 75
Panoramica 75
File di database e filegroup 77
Efficienza dello storage 82
Protezione dei dati 87
Disaster recovery 89
Configurazione dello storage sui sistemi ASA R2 110
Panoramica 110
File di database e filegroup 112
Protezione dei dati 116
Disaster recovery 117
Sicurezza dei dati 133
Copie Snapshot 133
Snapshot a prova di manomissione 133
Replica SnapMirror 134
Macchine virtuali di storage 134
RBAC amministrativo 134
Autenticazione multifattore (MFA) 134
RBAC API 135
Verifica multi-admin (MAV) 135
MySQL 136
Panoramica 136
Configurazione del database 136
Struttura dei file 136
Parametri di configurazione 139
innodb_log_file_size 140
innodb_flush_log_at_trx_commit 140
innodb_doublewrite 141
innodb_buffer_pool_size 141
innodb_flush_method 141

innodb_io_capacity . . . 142
innodb_lru_scan_depth . . . 143
open_file_limits . . . 143
Configurazione dell'host . . . 143
Containerizzazione . . . 143
NFSv3 tavoli con slot . . . 144
Scheduler i/O. . . . 144
Descrittori di file . . . 145
Configurazione dello storage . . . 145
NFS . . . 145
SAN . . . 146
Database Oracle . . . 149
Database Oracle su ONTAP . . . 149
Configurazione di ONTAP . . . 149
RAID . . . 149
Gestione della capacità . . . 150
Macchine virtuali di storage . . . 150
Gestione delle performance con QoS ONTAP . . . 151
Efficienza . . . 153
Thin provisioning . . . 157
Failover/switchover di ONTAP . . . 159
Configurazione del database . . . 161
Dimensioni dei blocchi . . . 161
db_file_multiblock_read_count . . . 161
filesystemio_options . . . 162
Timeout RAC . . . 163
Configurazione dell'host . . . 164
AIX . . . 164
HP-UX . . . 166
Linux . . . 168
ASMILib/AFD (driver filtro ASM) . . . 172
Microsoft Windows . . . 174
Solaris . . . 174
Configurazione di rete . . . 180
Interfacce logiche . . . 180
Configurazione TCP/IP ed ethernet . . . 184
Configurazione FC SAN . . . 186
Connessione di rete diretta . . . 186
Configurazione dello storage . . . 187
SAN FC . . . 187
NFS . . . 192
NVFAIL . . . 205
ASM Reclamation Utility (ASMRU) . . . 205
Virtualizzazione . . . 206
Supportabilità . . . 206

Presentazione storage . . . 206
Driver paravirtualizzati . . . 208
Overcommit RAM . . . 208
Striping dei datastore . . . 208
Tiering . . . 209
Panoramica . . . 209
Policy di tiering . . . 211
Strategie di tiering . . . 213
Interruzioni di accesso agli archivi di oggetti . . . 217
Data Protection Oracle . . . 217
Data Protection con ONTAP . . . 217
RTO, RPO e pianificazione SLA . . . 218
Disponibilità del database . . . 221
Checksum e integrità dei dati . . . 222
Elementi di base di backup e recovery . . . 227
Disaster recovery Oracle . . . 241
Panoramica . . . 241
MetroCluster . . . 242
Sincronizzazione attiva di SnapMirror . . . 261
Migrazione dei database Oracle . . . 295
Panoramica . . . 295
Pianificazione della migrazione . . . 296
Procedure . . . 299
Script di esempio . . . 402
Note aggiuntive . . . 414
Ottimizzazione delle prestazioni e benchmarking . . . 414
NFSv3 serrature obsolete . . . 417
Verifica dell’allineamento di WAFL . . . 418
PostgreSQL . . . 424
Panoramica . . . 424
Configurazione del database . . . 424
Architettura . . . 424
Parametri di inizializzazione . . . 425
Impostazioni . . . 426
Tablespace . . . 428
Configurazione dello storage . . . 428
NFS . . . 428
SAN . . . 430
Protezione dei dati . . . 432
Protezione DDTA nativa . . . 432
Snapshot . . . 433
Software per la data Protection . . . 434
SAP . . . 435
VMware . . . 436
VMware vSphere con ONTAP . . . 436

VMware vSphere con ONTAP . . . 436
Perché scegliere ONTAP per VMware vSphere? . . . 436
Storage unificato . . . 438
Strumenti di virtualizzazione per ONTAP . . . 439
Virtual Volumes (vVol) e Storage Policy Based Management (SPBM) . . . 442
Datastore e protocolli . . . 443
Configurazione di rete . . . 458
Clonazione di VM e datastore . . . 460
Protezione dei dati . . . 462
Qualità del servizio (QoS) . . . 465
Migrazione e backup del cloud . . . 470
Crittografia per i dati vSphere . . . 471
Active IQ Unified Manager . . . 472
Gestione basata su criteri di archiviazione e vVol . . . 474
VMware Storage Distributed Resource Scheduler . . . 476
Host ESXi consigliato e altre impostazioni ONTAP . . . 477
Volumi virtuali (vVol) con strumenti ONTAP 10 . . . 480
Panoramica . . . 480
Elenco di controllo . . . 486
Utilizzo di vVol con ONTAP . . . 489
Implementazione di vVol su sistemi AFF, ASA, ASA R2 e FAS . . . 494
Protezione di vVol . . . 505
Risoluzione dei problemi . . . 510
VMware Site Recovery Manager con ONTAP . . . 511
VMware Live Site Recovery con ONTAP . . . 511
Best practice per l'implementazione . . . 513
Best practice operative . . . 514
Topologie di replica . . . 518
Risoluzione dei problemi relativi a VLSRM/SRM quando si utilizza la replica vVols . . . 527
Ulteriori informazioni . . . 527
VSphere Metro Storage Cluster con ONTAP . . . 528
VSphere Metro Storage Cluster con ONTAP . . . 528
Panoramica della soluzione VMware vSphere . . . 531
Linee guida per la progettazione e l'implementazione di vMSC . . . 536
Resilienza per eventi pianificati e non pianificati . . . 547
Scenari di errore per vMSC con MetroCluster . . . 547
Sicurezza dei prodotti . . . 559
Strumenti ONTAP per VMware vSphere . . . 559
Plug-in di SnapCenter per VMware vSphere . . . 561
Guida alla protezione avanzata per gli strumenti ONTAP per VMware vSphere . . . 563
Guida alla protezione avanzata per gli strumenti ONTAP per VMware vSphere 9,13 . . . 563
Verifica dell'integrità dei tool ONTAP per i pacchetti di installazione di VMware vSphere 9,13 . . . 564
Porte e protocolli per gli strumenti ONTAP 9,13 . . . 566
Tool ONTAP per access point VMware vSphere 9,13 (utenti) . . . 567
ONTAP tools 9,13 Mutual TLS (autenticazione basata su certificato) . . . 568

ONTAP tools 9,13 certificato HTTPS . . . 574
Banner di accesso di ONTAP Tools 9,13 . . . 574
Timeout di inattività per gli strumenti ONTAP 9,13 . . . 575
Numero massimo di richieste simultanee per utente (protezione di rete/attacco DOS) Strumenti ONTAP per VMware vSphere 9,13 . . . 575
Configurazione del protocollo NTP (Network Time Protocol) per gli strumenti ONTAP 9,13 . . . 576
Criteri delle password per gli strumenti ONTAP 9,13 . . . 576
Note legali . . . 578
Copyright . . . 578
Marchi . . . 578
Brevetti . . . 578
Direttiva sulla privacy . . . 578
Open source . . . 578
ONTAP . . . 578
ONTAP Mediator per configurazioni IP MetroCluster . . . 578

ONTAP e applicazioni aziendali

EPIC

EPIC su ONTAP

La chiave della digital transformation sta facendo di più con i tuoi dati.

Questa documentazione sostituisce questo report tecnico precedentemente pubblicato *TR-3923: Best practice NetApp per Epic*.

Gli ospedali richiedono grandi quantità di dati per iniziare il viaggio della digital transformation. Parte del processo di trattamento dei pazienti, gestione degli orari del personale e delle risorse mediche è che le informazioni vengono raccolte ed elaborate. Tuttavia, molte azioni vengono ancora eseguite manualmente o tramite sistemi obsoleti. L'unica costante è che la quantità di dati continua a crescere in modo esponenziale e quindi diventa sempre più difficile da gestire.

La causa principale di questo problema è che i dati ospedalieri vengono spesso archiviati in silos di dati. Troppo tempo viene impiegato per le immissioni manuali e gli aggiornamenti che portano a burnout ed errori. Il presente documento riguarda una parte dei dati del settore sanitario, la Epic Electronic Health Records (EHR). Tuttavia, la strategia di gestione dei dati qui trattata può e deve essere applicata a tutti i dati sanitari. NetApp vanta una lunga storia di modernizzazione e semplificazione dell'infrastruttura digitale. La nostra infrastruttura dati intelligente costituisce la base della digital transformation.

NetApp offre una singola soluzione di gestione dei dati per tutte le esigenze del settore sanitario e possiamo guidare gli ospedali nel loro percorso verso la digital transformation. Costruendo una base con strutture e soluzioni intelligenti, l'assistenza sanitaria può trarre il pieno valore di queste preziose informazioni. Questo quadro può aiutare i medici a diagnosticare le malattie più velocemente e sviluppare piani di trattamento personalizzati per supportare meglio i processi decisionali in situazioni di emergenza. Inoltre, sarà possibile creare una propria infrastruttura dati intelligente e consentire all'ospedale di sbloccare i silos di dati, agevolare l'interoperabilità dei dati e proteggere le informazioni sensibili dei pazienti.

Utilizza questo documento come guida per costruire e implementare con successo Epic EHR. Invece di costruire diversi silos Epic, crea una singola infrastruttura dati Epic e trasforma il tuo ospedale.

Scopo

Questo documento descrive le Best practice per l'integrazione dello storage NetApp in un ambiente software Epic. Contiene le seguenti sezioni:

- Una conoscenza tecnica dell'ambiente software Epic e dei relativi requisiti di storage in diverse configurazioni.
- Considerazioni epiche sullo storage che descrivono importanti fattori decisionali per le soluzioni Epic.
- Consigli sullo storage NetApp, descrivendo le Best practice di configurazione dello storage NetApp per soddisfare i requisiti di storage Epic.

Scopo

Questo documento non tratta i seguenti argomenti:

- Requisiti quantitativi delle prestazioni e linee guida sul dimensionamento, risolti in "TR-3930i: Linee guida per il dimensionamento degli NetApp per Epic" (è richiesto l'accesso NetApp)

Pubblico

NetApp presuppone che il lettore disponga delle seguenti conoscenze di base:

- Una solida comprensione dei concetti SAN e NAS
- Familiarità tecnica con i sistemi di storage ONTAP
- Familiarità tecnica con la configurazione e l'amministrazione di ONTAP

EPIC su ONTAP

EPIC su ONTAP

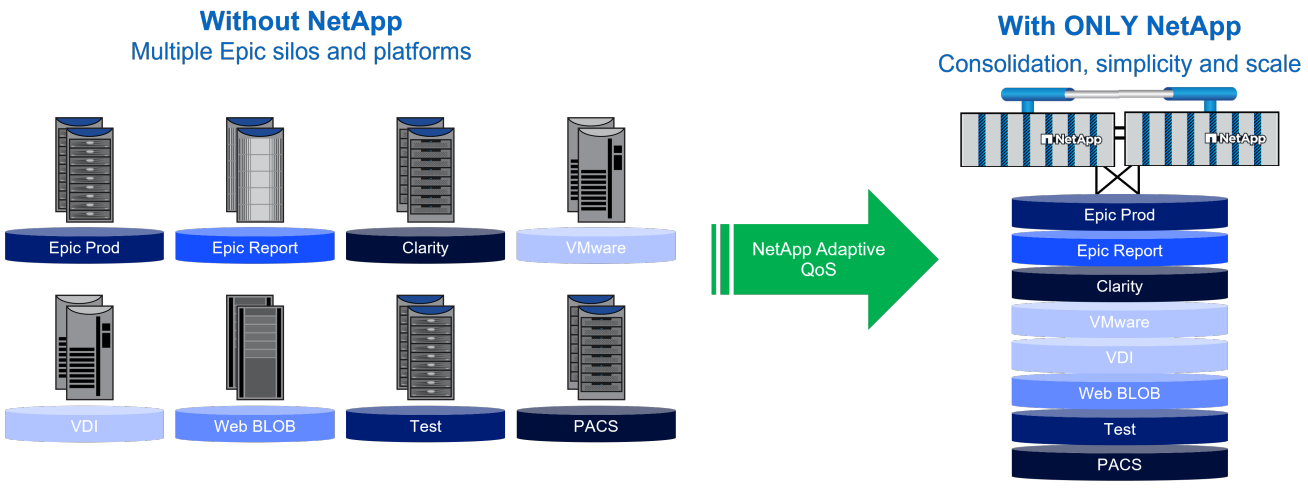
EPIC è più facile con ONTAP.

ONTAP è una piattaforma per la gestione dei dati che ti consente di consolidare i workload Epic e soddisfare al contempo tutti i tuoi requisiti in termini di performance, data Protection e gestione dei dati.

Solo NetApp ti consente di standardizzare tutti i carichi di lavoro del settore sanitario per SAN, NAS e oggetti su una singola piattaforma per la gestione dei dati ad alta disponibilità. ONTAP è la piattaforma software di storage più diffusa al mondo e viene fornita con quasi 30 anni di innovazione costante. Puoi affrontare tutte le sfide Epic con strumenti nativi per la gestione dei dati ONTAP e integrazione applicativa. Non è necessario acquistare numerosi strumenti di terze parti per colmare le lacune della soluzione.

Molti vendor di soluzioni storage offrono uno storage a blocchi tradizionale, affidabile e veloce. Funzionano bene, ma di norma sono implementate in silos per eseguire un singolo workload come produzione, report, Clarity, VDI, VMware e NAS. Ciascuna di queste unità utilizza hardware e strumenti di gestione diversi e generalmente vengono gestite da diversi gruppi IT. Questo approccio tradizionale aggiunge al problema più grande con l'assistenza sanitaria oggi - la complessità.

NetApp rende la gestione dei dati più semplice ed efficiente. Invece di sprecare denaro nel caso di silos di dimensioni eccessive, ONTAP utilizza l'innovazione e la tecnologia per offrire SLA coerenti e garantiti per ogni carico di lavoro su una singola piattaforma, su qualsiasi protocollo con la protezione integrata dei dati. Queste funzionalità e questi strumenti si estendono anche al cloud di tua scelta, come illustrato di seguito.



Disponibilità EPIC su ONTAP

Al cuore di ONTAP troviamo operazioni senza interruzioni che consentono di evitare costose interruzioni delle operazioni di business.

NetApp offre una disponibilità di oltre il 99,999999% in base ai dati di produzione, che vengono chiamati "a casa" tramite NetApp Active IQ. Ogni coppia ha nel cluster non ha single point of failure. ONTAP risale al 1992 e rappresenta il software per la gestione dei dati più diffuso al mondo, con una storia eccezionale di fornitura di storage affidabile. Ora, grazie al monitoring proattivo e alla risoluzione automatica del 97% dei problemi da parte di Active IQ, la disponibilità è maggiore e i casi di supporto sono notevolmente inferiori.

EPIC consiglia l'utilizzo di sistemi storage ha per ridurre i guasti dei componenti hardware. Questo suggerimento va dall'hardware di base (ad esempio gli alimentatori ridondanti) al networking (ad esempio il collegamento in rete multipercorso).

Quando è necessario aggiornare lo storage, scalare in verticale, scalare in orizzontale o ribilanciare i carichi di lavoro nel cluster, non ci sono effetti per la cura del paziente. È possibile spostare i dati, ma non interrompere mai più l'attenzione dei pazienti grazie a migrazioni o aggiornamenti integrali. Passa alla tecnologia di prossima generazione, a prova di futuro ed evita il lock-in hardware. NetApp offre anche una garanzia di disponibilità scritta del 100%.

Ulteriori informazioni sulle funzionalità di affidabilità, disponibilità, manutenzione e sicurezza di NetApp sono disponibili nel "Affidabilità, disponibilità, serviceability e sicurezza di NetApp ONTAP"white paper.

EPIC su ONTAP Consolidation

Una delle principali sfide nel settore sanitario è l'inefficienza degli ambienti a silos.

Le soluzioni a più punti vengono create da vari gruppi che impediscono il progresso. Disporre di una strategia unificata per la gestione dei dati offre efficienza per accelerare la trasformazione. Tecnologie dirompenti come la digitalizzazione delle cartelle cliniche dei pazienti, il ransomware e l'ai generativa, favoriscono tutto ciò che è necessario consolidare.

Con ONTAP puoi consolidare file/blocchi/oggetti e ciascuno dei tuoi workload di Tier 0/1/2/3, on-premise e nel cloud, il tutto in esecuzione su ONTAP.

EPIC su ONTAP Efficiency

EPIC viene eseguito su array all-flash in cui la maggior parte del costo è costituita dal disco. Pertanto, l'efficienza dello storage è fondamentale per il risparmio sui costi.

L'efficienza dello storage inline di NetApp ottiene risparmi leader del settore sullo storage senza effetti sulle performance e offriamo persino una garanzia di efficienza scritta con gli array all-flash.

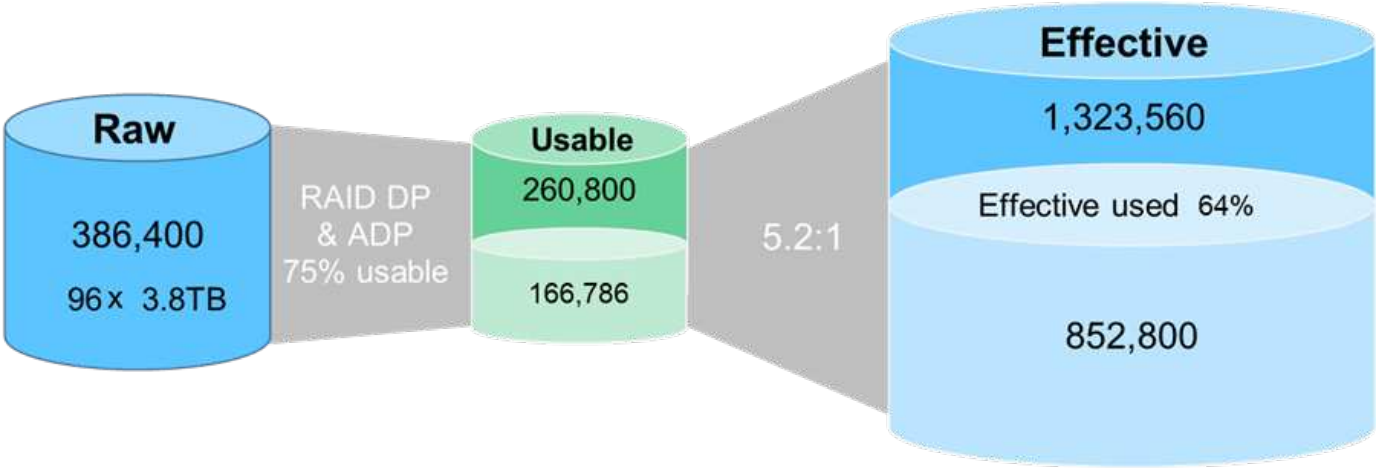
Nel calcolo dell'efficienza dello storage, è importante misurare la capacità da raw a utilizzabile su quella effettiva.

- **Capacità nativa** prima di applicare qualsiasi RAID, la dimensione del disco in base al numero di dischi.
- **Capacità utilizzabile** dopo l'applicazione del RAID, quanto spazio di archiviazione utilizzabile è disponibile.
- **Capacità effettiva** quantità di storage fornita e presentata all'host o al client.

La figura riportata di seguito rappresenta un calcolo dell'efficienza di esempio di un'implementazione Epic

tipica, che include tutti i carichi di lavoro che richiedono 852TB TB di storage effettivo e con un'efficienza di 5,2:1 che fornisce 1,32PB TB di dati effettivi totali.

In base al numero di dischi, la capacità da raw a utilizzabile varia leggermente.



NetApp non utilizza la tecnologia Snapshot di NetApp o il thin provisioning per calcolare l'efficienza nel programma di garanzia. In questo modo è possibile ottenere efficienze non realistiche pari a 30 o 100:1, cosa che non significa nulla per il dimensionamento della capacità dello storage reale.

Prestazioni epiche su ONTAP

ONTAP ha introdotto le tecnologie flash nel 2009 e ha supportato i dischi a stato solido dal 2010. Questa lunga esperienza con lo storage flash consente a NetApp di mettere a punto le funzionalità ONTAP per ottimizzare le performance SSD e aumentare la resistenza dei supporti flash mantenendo al contempo le funzionalità ricche di funzionalità di ONTAP.

A partire dall'anno 2020, tutti i workload Epic ODB devono trovarsi su storage all-flash. I carichi di lavoro EPIC operano tipicamente a circa 1.000-2.000 IOPS per terabyte di storage (blocco da 8k KB, rapporto di lettura e scrittura 75%/25% e 100% di random). EPIC è molto sensibile alla latenza e l'elevata latenza ha un effetto visibile sull'esperienza dell'utente finale, oltre che su task operativi come l'esecuzione di report, backup, controlli di integrità e tempi di refresh dell'ambiente.

- Il fattore limitante per gli array all-flash non è i dischi, ma piuttosto l'utilizzo sui controller.
- ONTAP utilizza un'architettura Active-Active. Per ottenere performance elevate, entrambi i nodi nella coppia ha scrivono sui dischi.
- Questo risultato è un utilizzo massimizzato della CPU, il singolo fattore più importante che consente a NetApp di pubblicare le migliori performance Epic nel settore.
- Le tecnologie NetApp RAID DP, ADP (Advanced Disk Partitioning) e WAFL soddisfano tutti i requisiti Epic. Tutti i carichi di lavoro distribuiscono i/o su tutti i dischi. Nessun collo di bottiglia.
- ONTAP è ottimizzato in funzione della scrittura; le scritture vengono riconosciute una volta scritte sulla NVRAM mirrorata prima di essere scritte su disco a una velocità della memoria inline.
- WAFL, NVRAM e l'architettura modulare consentono a NetApp di utilizzare il software per innovare con efficienze inline, crittografia e performance. Consentono inoltre a NetApp di introdurre nuove caratteristiche

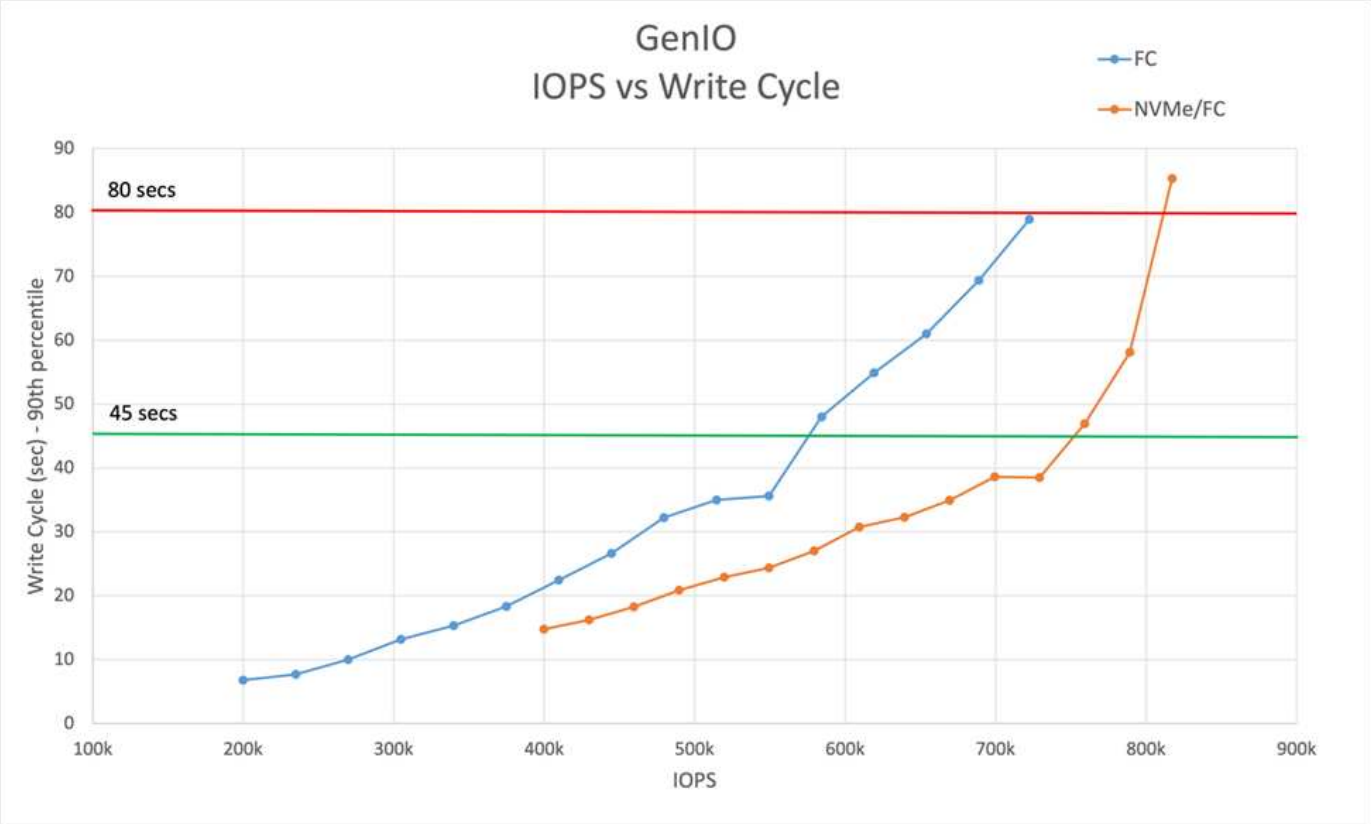
e funzionalità senza influire sulle prestazioni.

- Storicamente, ogni nuova versione di ONTAP aumenta le prestazioni e l'efficienza del 30-50%. Le prestazioni sono ottimali con ONTAP.

NVMe

Quando le performance sono di primaria importanza, NetApp supporta anche NVMe/FC, il protocollo FC SAN di nuova generazione.

Come si può notare nella figura sotto, i nostri test Genio hanno raggiunto un numero molto maggiore di IOPS utilizzando il protocollo NVMe/FC rispetto al protocollo FC. La soluzione connessa NVMe/FC ha raggiunto oltre 700k IOPS prima di superare la soglia del ciclo di scrittura di 45 secondi. Sostituendo i comandi SCSI con NVMe, puoi anche ridurre significativamente l'utilizzo sull'host.



EPIC su scalabilità ONTAP

La Epic hardware Configuration Guide registra una crescita annua di circa il 20% per 3 anni. Tuttavia, anche gli ambienti possono crescere inaspettatamente.

NetApp può scalare perfettamente performance e capacità fino a 12 nodi per cluster NAS, SAN e a oggetti. Potrai quindi scalare in verticale e in orizzontale senza interruzioni seguendo la crescita del tuo business.

EPIC Iris offre ulteriori funzionalità di scalabilità. Consente ai clienti più grandi con diverse istanze Epic di consolidarsi in una singola istanza. Il "Epic sull'architettura verificata NetApp su SAN moderna" documento dimostra che Epic è in grado di scalare perfettamente i workload consolidati fino a 720K milioni di IOPS su una singola ha e di scalare in orizzontale fino a oltre 4M milioni di IOPS in un cluster. Puoi scalare in verticale senza interruzioni aggiornando i controller o aggiungendo dischi ai cluster esistenti.

Inoltre, è possibile spostare i dati NAS, SAN e a oggetti senza interruzioni tra i nodi del cluster. Ogni coppia ha nel cluster può essere qualsiasi combinazione di tipi e dimensioni di sistemi ONTAP FAS e AFF. Puoi bilanciare i carichi di lavoro in un singolo cluster per massimizzare l'investimento storage.

ONTAP offre anche la possibilità di utilizzare lo storage a oggetti su StorageGRID o nel cloud come destinazione di backup e/o destinazione del tiering automatico del cold storage. Questa funzionalità ti consente di liberare automaticamente sugli oggetti i costosi dischi all-flash, le snapshot di Tier e i dati cold.

Il risultato è che Epic funziona semplicemente meglio con il portfolio di prodotti NetApp, sfruttando ONTAP, diversi protocolli, StorageGRID e il cloud di tua scelta. Questi prodotti offrono opzioni di disaster recovery, archiviazione, analisi, tiering e altro ancora.

Configurazione straordinaria dell'efficienza dello storage

Uno snapshot è una copia point-in-time di un volume di sola lettura.

Uno snapshot inserisce un blocco logico in tutti i blocchi del file system attivo. Le copie Snapshot di NetApp ONTAP sono near-Instant e non utilizzano storage aggiuntivo.

Write ANHERE file Layout, o WAFL, è un file system di sola scrittura; non esegue io aggiuntivi, come la copia dei dati in un blocco protetto da snapshot prima di essere sovrascritti. Non viene mai spostato i dati; pertanto, le snapshot non hanno alcun effetto sulla capacità o sulle performance dello storage. Le snapshot consentono enormi risparmi in termini di storage, migliorando al tempo stesso la soluzione di backup.

FlexClone

Un volume NetApp ONTAP FlexClone è un clone di un volume esistente o di uno Snapshot di un volume esistente. Si tratta di un volume ONTAP come qualsiasi altro e può essere esso stesso clonato, protetto con Snapshot e configurato con una policy di QoS.

Come per le Snapshot, un volume FlexClone non richiede spazio aggiuntivo al momento della creazione. Solo le modifiche al clone richiedono capacità aggiuntiva.

EPIC richiede da 10 a 30 copie dei database di produzione per vari requisiti operativi, come backup in streaming, controlli dell'integrità e ambienti di aggiornamento dello staging. L'esigenza di una soluzione costruita su FlexClone Volumes è aumentata con il passaggio ad aggiornamenti più frequenti.

> Una soluzione di backup Epic completamente automatizzata e una soluzione di refresh Epic sono fornite da NetApp come parte della soluzione utilizzando Ansible e i tool NetApp nativi.

EPIC su ONTAP Security

Oggi la sicurezza è la preoccupazione principale per le organizzazioni e i dirigenti del settore sanitario. Non è mai stato così difficile da gestire e le organizzazioni devono affrontare problemi come compliance, governance dei dati, protezione antivirus e ransomware.

Una guida completa a Epic e alla sicurezza dello storage esulano dall'ambito di questo documento; tuttavia, "Guida alla protezione per ONTAP" elenca tutte le funzionalità di sicurezza estese e avanzate disponibili con ONTAP.

NetApp Active IQ Unified Manager monitora le violazioni di sicurezza in base alle informazioni incluse nel "TR-4569" e le segnala nella dashboard per semplificare la gestione della sicurezza. Questi strumenti possono

aiutare la vostra organizzazione a raggiungere gli obiettivi di sicurezza per proteggere, rilevare e risolvere gli attacchi.

NetApp ha inoltre collaborato con i vendor di sicurezza per fornire l'integrazione tramite "FPolicy di NetApp" software per migliorare la tua offerta di sicurezza. Inoltre, "Autenticazione multifattore (MFA)" può essere aggiunto per proteggere l'ambiente Epic da accessi non autorizzati con credenziali trapelate.

Infine, le copie snapshot native di ONTAP e le tecnologie immutabili SnapLock con "Cyber vault di ONTAP", offrono un'esclusiva funzionalità air gap per proteggere le cartelle cliniche dei pazienti dal ransomware. Consultare la documentazione NetApp a "La soluzione NetApp per ransomware". Per un approccio più strategico alla protezione, vedere "NetApp e zero trust".

Architettura e design epici

Architettura epica

Questa sezione descrive l'ambiente software Epic e i componenti chiave che richiedono lo storage. Fornisce importanti considerazioni per guidare la progettazione dello storage.

EPIC, che ha sede a Verona, Wisconsin, produce software per gruppi medici, ospedali e organizzazioni sanitarie integrate di medie e grandi dimensioni. I clienti includono anche ospedali comunitari, strutture accademiche, organizzazioni per bambini, fornitori di reti di sicurezza e sistemi multi-ospedalieri. Il software integrato di livello epico comprende funzioni cliniche, di accesso e di guadagno e si estende a casa.

Non rientra nello scopo di questo documento descrivere l'ampia gamma di funzioni supportate dal software Epic. Dal punto di vista del sistema storage, tuttavia, tutto il software Epic condivide un singolo database incentrato sul paziente per ogni implementazione. EPIC sta passando dal database InterSystems Caché al nuovo database InterSystems Iris. Poiché i requisiti di storage sono gli stessi per Caché e Iris, nel resto del documento faremo riferimento al database come Iris. Iris è disponibile per i sistemi operativi AIX e Linux.

IRIS intersistemi

InterSystems Iris è il database utilizzato dall'applicazione Epic. In questo database, il server dati è il punto di accesso per i dati memorizzati in modo permanente. Il server applicazioni gestisce le query del database ed esegue le richieste di dati al server dati. Per la maggior parte degli ambienti software Epic, l'utilizzo dell'architettura SMP (Symmetric Multiprocessor) in un unico server di database è sufficiente per soddisfare le richieste di database delle applicazioni Epic. Nelle distribuzioni di grandi dimensioni, è possibile supportare un modello distribuito utilizzando il protocollo ECP (Enterprise Caché Protocol) di InterSystems.

L'utilizzo di hardware in cluster abilitato per il failover consente a un server dati in standby di accedere allo stesso storage del server dati primario. Consente inoltre al server dati di standby di assumersi le responsabilità di elaborazione durante un guasto hardware.

Inoltre, fornisce tecnologie in grado di soddisfare i requisiti di replica dei dati, disaster recovery e alta disponibilità (ha). La tecnologia di replica di InterSystems viene utilizzata per replicare un database Iris in modo sincrono o asincrono da un server dati primario a uno o più server dati secondari. NetApp SnapMirror viene utilizzato per replicare lo storage WebBLOB o per il backup e il disaster recovery.

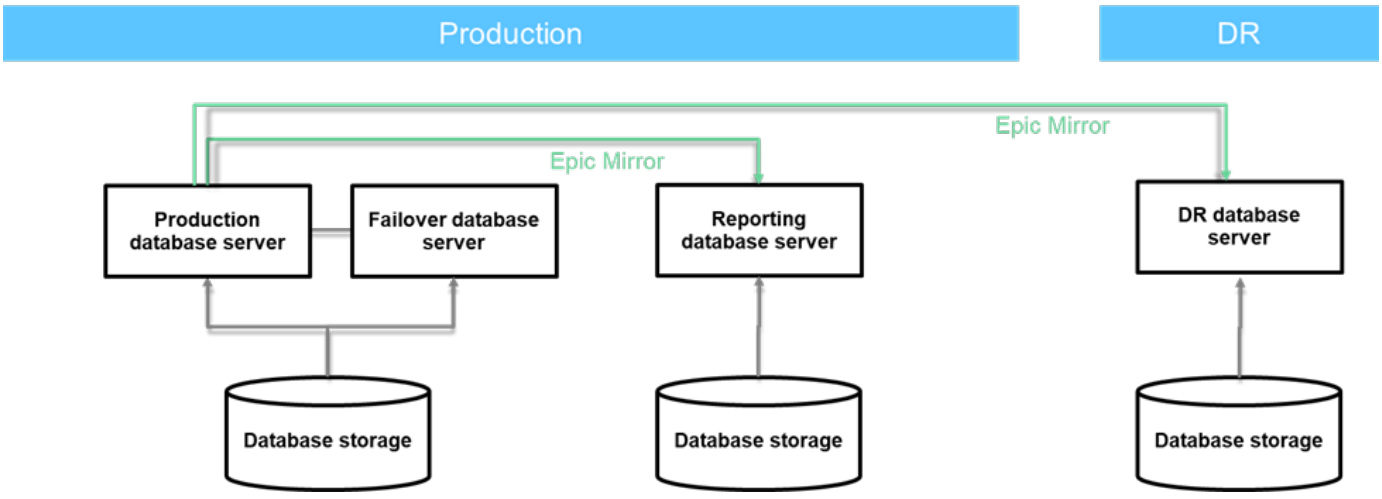
Il database Iris aggiornato presenta numerosi vantaggi:

- Maggiore scalabilità e possibilità per le organizzazioni più grandi con diverse istanze Epic di consolidarsi in un'unica istanza più ampia.
- Un periodo di validità delle licenze in cui i clienti possono ora passare da AIX a Red Hat Enterprise Linux

(RHEL) senza dover pagare una nuova licenza per la piattaforma.

Utilizzo dello storage e dei server del database Caché

- **Produzione** negli ambienti software Epic viene implementato un singolo database incentrato sul paziente. Nei requisiti hardware di Epic, il server fisico che ospita il server primario dei dati IRIS di lettura/scrittura è chiamato server di database di produzione. Questo server richiede uno storage all-flash dalle performance elevate per i file appartenenti all'istanza del database primario. Per l'alta disponibilità, Epic supporta l'utilizzo di un server di database di failover che ha accesso agli stessi file. Iris utilizza Epic Mirror per la replica nel report di sola lettura, il disaster recovery e il supporto delle copie di sola lettura. Ogni tipo di server di database può essere impostato sulla modalità di lettura/scrittura per motivi di continuità aziendale.
- **Report** Un server di database mirror per la creazione di report fornisce l'accesso in sola lettura ai dati di produzione. Ospita un server di dati Iris configurato come mirror di backup del server di dati Iris di produzione. Il server del database di reporting ha gli stessi requisiti di capacità di archiviazione del server del database di produzione. Reporting delle performance di scrittura è lo stesso della produzione, ma le caratteristiche del carico di lavoro in lettura sono diverse e dimensionate in modo diverso.
- **Supporta la sola lettura** questo server database è opzionale e non è mostrato nella figura seguente. È inoltre possibile implementare un server database mirror per supportare Epic supporta funzionalità di sola lettura, in cui viene fornito l'accesso a una copia di produzione in modalità di sola lettura. Questo tipo di server di database può essere impostato sulla modalità di lettura/scrittura per motivi di continuità aziendale.
- **Disaster Recovery** per soddisfare gli obiettivi di business continuity e disaster recovery, un server di database mirror per il disaster recovery viene comunemente installato in un sito geograficamente separato dai server di database mirror per la produzione e/o la creazione di rapporti. Un server di database mirror per il disaster recovery ospita anche un server di dati Iris configurato come mirror di backup del server di dati Iris di produzione. Se il sito di produzione non è più disponibile per un lungo periodo di tempo, è possibile configurare questo server di database mirror per fungere da istanza di lettura/scrittura speculare (SRW). Il server del database mirror di backup ha gli stessi requisiti di archiviazione dei file del server del database di produzione. Al contrario, lo storage del database del mirroring del backup è dimensionato allo stesso modo dello storage di produzione, dal punto di vista delle prestazioni per la business continuity.



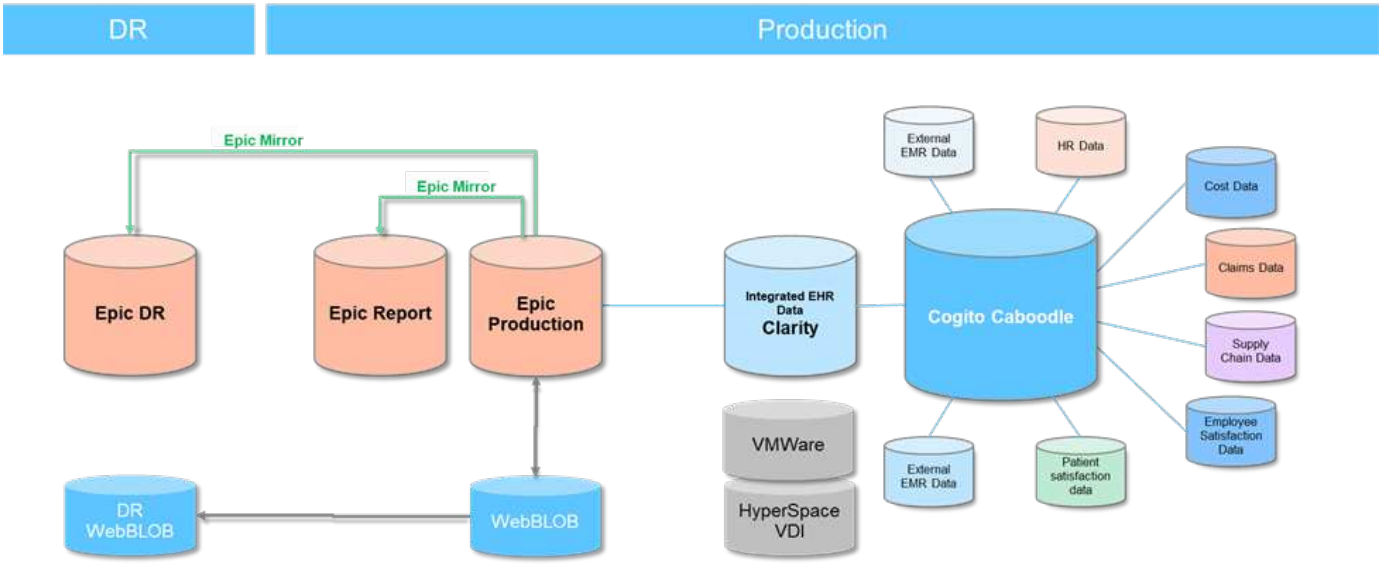
- **Test** le organizzazioni sanitarie spesso distribuiscono ambienti di sviluppo, test e staging. Anche i server di dati IRIS aggiuntivi per questi ambienti richiedono uno storage che può essere gestito dallo stesso sistema di storage. EPIC ha requisiti e vincoli specifici per fornire storage aggiuntivo da un sistema storage condiviso. Questi requisiti specifici sono affrontati genericamente dalle Best practice contenute in questo documento.

Oltre ai server di dati ODB Iris, gli ambienti software Epic includono in genere altri componenti, come quelli seguenti e come mostrato nella figura seguente:

- Un server di database Oracle o Microsoft SQL Server come back-end degli strumenti di reporting aziendale di Epic Clarity

> Clarity viene utilizzato per generare rapporti sui dati estratti giornalmente dal database Iris di reporting.

- Server WebBLOB (SMB)
- Server database multifunzione
- Macchine virtuali polivalenti (VM)
- Spazio ipertestuale per accesso client



I requisiti di storage di tutti questi workload, pool, protocolli NAS e SAN multipli possono essere consolidati e ospitati da un singolo cluster ONTAP. Questo consolidamento consente alle organizzazioni del settore sanitario di disporre di una singola strategia di gestione dei dati per tutti i workload Epic e non Epic.

Carichi di lavoro del database operativi

Ogni server di database Epic esegue l'i/o sui seguenti tipi di file:

- File di database
- File journal
- File dell'applicazione

Il carico di lavoro di un singolo server di database dipende dal suo ruolo nell'ambiente software Epic. Ad esempio, i file di database di produzione sono in genere interessati ai carichi di lavoro più impegnativi, costituiti al 100% da richieste i/o casuali. Il carico di lavoro di qualsiasi database mirror è generalmente meno impegnativo e presenta meno richieste di lettura. I carichi di lavoro dei file di giornale sono principalmente sequenziali.

EPIC mantiene un modello di carico di lavoro per il benchmark delle performance dello storage e i carichi di lavoro del cliente. Per ulteriori informazioni sul modello di workload Epic, sui risultati dei benchmark e sulle

linee guida sull'utilizzo dei tool di dimensionamento NetApp per dimensionare correttamente lo storage per gli ambienti Epic, consulta "TR-3930i: Linee guida per il dimensionamento degli NetApp per Epic" (è richiesto l'accesso NetApp).

EPIC offre inoltre a ciascun cliente una guida alla configurazione dell'hardware customizzata che contiene le proiezioni di i/o e i requisiti della capacità dello storage. I requisiti di storage finali possono includere ambienti di sviluppo, test e/o staging, nonché tutti gli altri carichi di lavoro secondari che potrebbero essere consolidati. I clienti possono utilizzare la guida alla configurazione hardware per comunicare a NetApp i requisiti totali di storage. Questa guida contiene tutti i dati necessari per dimensionare un'implementazione Epic.

Durante la fase di implementazione, Epic mette a disposizione una Database Storage Layout Guide, che fornisce dettagli più granulari a livello di LUN, da utilizzare per una progettazione dello storage avanzata. Tenere presente che la Guida al layout dello storage del database è una soluzione di archiviazione generica e non specifica per NetApp. Utilizza questa guida per determinare il miglior layout di storage su NetApp.

Dimensioni epiche

Una delle principali considerazioni sull'architettura per il dimensionamento di un ambiente storage Epic è la dimensione del database ODB.

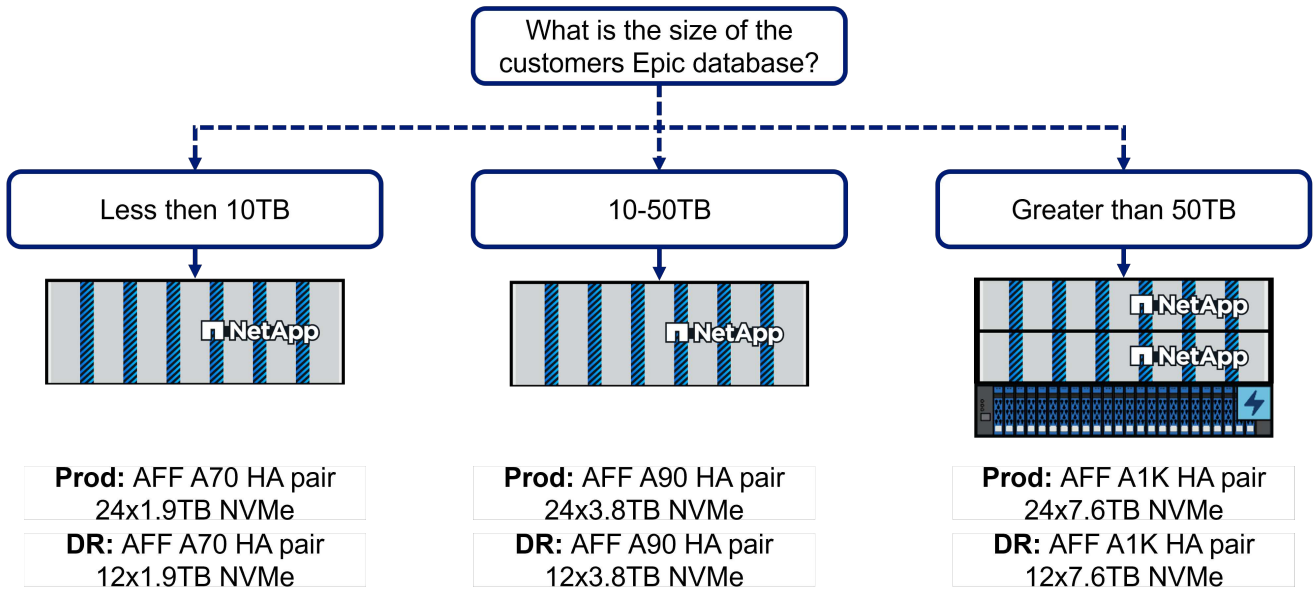
È possibile selezionare un'architettura storage Epic di piccole-medie-grandi dimensioni seguendo il diagramma riportato di seguito. Questi progetti includono l'esecuzione di tutti i carichi di lavoro elencati nella Guida alla configurazione dell'hardware. L'albero di dimensionamento si basa sui dati di oltre 100 guide alla configurazione hardware e dovrebbe essere una stima accurata.

È importante notare che si tratta solo di un punto di partenza. Devi collaborare con il nostro Alliance team Epic per confermare eventuali design Epic. Il team può essere raggiunto a EPIC@NetApp.com. Ogni implementazione deve soddisfare le richieste dei clienti, rispettando al contempo le Best practice consigliate da Epic e NetApp.

- Piccola architettura Epic con un database Epic inferiore a 10TB
- Architettura Epic media con un database Epic dal 10TB al 50TB
- Architettura Epic di grandi dimensioni con un database Epic da oltre 50TB TB

Epic sizing decision tree

Work with the NetApp Epic alliance team to validate designs



Requisiti di storage EPIC

Vengono generalmente fornite risorse di storage dedicate per il database di produzione, mentre le istanze dei database mirrorati condividono risorse di storage secondarie con altri componenti software Epic, come i tool di reporting Clarity.

Altre risorse di storage di software, ad esempio quelle utilizzate per file di sistema e applicazioni, vengono fornite anche dalle risorse di storage secondario.

Oltre alle considerazioni sul dimensionamento, Epic prevede le seguenti regole aggiuntive sul layout dello storage e considerazioni chiave:

- A partire dal 2020, tutti i carichi di lavoro del database operativo (ODB) devono trovarsi su array all-flash.
- EPIC consiglia di collocare ciascun pool di storage su hardware fisico separato, inclusi pool1, pool2, pool3, NAS1 ed NAS2.

> Un nodo in un cluster può essere considerato come un pool di storage. Con ONTAP 9.4 o versioni successive e AQoS, è possibile creare pool protetti utilizzando criteri.

- Nuovi suggerimenti per il backup su Epic 3-2-1.
  a. Copia situata nel sito remoto (disaster recovery)
  b. Una delle copie deve trovarsi su una piattaforma di storage diversa rispetto alla copia primaria
  c. Copie dei dati

> I clienti che utilizzano NetApp SnapMirror per eseguire il backup di NetApp non soddisfano i consigli 3-2-1-5. Il motivo è che ONTAP to ONTAP non soddisfa il secondo requisito sopra elencato. Puoi utilizzare SnapMirror direttamente da ONTAP per lo storage a oggetti on-premise (ad esempio tramite StorageGRID) o nel cloud per soddisfare i requisiti Epic.

Per ulteriori informazioni sugli obblighi di storage, consultare le seguenti guide Epic disponibili in Galaxy:

- Considerazioni sulla SAN
- Stato dei prodotti di storage e della tecnologia (SPATS)
- Guida alla configurazione dell'hardware

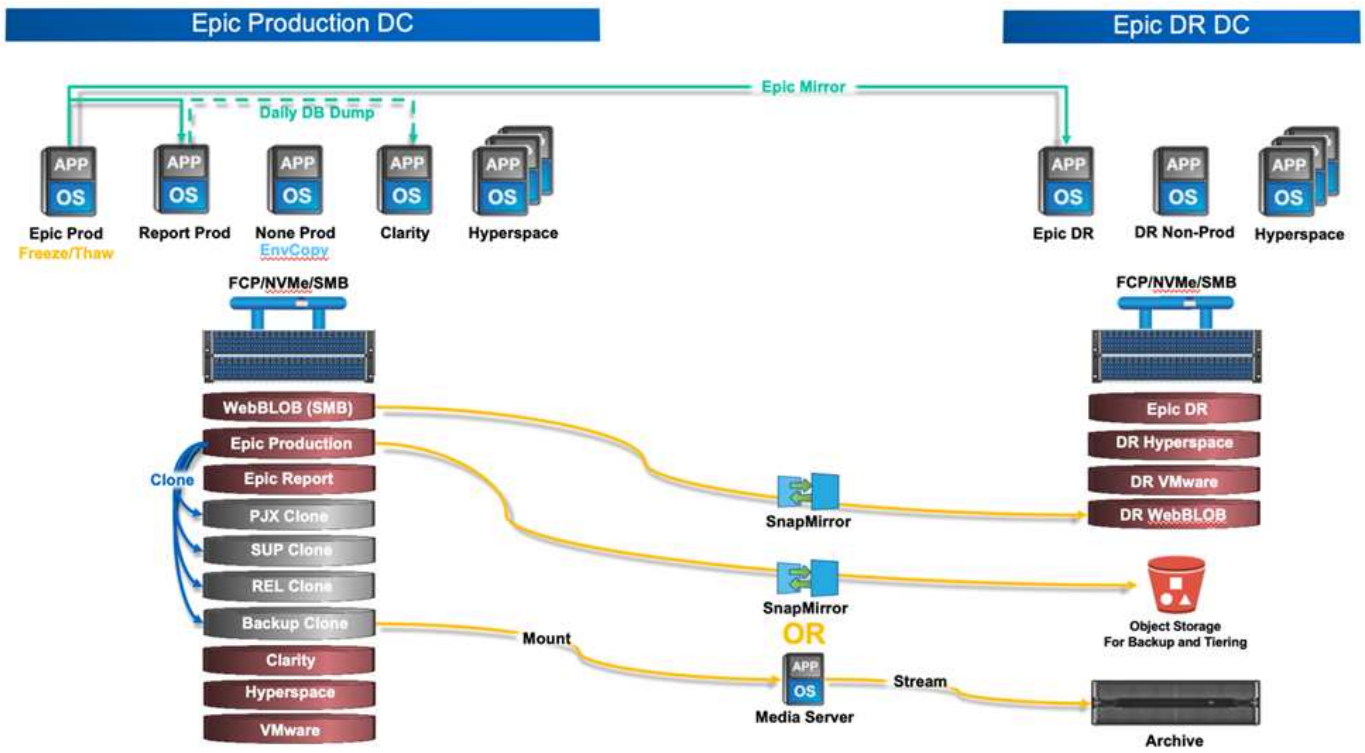
Architettura EPIC a quattro nodi

Le figure qui sotto mostrano il layout dello storage per un'architettura a quattro nodi: Una coppia ha in produzione e una coppia ha in disaster recovery. La dimensione dei controller e il numero di dischi si basano su quest'ultima immagine di dimensionamento.

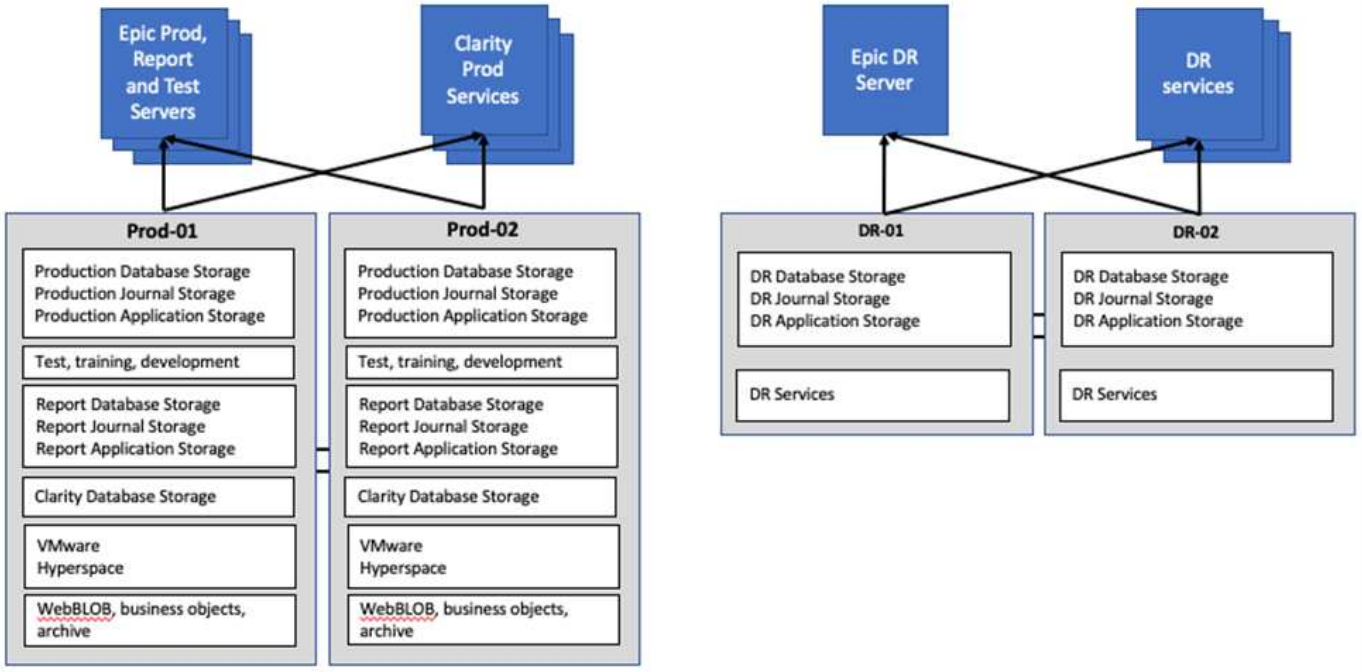
NetApp garantisce prestazioni di livello minimo accettando le policy AQoS consigliate da SLM. EPIC supporta il consolidamento dei pool di storage su ONTAP su hardware notevolmente inferiore. Per ulteriori informazioni, consulta il documento Epic Quarterly SPATS. Fondamentalmente, i sistemi pool1, pool2 e NAS1 (elencati nella Epic hardware Configuration Guide) possono essere eseguiti su una singola coppia ha, con i carichi di lavoro distribuiti in modo uniforme sui due controller. Nelle operazioni di disaster recovery, anche il pool Epic 3 e la NAS 3 sono divisi tra i due controller nella coppia ha.

Gli ambienti di copia completa di test (come SUP, REL e PJX) sono clonati da Epic Production, Epic Report o Epic Disaster Recovery. Per informazioni su backup e refresh Epic, consulta la sezione intitolata, "Gestione dei dati".

Architettura a quattro nodi



Posizionamento dei carichi di lavoro a quattro nodi

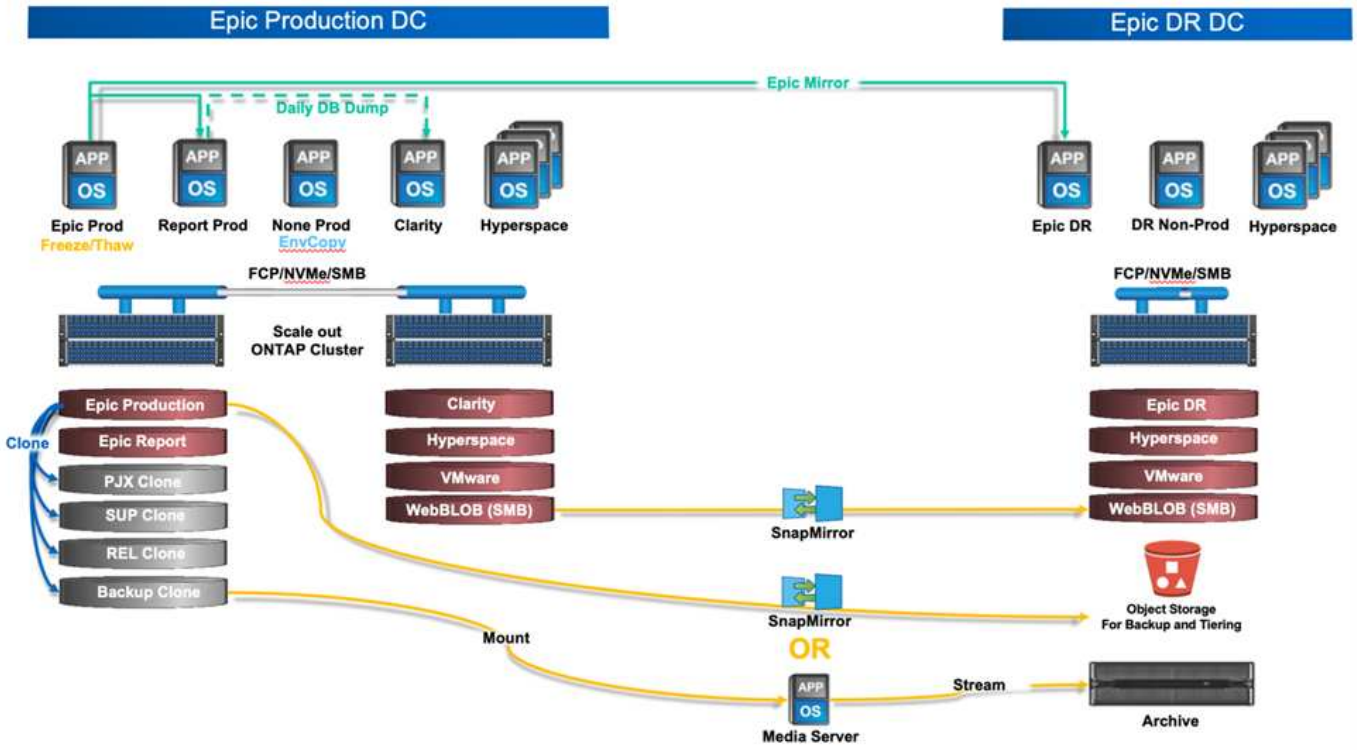


Architettura EPIC a sei nodi

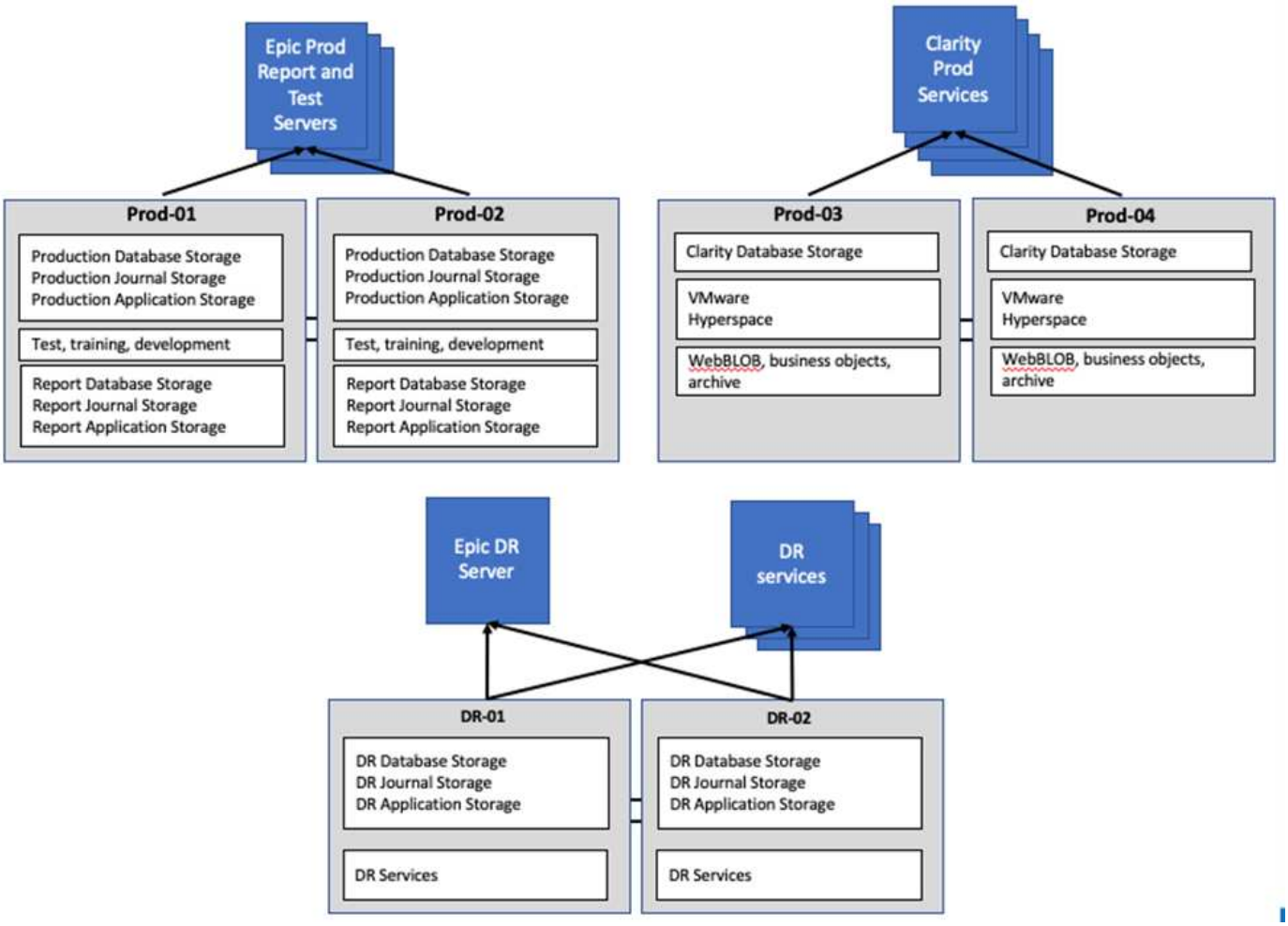
I clienti potrebbero voler iniziare con un design a sei nodi oppure scalare in orizzontale in maniera perfetta da quattro a sei nodi in base alla crescita della domanda. Con lo scale-out è possibile spostare senza interruzioni i carichi di lavoro tra i nodi e ribilanciare nell'intero cluster.

Questa architettura offre il miglior bilanciamento di performance e capacità nel cluster. Produzione di EPIC, Epic Report ed Epic Test vengono eseguiti tutti sulla prima coppia ha. La seconda coppia ha è utilizzata per Clarity, Hyperspace, VMware, NAS1 e i restanti workload Epic. Il disaster recovery corrisponde all'architettura a quattro nodi descritta nella sezione precedente.

Architettura a sei nodi



Posizionamento dei carichi di lavoro a sei nodi

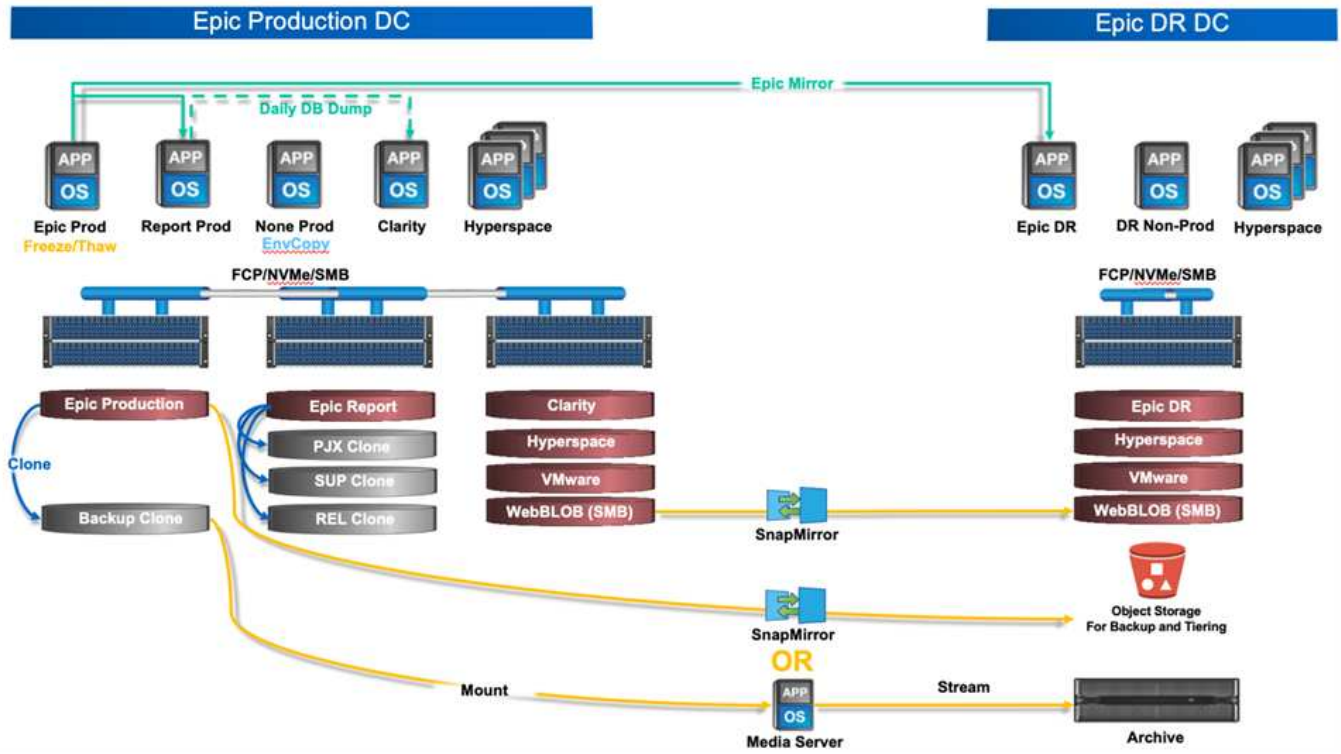


Architettura EPIC a otto nodi

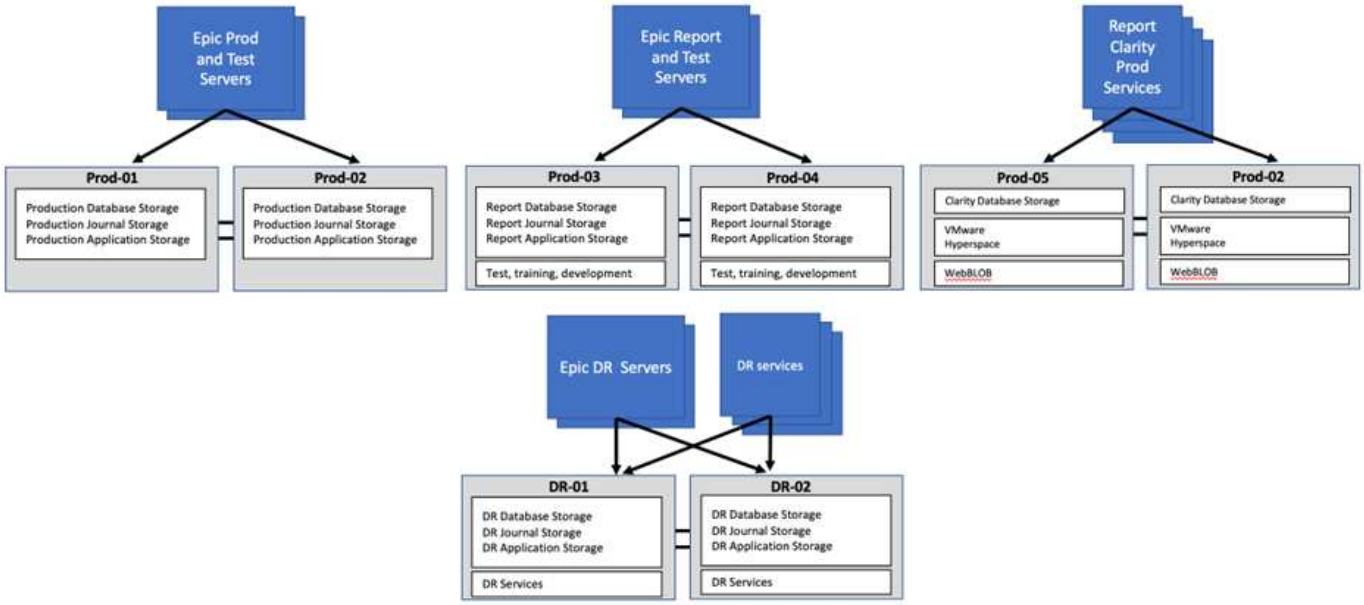
Le figure qui sotto mostrano l'architettura scale-out a otto nodi. Ancora una volta, puoi iniziare con quattro nodi e scalare fino a sei nodi e continuare con la scalabilità fino a otto nodi e oltre. Questa architettura offre il miglior equilibrio di performance e capacità sui sei nodi di produzione.

Gli ambienti di test vengono clonati da Report anziché dalla produzione in questa progettazione. Questo alleggerisce il carico degli ambienti di test e dei controlli dell'integrità dalla produzione.

Architettura a otto nodi



Posizionamento dei carichi di lavoro a otto nodi



Configurazione e Best practice

EPIC su ONTAP - Utilità host

Le utilità host NetApp sono pacchetti software per vari sistemi operativi che contengono utility di gestione come il `sanlun` binario CLI, i driver multipath e altri file importanti necessari per le operazioni SAN corrette.

NetApp consiglia di installare le utilità host NetApp sugli host connessi ai sistemi di archiviazione NetApp e che accedono ad essi. Per ulteriori informazioni, vedere "Tool di matrice di interoperabilità" e "Host SAN" la documentazione.

Con AIX, è particolarmente importante che le utilità host siano installate prima di rilevare i LUN. In questo modo si garantisce che il comportamento del multipathing LUN sia configurato correttamente. Se il rilevamento è stato eseguito senza le utilità host, è necessario deconfigurare i LUN dal sistema utilizzando il `rmdev -dl` comando e quindi eseguire nuovamente il rilevamento tramite `cfgmgr` o un riavvio.

Configurazione EPIC di LUN e volume

Il documento Epic Database Storage Layout Recommendations fornisce indicazioni sulle dimensioni e sul numero di LUN per ogni database.

È importante rivedere questo documento con il supporto degli Epic DBA e Epic così come finalizzare il numero di LUN e le dimensioni delle LUN che potrebbero essere necessarie una modifica. Questi consigli relativi allo storage sono importanti in termini di profondità della coda degli HBA, prestazioni dello storage, facilità di funzionamento e facilità di espansione.

Per la profondità della coda del sistema operativo del server, utilizzare un minimo di otto LUN (un LUN per volume) per un database. Aumentare il numero di LUN in base al numero di nodi nel cluster ONTAP. Ad esempio, Aggiungi 4 LUN quando utilizzi un cluster da 4 nodi (2 coppie ha). Per gli ambienti più grandi potrebbe essere necessario un numero maggiore di LUN, ovvero utilizzare lo stesso numero di volumi (otto

totali distribuiti nel nodo storage) e aggiungere LUN in multipli di due nei nodi e volumi del cluster. Questo approccio ti consente di scalare facilmente il tuo ambiente Epic.

Esempio 1: Cluster ONTAP a 2 nodi

2 nodi, 1 ha Pair 8 volumi, 4 volumi per nodo 8 LUN, una LUN per volume che aggiunge 2 LUN aggiuntive, una su node01 in volume01, una su node02 in volume02.

Esempio 2: Cluster ONTAP a 4 nodi

4 nodi, 2 ha Pair 8 volumi, 2 volumi per nodo 8 LUN, una LUN per volume che aggiunge 4 LUN aggiuntive, una su node01 in volume01, una su node02 in volume02, una su node03 in volume03, una su node04 in volume04.

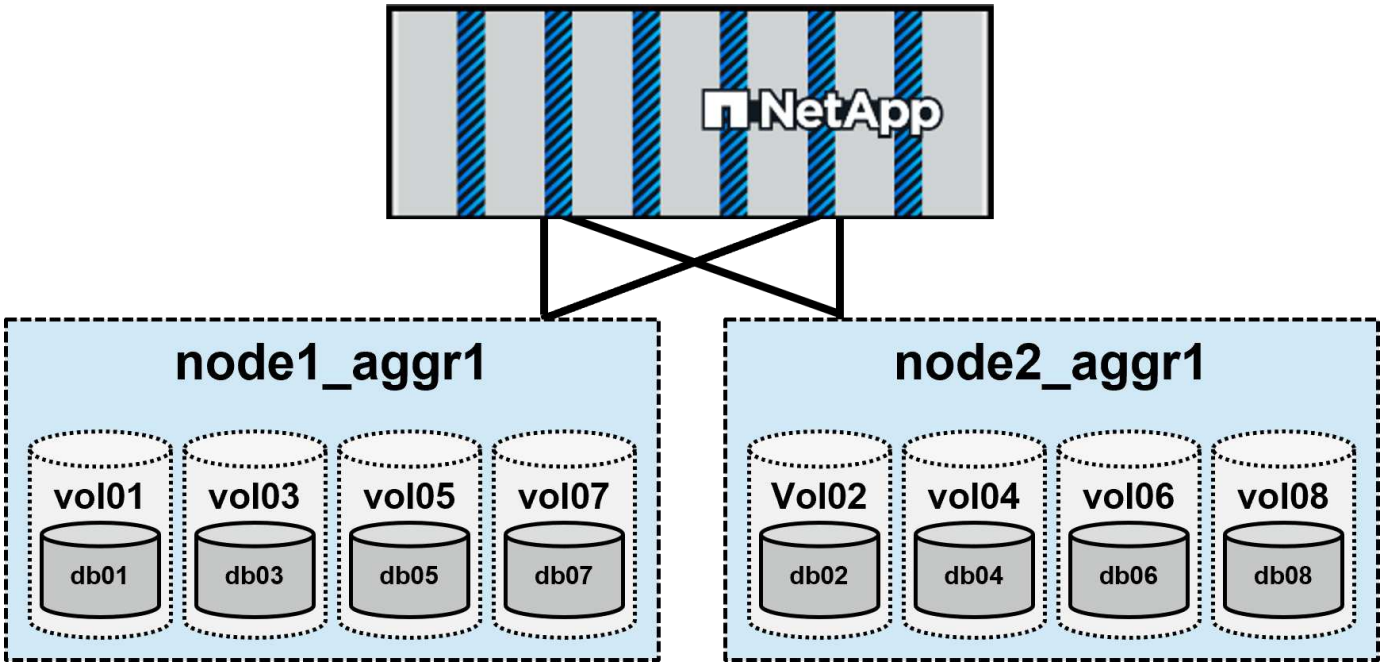
Per massimizzare le performance di un carico di lavoro, ad esempio Epic ODB o Clarity, ogni layout funziona al meglio anche per lo storage NetApp. Utilizzando otto volumi, l'io in scrittura viene distribuito in modo uniforme tra i controller, massimizzando l'utilizzo della CPU. Per la replica e il backup, è consigliabile limitare il numero di volumi a otto per semplificare le operazioni.

Opzioni di scalabilità

Se il server richiede più storage, l'opzione più semplice è aumentare le LUN che contengono volumi. La seconda opzione consiste nell'aggiungere LUN ai gruppi di volumi in multipli di due alla volta (uno per volume per nodo).

Esempio:

Volume e layout di 8 LUN



Se in un ambiente di grandi dimensioni che richiede più di 4 nodi o 8 LUN, consulta il nostro Alliance team Epic per confermare le progettazioni delle LUN. Il team può essere raggiunto a EPIC@NetApp.com.

Best practice

- Utilizzare 8 LUN in volumi 8 per iniziare, aggiungendo 2 LUN alla volta, in tutti i nodi del cluster.
- Bilancia i carichi di lavoro nella coppia ha per massimizzare performance ed efficienza.
- Creare LUN delle dimensioni previste per 3 anni di crescita. (Per conoscere le dimensioni massime dei LUN, consulta la"Documentazione ONTAP").
- USA volumi e LUN con thin provisioning.
- Utilizzare un minimo di otto LUN del database, due LUN di journal e due LUN delle applicazioni. Questa configurazione massimizza le prestazioni dello storage e la profondità della coda del sistema operativo. Se necessario, è possibile utilizzarne di più per motivi di capacità o di altro tipo.
- Se non è necessario aggiungere LUN ai gruppi di volumi, aggiungere otto LUN alla volta.
- I gruppi di coerenza (CGS) sono necessari per il gruppo di volumi e LUN di cui eseguire il backup insieme.
- Non utilizzare la QoS durante Genio o le prestazioni i/O.
- Dopo il test genio o Clarity, NetApp consiglia di eliminare lo storage e di eseguire nuovamente il provisioning prima di caricare i dati di produzione.
- È importante che `-space-allocation` l'opzione abilitato sia impostata sui LUN. In caso contrario, i dati eliminati sulle LUN non verranno rilevati da ONTAP e potrebbero causare problemi di capacità. Per ulteriori informazioni, consulta la Epic Storage Configuration Quick Reference Guide.

Protocolli EPIC e file

È supportata la combinazione di NAS e SAN sullo stesso array all-flash.

NetApp consiglia l'utilizzo di volumi FlexGroup per condivisioni NAS, ad esempio WebBLOB (se disponibile).

WebBLOB offre fino al 95% di dati cold. Puoi facoltativamente liberare spazio sul tuo array all-flash e sfruttare il tiering dei backup e dei dati cold nello storage a oggetti on-premise o nel cloud utilizzando la "FabricPool" funzionalità di ONTAP. Tutto ciò può essere realizzato senza alcun effetto significativo sulle prestazioni. FabricPool è una funzione inclusa in ONTAP. I clienti possono generare un report di dati cold (o inattivi) per esaminare quanti benefici possono essere realizzati abilitando FabricPool. Puoi impostare l'età dei dati da Tier tramite le policy. I clienti EPIC hanno ottenuto risparmi significativi grazie a questa funzionalità.

Gestione epica delle performance

La maggior parte degli array all-flash offre le performance richieste per i workload Epic. Il fattore di differenziazione di NetApp è la sua capacità di impostare policy sulle performance a livello minimo e garantire performance coerenti per ogni applicazione.

Qualità del servizio (QoS)

NetApp consiglia di utilizzare la qualità del servizio. Il beneficio della qualità del servizio è la capacità di consolidare tutti i workload Epic. Tutti i protocolli e i pool di storage possono risiedere su meno hardware. Non hai bisogno di separare pool di storage.

- NetApp consiglia di assegnare tutti i workload nel cluster a una policy di qualità del servizio per una migliore gestione dello spazio nel cluster.
- NetApp consiglia di eseguire un bilanciamento uniforme dei workload nella coppia ha.

- Non utilizzare i criteri QoS durante l'esecuzione di test di i/o; in caso contrario, il test Genio non riuscirà. Analizza i diversi workload di produzione per 2-4 settimane prima di assegnare qualsiasi policy QoS.

EPIC su ONTAP - protocolli

FCP è il protocollo preferito per la presentazione dei LUN.

NetApp Recommended Single Initiator zoning: Un iniziatore per zona con tutte le porte di destinazione richieste sullo storage utilizzando nomi di porte internazionali (WWPN). La presenza di più di un iniziatore in una singola zona può causare un crosstalk intermittente dell'HBA, che causa un'interruzione significativa.

Dopo aver creato il LUN, associare il LUN al gruppo iniziatore (igroup) contenente i WWPN dell'host per abilitare l'accesso.

NetApp supporta anche l'utilizzo di NVMe/FC (se disponi di versioni dei sistemi operativi AIX e RHEL che sono capaci) e migliora le performance. FCP e NVMe/FC possono coesistere sullo stesso fabric.

Configurazione straordinaria dell'efficienza dello storage

Le efficienze inline di ONTAP sono attive per impostazione predefinita e funzionano indipendentemente dal protocollo storage, dall'applicazione o dal Tier storage.

L'efficienza riduce la quantità di dati scritti sul costoso storage flash e limita il numero di dischi richiesti. ONTAP preserva l'efficienza con la replica. Ognuna di queste efficienze ha poco o nessun effetto sulle performance, anche per un'applicazione sensibile alla latenza come Epic.

NetApp consiglia di attivare tutte le impostazioni di efficienza per ottimizzare l'utilizzo del disco. Queste impostazioni sono attivate per impostazione predefinita nei sistemi basati su AFF e ASA.

Le seguenti caratteristiche rendono possibile questa efficienza dello storage:

- La deduplica consente di risparmiare spazio sullo storage primario rimuovendo copie ridondanti dei blocchi di un volume che ospita le LUN. Questa opzione consigliata è attivata per impostazione predefinita.
- La compressione inline riduce la quantità di dati da scrivere su disco e con i workload Epic vengono realizzati considerevoli risparmi di spazio. Questa opzione consigliata è attivata per impostazione predefinita.
- La inline compaction impiega blocchi da 4K KB e li combina in un singolo blocco. Questa opzione consigliata è attivata per impostazione predefinita.
- La replica con risorse limitate è al centro del portfolio di software per la protezione dei dati di NetApp, che include il software NetApp SnapMirror. La thin Replication SnapMirror protegge i dati business-critical riducendo i requisiti di capacità dello storage. **NetApp consiglia** di attivare questa opzione.
- Deduplica aggregata. La deduplica è sempre stata a livello di volume. Con ONTAP 9.2, è stata resa disponibile la deduplica aggregata, che offre ulteriori risparmi in termini di riduzione dei dischi. La deduplica dell'aggregato post-processo è stata aggiunta con ONTAP 9.3. **NetApp consiglia** di attivare questa opzione.

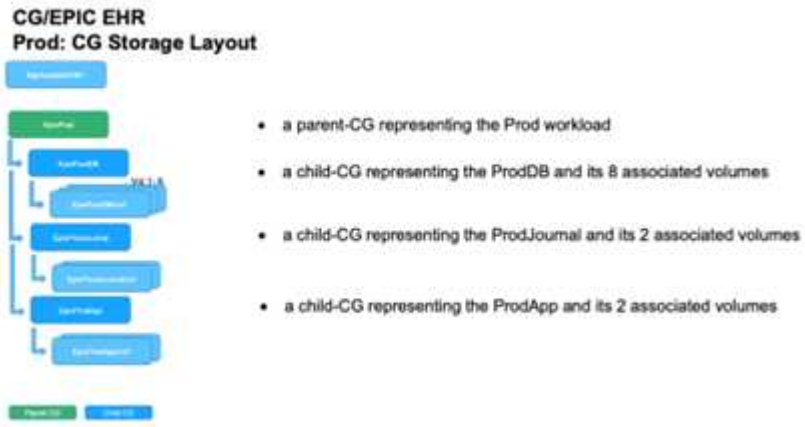
Configurazione straordinaria dell'efficienza dello storage

Le applicazioni con storage distribuito su più di un volume con una o più LUN di quantità appropriate per il carico di lavoro richiedono un backup dei contenuti, garantendo una protezione dei dati coerente.

I gruppi di coerenza (in breve CGS) offrono queste funzionalità e molto altro ancora. Possono essere utilizzate di notte per creare snapshot coerenti on-demand o pianificate utilizzando una policy. Tutto questo è utile per ripristinare, clonare e persino replicare i dati.

Per ulteriori informazioni sul CGS, fare riferimento alla "Panoramica dei gruppi di coerenza"

Una volta eseguito il provisioning di volumi e LUN come descritto nelle precedenti sezioni di questo documento, è possibile configurarli in un set di CGS. La procedura consigliata consiste nel configurarli come illustrato nella figura seguente:



Snapshot di gruppo di coerenza

È necessario impostare una pianificazione degli snapshot del CG notturno su ciascuno dei CGS figlio associati ai volumi che forniscono spazio di archiviazione per il database di produzione. Ciò comporterà una nuova serie di backup coerenti di questi CGS ogni notte. Questi possono essere utilizzati per il cloning del database di produzione per l'utilizzo in ambienti non di produzione, come test e sviluppo. NetApp ha sviluppato workflow Ansible automatizzati basati su CG proprietari per Epic per automatizzare il backup dei database di produzione, il refresh e gli ambienti di test.

Gli snapshot CG possono essere utilizzati per supportare le operazioni di ripristino del database in produzione di Epic.

Per i volumi SAN, disattiva la policy di snapshot predefinita su ciascun volume utilizzato per CGS. Questi snapshot sono generalmente gestiti dall'applicazione di backup in uso o dal servizio di automazione Ansible Epic di NetApp.

Per volumi SAN, disattivare la policy di snapshot predefinita su ogni volume. Questi snapshot vengono in genere gestiti da un'applicazione di backup o dall'automazione Epic Ansible.[NS2]

I set di dati WebBLOB e VMware devono essere configurati come soli volumi, non associati a CGS. Puoi utilizzare SnapMirror per mantenere le snapshot su sistemi storage separati dalla produzione.

Al termine, la configurazione sarà la seguente:

Dimensionamento dello storage per Epic

Devi collaborare con il nostro Alliance team Epic per confermare eventuali design Epic. Il team può essere raggiunto a EPIC@NetApp.com. Ogni implementazione deve soddisfare le richieste dei clienti, rispettando al contempo le Best practice consigliate da Epic e NetApp.

Per informazioni su come utilizzare i tool di dimensionamento NetApp per determinare le dimensioni e il numero corretti dei gruppi RAID per le esigenze di storage dell'ambiente software Epic, consulta "TR-3930i: Linee guida per il dimensionamento degli NetApp per Epic" (è richiesto l'accesso a NetApp).

È richiesto l'accesso al Field Portal di NetApp.

Informazioni aggiuntive per Epic su ONTAP

Per ulteriori informazioni sulle informazioni descritte in questo documento, consultare i seguenti documenti e/o siti Web:

- "Documentazione sui prodotti NetApp"
- "Documentazione di ONTAP 9"
- "Gruppi di coerenza"
- "Risorse per la documentazione di ONTAP e ONTAP System Manager"
- "TR-3930i: Linee guida per il dimensionamento degli NetApp per Epic" (Accesso NetApp richiesto)

Documenti di orientamento epici per i clienti

EPIC fornisce ai clienti i seguenti documenti per la guida su server, storage e rete. Questi documenti sono citati in questo report tecnico.

- Considerazioni sulla Storage Area Network
- Business Continuity Guida alle soluzioni tecniche
- All-Flash - Manuale della strategia per l'architettura di riferimento
- Stato dei prodotti di storage e della tecnologia
- Considerazioni epiche sul cloud

- Guida alla configurazione dell’hardware (specifica del cliente)
- Consigli sul layout dello storage per il database (specifici del cliente)

Hyper-V.

Linee guida per l’implementazione e Best practice per lo storage

Panoramica

Microsoft Windows Server è un sistema operativo di classe Enterprise che copre networking, sicurezza, virtualizzazione, cloud privato, cloud ibrido, infrastruttura desktop virtuale, protezione degli accessi, protezione delle informazioni, servizi web, infrastruttura della piattaforma applicativa, e molto altro ancora.

Questa documentazione sostituisce i report tecnici pubblicati in precedenza *TR-4568: Linee guida per la distribuzione di NetApp e Best practice per lo storage in Windows Server*

ONTAP viene eseguito sugli storage controller NetApp. È disponibile in più formati.

- Un’architettura unificata che supporta protocolli di file, oggetti e blocchi. In questo modo, gli storage controller fungono da dispositivi NAS e SAN e da archivi di oggetti
- Un array All SAN (ASA) incentrato solo sui protocolli a blocchi e che ottimizza i tempi di ripresa dell’i/o (IORT) aggiungendo multipathing Active-Active simmetrico per gli host di connessione
- Un’architettura unificata software-defined
  - ONTAP Select in esecuzione su VMware vSphere o KVM
  - Cloud Volumes ONTAP in esecuzione come istanza cloud nativa
- Offerte di first party degli hyperscale cloud provider
  - Amazon FSX per NetApp ONTAP
  - Azure NetApp Files
  - Google Cloud NetApp Volumes

ONTAP offre funzionalità di efficienza dello storage di NetApp come la tecnologia NetApp Snapshot, il cloning, la deduplica, il thin provisioning, la replica con risorse limitate, la compressione, il virtual storage tiering e molto altro ancora con performance ed efficienza migliorate.

Insieme, Windows Server e ONTAP sono in grado di operare in ambienti di grandi dimensioni e offrire un valore immenso al consolidamento dei data center e alle implementazioni di cloud privato o ibrido. Questa combinazione offre anche carichi di lavoro senza interruzioni in modo efficiente e supporta una scalabilità perfetta.

Pubblico previsto

Il presente documento è destinato agli architetti di sistema e di storage che progettano soluzioni di storage NetApp per server Windows.

Nel presente documento si presuppone quanto segue:

- Il lettore ha una conoscenza generale delle soluzioni hardware e software NetApp. Vedere "Guida all’amministrazione di sistema per gli amministratori di cluster" per ulteriori informazioni.

- Il lettore vanta una conoscenza generale dei protocolli di accesso ai blocchi, quali iSCSI, FC e il protocollo di accesso ai file SMB/CIFS. Vedere "Gestione DELLE SAN di Clustered Data ONTAP" Per informazioni relative alle SAN. Vedere "Gestione NAS" Per informazioni relative a CIFS/SMB.
- Il lettore possiede una conoscenza generale del sistema operativo Windows Server e di Hyper-V.

Per una matrice completa e aggiornata regolarmente di configurazioni SAN e NAS testate e supportate, consulta il "Tool di matrice di interoperabilità (IMT)" sito di supporto NetApp. Con IMT, è possibile determinare le versioni esatte dei prodotti e delle funzionalità supportate per il proprio ambiente specifico. NetApp IMT definisce i componenti e le versioni del prodotto compatibili con le configurazioni supportate da NetApp. I risultati specifici dipendono dall'installazione di ciascun cliente in conformità alle specifiche pubblicate.

Storage NetApp e ambiente Windows Server

Come indicato nella "Panoramica", I controller di storage NetApp forniscono un'architettura realmente unificata che supporta protocolli di file, blocchi e oggetti. Sono inclusi SMB/CIFS, NFS, NVMe/TCP, NVMe/FC, iSCSI, FC(FCP) e S3, inoltre, creano un accesso unificato a client e host. Lo stesso storage controller può offrire simultaneamente un servizio di storage a blocchi sotto forma di LUN SAN e un file service come NFS e SMB/CIFS. ONTAP è disponibile anche come All SAN Array (ASA) in grado di ottimizzare l'accesso host attraverso un multipathing Active-Active simmetrico con iSCSI e FCP, mentre i sistemi ONTAP unificati utilizzano un multipathing Active-Active asimmetrico. In entrambe le modalità, ONTAP utilizza ANA per la gestione multipath NVMe over Fabrics (NVMe-of).

Uno storage controller NetApp che esegue ONTAP può supportare i seguenti carichi di lavoro in un ambiente Windows Server:

- Macchine virtuali in hosting sulle condivisioni SMB 3,0 sempre disponibili
- VM ospitate su LUN CSV (Cluster Shared Volume) in esecuzione su iSCSI o FC
- Database SQL Server su condivisioni SMB 3,0
- Database SQL Server su NVMe-of, iSCSI o FC
- Altri workload delle applicazioni

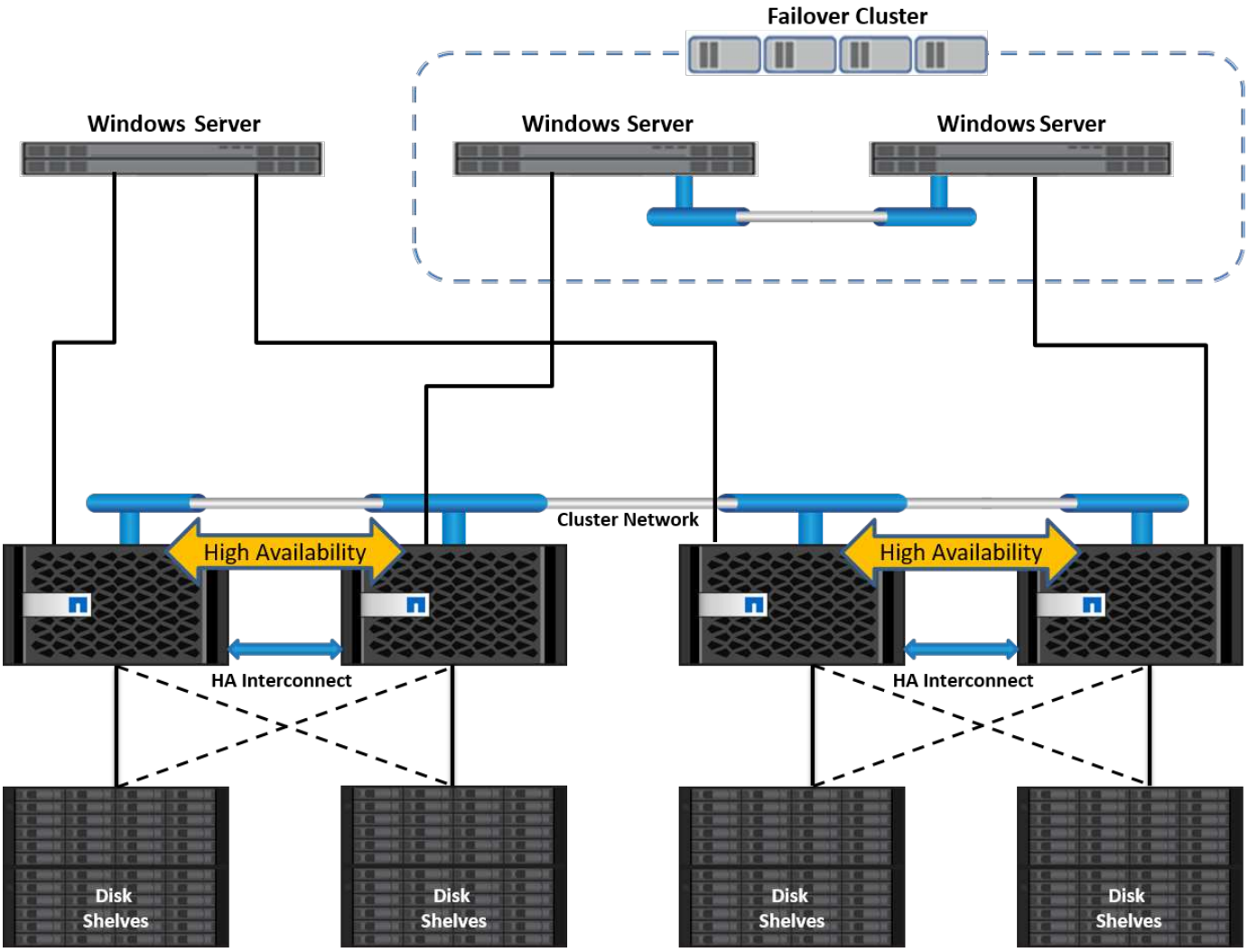
Inoltre, le funzionalità di efficienza dello storage di NetApp come la deduplica, le copie NetApp FlexClone®, la tecnologia Snapshot di NetApp, il thin provisioning, la compressione e il tiering dello storage offrono un valore significativo per i workload in esecuzione su Windows Server.

Gestione dei dati ONTAP

ONTAP è un software di gestione eseguito su uno storage controller del NetApp. Detto nodo, uno storage controller NetApp è un dispositivo hardware dotato di processore, RAM e NVRAM. Il nodo può essere connesso a dischi SATA, SAS o SSD, o a una combinazione di questi dischi.

I nodi multipli vengono aggregati in un sistema in cluster. I nodi nel cluster comunicano continuamente tra loro per coordinare le attività del cluster. I nodi possono anche spostare i dati in modo trasparente da nodo a nodo utilizzando percorsi ridondanti verso un cluster Interconnect dedicato composto da due switch Ethernet 10Gb. I nodi nel cluster possono sostituirsi l'uno all'altro per fornire alta disponibilità in qualsiasi scenario di failover. I cluster vengono amministrati su un intero cluster piuttosto che su un singolo nodo e i dati vengono distribuiti da una o più Storage Virtual Machine (SVM). Un cluster deve avere almeno una SVM per fornire i dati.

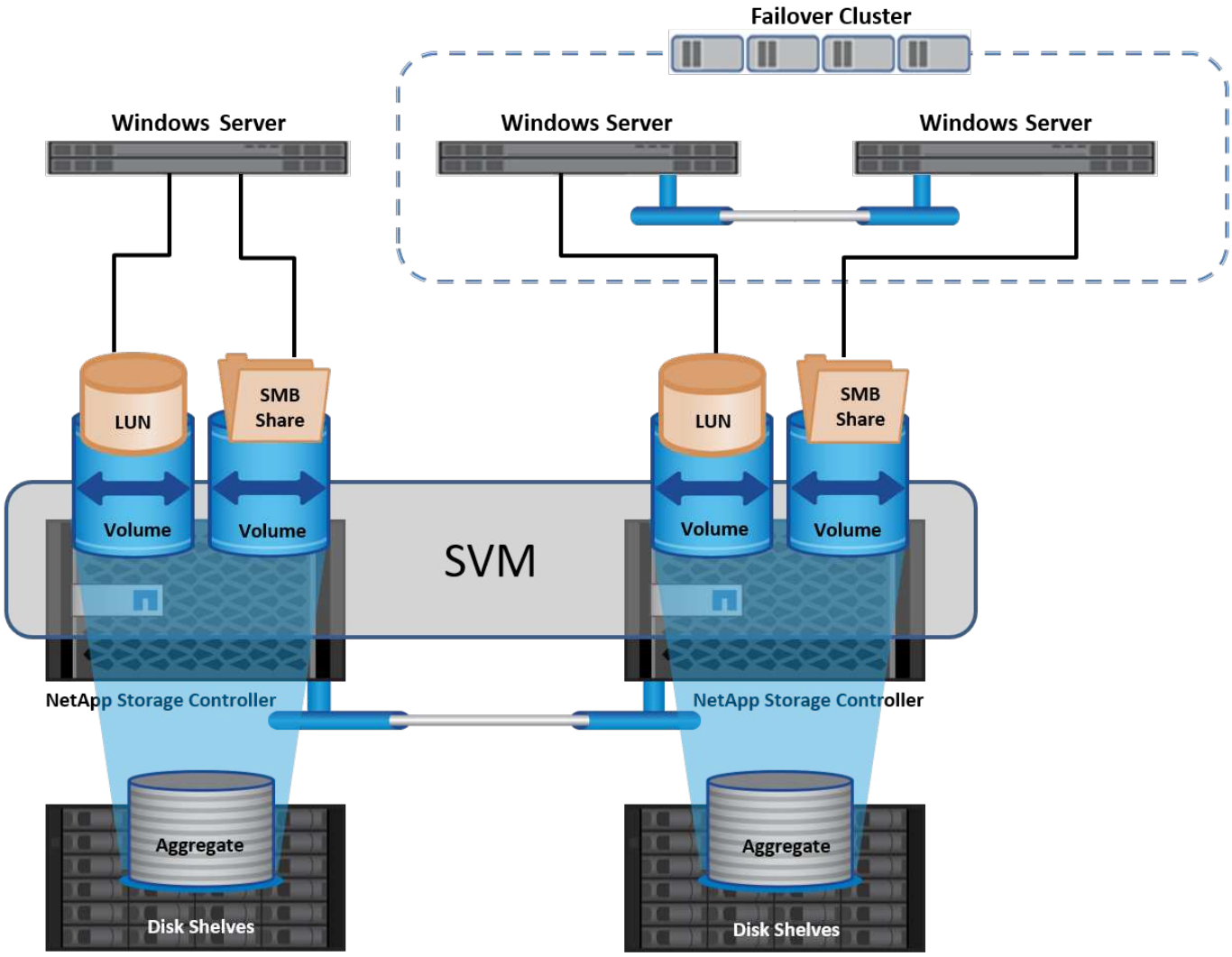
L'unità di base di un cluster è il nodo, che viene aggiunto al cluster nell'ambito di una coppia ha (high Availability). Le coppie HA offrono un'elevata disponibilità comunicando tra loro in un'interconnessione ha (separata dal cluster Interconnect dedicato) e mantenendo le connessioni ridondanti ai dischi della coppia ha. I dischi non sono condivisi tra coppie ha, anche se gli shelf potrebbero contenere dischi appartenenti a uno dei membri di una coppia ha. La figura seguente illustra una distribuzione dello storage NetApp in un ambiente Windows Server.



Macchine virtuali di storage

ONTAP SVM è uno storage server logico che offre l'accesso ai dati di LUN e/o un namespace NAS da una o più interfacce logiche (LIF). La SVM è quindi l'unità di base di segmentazione storage per la multitenancy sicura in ONTAP. Ciascuna SVM è configurata in modo da gestire i volumi storage forniti da un aggregato fisico e da interfacce logiche (LIF) assegnate a una rete Ethernet fisica o a porte di destinazione FC.

I dischi logici (LUN) o le condivisioni CIFS vengono creati all'interno dei volumi di una SVM e vengono mappati agli host e ai cluster Windows per fornire loro spazio di storage, come illustrato nella seguente figura. Le SVM sono indipendenti dai nodi e basate sul cluster e possono utilizzare risorse fisiche come volumi o porte di rete in qualsiasi punto del cluster.



Provisioning dello storage NetApp per Windows Server

È possibile eseguire il provisioning dello storage su Windows Server in ambienti SAN e NAS. In un ambiente SAN, lo storage viene fornito come dischi dalle LUN sul volume NetApp come storage a blocchi. In un ambiente NAS, lo storage viene fornito come condivisioni CIFS/SMB sui volumi NetApp come file storage. I dischi e le condivisioni possono essere applicati in Windows Server nel modo seguente:

- Storage per host Windows Server per workload dell'applicazione
- Stoccaggio per Nano Server e container
- Storage per singoli host Hyper-V per archiviare le macchine virtuali
- Storage condiviso per i cluster Hyper-V sotto forma di CSV per archiviare le VM
- Storage per database SQL Server

Gestione dello storage NetApp

Per connettere, configurare e gestire lo storage NetApp da Windows Server 2016, utilizzare uno dei seguenti metodi:

- **Secure Shell (SSH).** utilizzare qualsiasi client SSH su Windows Server per eseguire i comandi CLI di NetApp.

- **System Manager.** questo è il prodotto di gestibilità basato su GUI di NetApp.
- **Toolkit PowerShell NetApp.** questo è il toolkit PowerShell di NetApp per l'automazione e l'implementazione di script e workflow personalizzati.

Toolkit PowerShell NetApp

Il NetApp PowerShell Toolkit (PSTK) è un modulo PowerShell che offre automazione end-to-end e consente l'amministrazione dello storage di ONTAP. Il modulo ONTAP contiene oltre 2.000 cmdlet e aiuta nell'amministrazione di NetApp AFF, FAS, commodity hardware e risorse cloud.

Cose da ricordare

- NetApp non supporta gli spazi di archiviazione di Windows Server. Gli spazi di archiviazione sono utilizzati solo per JBOD (solo un gruppo di dischi) e non funzionano con alcun tipo di RAID (DAS (Direct-Attached Storage) o SAN).
- I pool di storage in cluster in Windows Server non sono supportati da ONTAP.
- NetApp supporta il formato VHDX (Virtual Hard Disk Format) condiviso per il clustering guest in ambienti SAN Windows.
- Windows Server non supporta la creazione di pool di storage utilizzando LUN iSCSI o FC.

Ulteriori letture

- Per ulteriori informazioni sul toolkit PowerShell di NetApp, visitare il "Sito di supporto NetApp".
- Per informazioni sulle Best practice del toolkit PowerShell di NetApp, vedere "TR-4475: Guida alle Best practice per il toolkit PowerShell di NetApp".

Best practice per il networking

Le reti Ethernet possono essere ampiamente segregate nei seguenti gruppi:

- Una rete client per le VM
- Un'altra rete di storage (connessione iSCSI o SMB ai sistemi di storage)
- Una rete di comunicazione cluster (heartbeat e altre comunicazioni tra i nodi del cluster)
- Una rete di gestione (per monitorare e risolvere i problemi del sistema)
- Una rete di migrazione (per la migrazione live dell'host)
- Replica VM (replica Hyper-V)

Best practice

- NetApp consiglia di disporre di porte fisiche dedicate per ciascuna delle funzionalità precedenti per l'isolamento e le prestazioni della rete.
- Per ciascuno dei precedenti requisiti di rete (ad eccezione dei requisiti di storage), è possibile aggregare più porte di rete fisiche per distribuire il carico o fornire la tolleranza agli errori.
- NetApp consiglia di creare uno switch virtuale dedicato sull'host Hyper-V per la connessione dello storage guest all'interno della macchina virtuale.
- Accertarsi che i percorsi dei dati iSCSI host e guest di Hyper-V utilizzino porte fisiche e switch virtuali diversi per un isolamento sicuro tra l'host e l'host.
- NetApp consiglia di evitare il raggruppamento delle schede di rete per le schede di rete iSCSI.

- NetApp consiglia di utilizzare MPIO (ONTAP Multipath Input/Output) configurato sull'host a scopo di storage.
- NetApp consiglia di utilizzare MPIO all'interno di una macchina virtuale guest se si utilizzano initiator iSCSI guest. L'utilizzo di MPIO deve essere evitato all'interno del guest se si utilizzano dischi pass-through. In questo caso, è sufficiente installare MPIO sull'host.
- NetApp consiglia di non applicare policy di QoS allo switch virtuale assegnato alla rete di storage.
- NetApp consiglia di non utilizzare l'indirizzamento IP privato automatico (APIPA) su schede di rete fisiche, poiché APIPA non è instradabile e non è registrato nel DNS.
- NetApp consiglia di attivare frame jumbo per reti CSV, iSCSI e di migrazione live per aumentare la capacità di trasmissione e ridurre i cicli della CPU.
- NetApp consiglia di deselezionare l'opzione Consenti al sistema operativo di gestione di condividere questa scheda di rete per lo switch virtuale Hyper-V per creare una rete dedicata per le VM.
- NetApp consiglia di creare percorsi di rete ridondanti (switch multipli) per la migrazione live e la rete iSCSI per garantire resilienza e qualità del servizio.

Provisioning negli ambienti SAN

Le SVM di ONTAP supportano i protocolli di blocco iSCSI ed FC. Quando viene creata una SVM con protocollo a blocchi iSCSI o FC, la SVM ottiene rispettivamente un iSCSI Qualified Name (IQN) o un FC Worldwide Name (WWN). Questo identificatore presenta una destinazione SCSI per gli host che accedono allo storage a blocchi NetApp.

Provisioning del LUN NetApp su server Windows

Prerequisiti

L'utilizzo dello storage NetApp in ambienti SAN in Windows Server presenta i seguenti requisiti:

- Un cluster NetApp è configurato con uno o più storage controller NetApp.
- Il cluster NetApp o gli storage controller dispongono di una licenza iSCSI valida.
- Sono disponibili porte configurate iSCSI e/o FC.
- Lo zoning FC viene eseguito su uno switch FC per la connettività FC.
- Viene creato almeno un aggregato.
- Una SVM deve avere una LIF per rete Ethernet o fabric Fibre Channel su ogni storage controller che fornirà dati tramite iSCSI o Fibre Channel.

Implementazione

1. Crea una nuova SVM con protocollo a blocchi iSCSI e/o FC abilitato. È possibile creare una nuova SVM utilizzando uno dei seguenti metodi:
    - Comandi CLI sullo storage NetApp
    - Gestore di sistema di ONTAP
    - Toolkit PowerShell NetApp
2. Configurare il protocollo iSCSI e/o FC.
3. Assegna una SVM con LIF a ciascun nodo del cluster.

4. Avviare il servizio iSCSI e/o FC sulla SVM.

   .

5. Creare set di porte iSCSI e/o FC usando i LIF SVM.
6. Creare un gruppo iniziatore iSCSI e/o FC per Windows utilizzando il set di porte creato.
7. Aggiungere un iniziatore al gruppo iniziatore. L'iniziatore è l'IQN per iSCSI e WWPN per FC. È possibile eseguire una query da Windows Server eseguendo il cmdlet Get-InitatorPort di PowerShell.

```
# Get the IQN for iSCSI
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'iSCSI'} | Select-
Object -Property NodeAddress
```

```
# Get the WWPN for FC
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'Fibre Channel'} |
Select-Object -Property PortAddress
```

```
# While adding initiator to the initiator group in case of FC, make sure
to provide the initiator(PortAddress) in the standard WWPN format
```

IQN per iSCSI su Windows Server può anche essere controllato nella configurazione delle proprietà dell'iniziatore iSCSI.

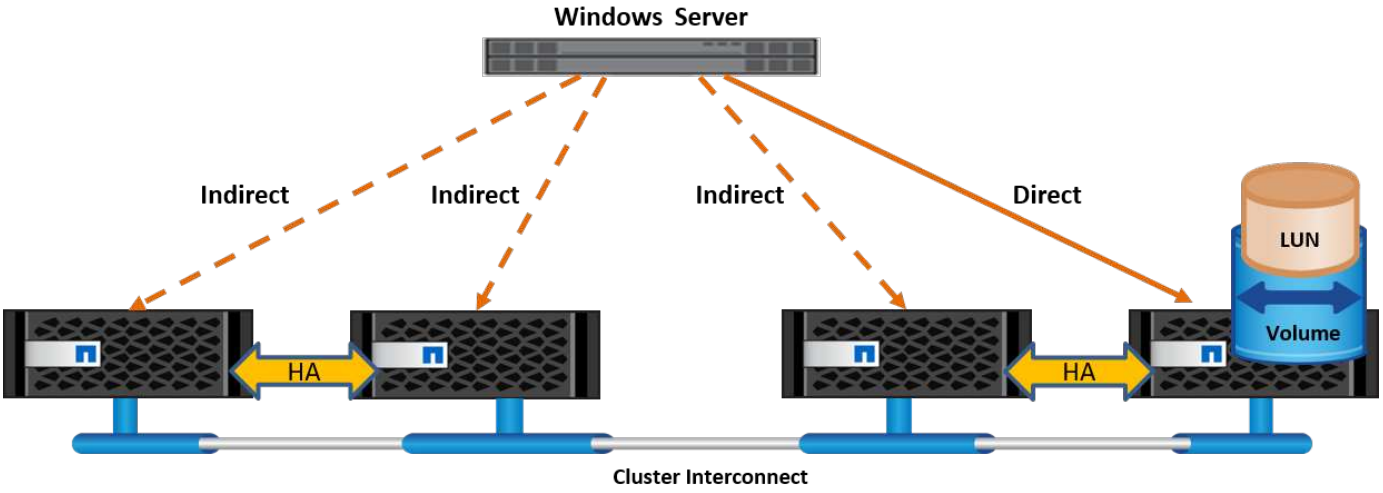
- Creare una LUN mediante la procedura guidata Crea LUN e associarla al gruppo iniziatore creato.

Integrazione host

Windows Server utilizza l'estensione MPIO ALUA (Asymmetric Logical Unit Access) per determinare i percorsi diretti e indiretti verso i LUN. Anche se ogni LIF di proprietà di una SVM accetta richieste di lettura/scrittura per le proprie LUN, solo uno dei nodi del cluster è effettivamente proprietario dei dischi che supportano tale LUN in un dato momento. In questo modo i percorsi disponibili per un LUN vengono suddivisi in due tipi, diretto o indiretto, come illustrato nella figura seguente.

Un percorso diretto per una LUN è un percorso su cui le LIF di una SVM e la LUN a cui si accede risiedono nello stesso nodo. Per passare da una porta di destinazione fisica a un disco, non è necessario attraversare il cluster Interconnect.

I percorsi indiretti sono percorsi di dati su cui si trovano le LIF di una SVM e la LUN a cui si accede su nodi diversi. I dati devono attraversare il cluster Interconnect per passare da una porta di destinazione fisica al disco.



MPIO

ONTAP offre uno storage ad alta disponibilità in cui possono esistere più percorsi dallo storage controller a Windows Server. Il multipathing consente di utilizzare più percorsi dei dati da un server a uno storage array. Il multipathing protegge da guasti hardware (rottura dei cavi, guasto di switch e HBA (host Bus Adapter) e così via) e può offrire limiti di performance più elevati utilizzando le prestazioni aggregate di più connessioni. Quando un percorso o una connessione non è disponibile, il software multipathing sposta automaticamente il carico su uno degli altri percorsi disponibili. La funzione MPIO unisce i diversi percorsi fisici allo storage come unico percorso logico utilizzato per l'accesso ai dati allo scopo di garantire la resilienza dello storage e il bilanciamento del carico. Per utilizzare questa funzione, la funzione MPIO deve essere attivata su Windows Server.

Attiva MPIO

Per attivare MPIO su Windows Server, attenersi alla seguente procedura:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Avviare Server Manager.
3. Nella sezione Gestione, fare clic su Aggiungi ruoli e funzioni.
4. Nella pagina Select Features (Seleziona funzioni), selezionare Multipath i/O.

Configurare MPIO

Quando si utilizza il protocollo iSCSI, è necessario indicare a Windows Server di applicare il supporto multipath ai dispositivi iSCSI nelle proprietà MPIO.

Per configurare MPIO su Windows Server, attenersi alla procedura illustrata di seguito:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Avviare Server Manager.
3. Nella sezione Tools (Strumenti), fare clic su MPIO.
4. In Proprietà MPIO su rileva percorsi multipli, selezionare Aggiungi supporto per dispositivi iSCSI e fare clic su Aggiungi. Viene quindi richiesto di riavviare il computer.
5. Riavviare Windows Server per vedere il dispositivo MPIO elencato nella sezione MPIO Devices (dispositivi MPIO) delle proprietà MPIO.

Configurare iSCSI

Per rilevare lo storage a blocchi iSCSI su Windows Server, attenersi alla seguente procedura:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Avviare Server Manager.
3. Nella sezione Strumenti, fare clic su iSCSI Initiator.
4. Nella scheda rilevamento, fare clic su rileva portale.
5. Fornisci l'indirizzo IP delle LIF associate alla SVM creata per lo storage NetApp per il protocollo SAN. Fare clic su Avanzate, configurare le informazioni nella scheda Generale, quindi fare clic su OK.
6. L'iniziatore iSCSI rileva automaticamente la destinazione iSCSI e la elenca nella scheda Destinazioni.
7. Selezionare la destinazione iSCSI nelle destinazioni rilevate. Fare clic su Connect (Connetti) per aprire la finestra Connect to Target (Connetti a destinazione).
8. È necessario creare sessioni multiple dall'host Windows Server alle LIF iSCSI di destinazione sul cluster storage NetApp. A tale scopo, attenersi alla seguente procedura:
9. Nella finestra connessione a destinazione, selezionare Enable MPIO (attiva MPIO) e fare clic su Advanced (Avanzate).
10. In Impostazioni avanzate nella scheda Generale, selezionare la scheda locale come Microsoft iSCSI Initiator e selezionare l'IP iniziatore e l'IP del portale di destinazione.
11. È inoltre necessario effettuare la connessione utilizzando il secondo percorso. Pertanto, ripetere i passi da 5 a 8, ma questa volta selezionare l'IP iniziatore e l'IP del portale di destinazione per il secondo percorso.
12. Selezionare la destinazione iSCSI nelle destinazioni rilevate nella finestra principale di iSCSI Properties e fare clic su Properties.
13. La finestra Proprietà mostra che sono state rilevate più sessioni. Selezionare la sessione, fare clic su Devices (periferiche), quindi fare clic sul pulsante MPIO per configurare il criterio di bilanciamento del carico. Vengono visualizzati tutti i percorsi configurati per il dispositivo e tutti i criteri di bilanciamento del carico sono supportati. In genere, NetApp consiglia di eseguire il round robin con il sottoinsieme e questa impostazione è quella predefinita per gli array con ALUA attivato. Round robin è l'impostazione predefinita per gli array Active-Active che non supportano ALUA.

Rileva lo storage a blocchi

Per rilevare lo storage a blocchi iSCSI o FC su Windows Server, attenersi alla seguente procedura:

1. Fare clic su Gestione computer nella sezione Strumenti di Gestione server.
2. In Gestione computer, fare clic sulla sezione Gestione disco in archiviazione, quindi fare clic su altre azioni e ripetere la scansione dei dischi. In questo modo vengono visualizzati i LUN iSCSI raw.
3. Fare clic sul LUN rilevato e renderlo online. Quindi selezionare Initialize Disk (Inizializza disco) utilizzando la partizione MBR o GPT. Creare un nuovo volume semplice fornendo le dimensioni del volume e la lettera dell'unità e formattarlo utilizzando FAT, FAT32, NTFS o il file system resiliente (Refs).

Best practice

- NetApp consiglia di attivare il thin provisioning sui volumi che ospitano le LUN.
- Per evitare problemi di multipathing, NetApp consiglia di utilizzare tutte le 10Gb sessioni o tutte le 1Gb sessioni a un determinato LUN.
- NetApp consiglia di confermare l'abilitazione di ALUA nel sistema storage. ALUA è attivato per impostazione predefinita su ONTAP.

- Nell'host del server Windows a cui è mappata la LUN NetApp, attivare il servizio iSCSI (TCP-in) per il servizio in entrata e il servizio iSCSI (TCP-out) per il servizio in uscita nelle impostazioni del firewall. Queste impostazioni consentono il passaggio del traffico iSCSI da e verso l'host Hyper-V e il controller NetApp.

Provisioning delle LUN NetApp sul server Nano

Prerequisiti

Oltre ai prerequisiti menzionati nella sezione precedente, il ruolo di archiviazione deve essere abilitato dal lato server Nano. Ad esempio, Nano Server deve essere distribuito utilizzando l'opzione -Storage. Per distribuire Nano Server, vedere la sezione ""Distribuire Nano Server.""

Implementazione

Per eseguire il provisioning dei LUN NetApp su un server nano, attenersi alla seguente procedura:

1. Connettersi al Nano Server in modalità remota seguendo le istruzioni riportate nella sezione ""Connettersi a Nano Server"."
2. Per configurare iSCSI, eseguire i seguenti cmdlet PowerShell sul Nano Server:

```
# Start iSCSI service, if it is not already running
Start-Service msiscsi
```

```
# Create a new iSCSI target portal
New-IscsiTargetPortal -TargetPortalAddress <SVM LIF>
```

```
# View the available iSCSI targets and their node address
Get-IscsiTarget
```

```
# Connect to iSCSI target
Connect-IscsiTarget -NodeAddress <NodeAddress>
```

```
# NodeAddress is retrived in above cmdlet Get-IscsiTarget
# OR
Get-IscsiTarget | Connect-IscsiTarget
```

```
# View the established iSCSI session
Get-IscsiSession
```

```
# Note the InitiatorNodeAddress retrieved in the above cmdlet Get-
IscsiSession. This is the IQN for Nano server and this needs to be added
in the Initiator group on NetApp Storage
```

```
# Rescan the disks
Update-HostStorageCache
```

3. Aggiungere un iniziatore al gruppo iniziatore.

```
Add the InitiatorNodeAddress retrieved from the cmdlet Get-IscsiSession
to the Initiator Group on NetApp Controller
```

4. Configurare MPIO.

```
# Enable MPIO Feature
Enable-WindowsOptionalFeature -Online -FeatureName MultipathIo
```

```
# Get the Network adapters and their IPs
Get-NetIPAddress -AddressFamily IPv4 -PrefixOrigin <Dhcp or Manual>
```

```
# Create one MPIO-enabled iSCSI connection per network adapter
Connect-IscsiTarget -NodeAddress <NodeAddress> -IsPersistent $True
-IsMultipathEnabled $True -InitiatorPortalAddress <IP Address of
ethernet adapter>
```

```
# NodeAddress is retrieved from the cmdlet Get-IscsiTarget
# IPs are retrieved in above cmdlet Get-NetIPAddress
```

```
# View the connections
Get-IscsiConnection
```

5. Rileva lo storage a blocchi.

```
# Rescan disks
Update-HostStorageCache
```

```
# Get details of disks
Get-Disk
```

```
# Initialize disk
Initialize-Disk -Number <DiskNumber> -PartitionStyle <GPT or MBR>
```

```
# DiskNumber is retrived in the above cmdlet Get-Disk
# Bring the disk online
Set-Disk -Number <DiskNumber> -IsOffline $false
```

```
# Create a volume with maximum size and default drive letter
New-Partition -DiskNumber <DiskNumber> -UseMaximumSize
-AssignDriveLetter
```

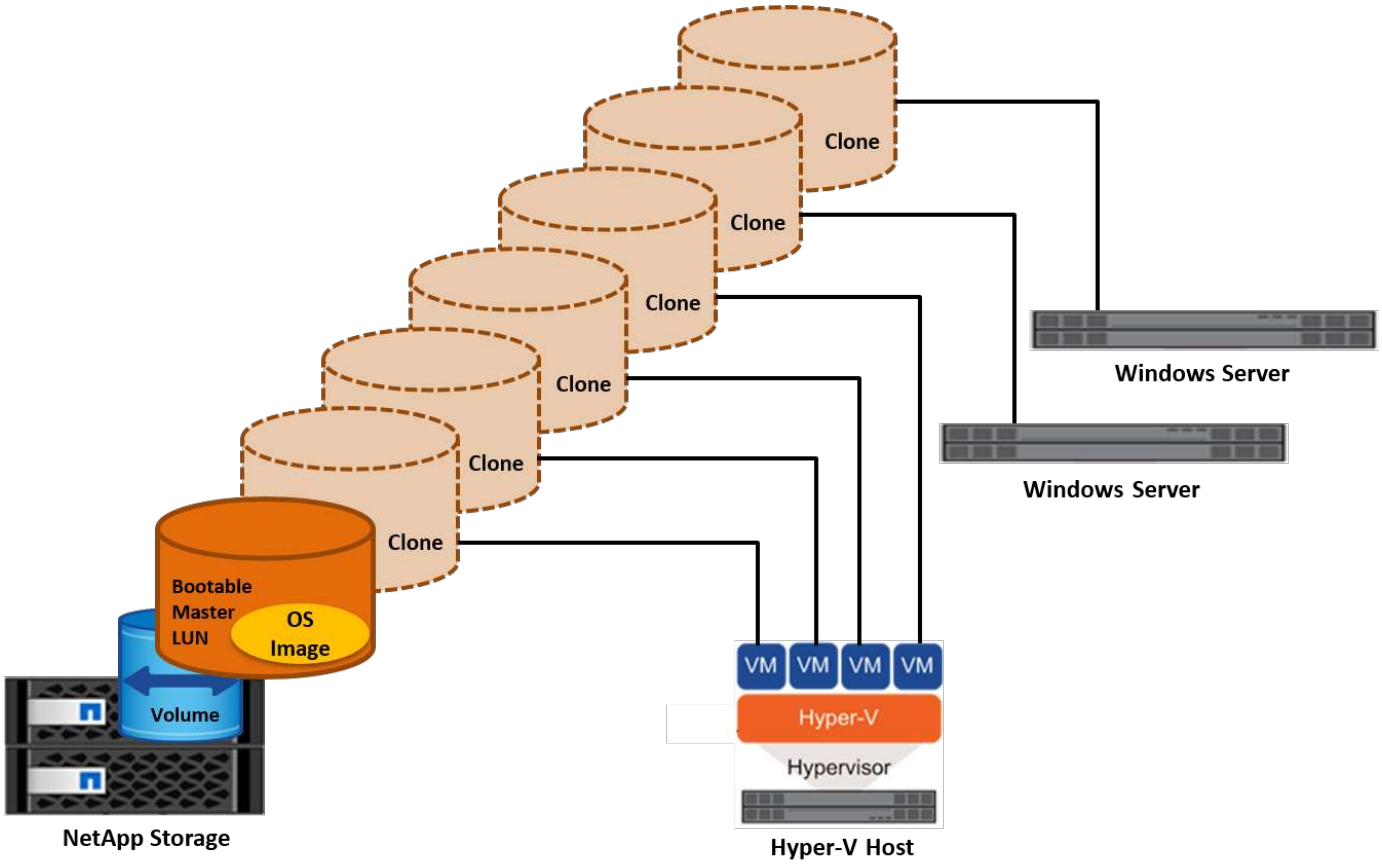
```
# To choose the size and drive letter use -Size and -DriveLetter
parameters
# Format the volume
Format-Volume -DriveLetter <DriveLetter> -FileSystem <FAT32 or NTFS or
REFS>
```

Avvio da SAN

Un host fisico (server) o una macchina virtuale Hyper-V può avviare il sistema operativo Windows Server direttamente da una LUN NetApp invece del disco rigido interno. Nell'approccio all'avvio da SAN, l'immagine del sistema operativo da cui eseguire l'avvio risiede su una LUN NetApp collegata a un host fisico o a una VM. Per un host fisico, l'HBA dell'host fisico è configurato per utilizzare il LUN NetApp per l'avvio. Per una VM, la LUN NetApp è collegata come disco pass-through per l'avvio.

Approccio FlexClone di NetApp

Grazie alla tecnologia NetApp FlexClone, è possibile clonare immediatamente le LUN di avvio con un'immagine del sistema operativo e allegarle ai server e alle macchine virtuali per fornire rapidamente immagini del sistema operativo pulite, come illustrato nella figura seguente.



Avvio da SAN per l'host fisico

Prerequisiti

- L'host fisico (server) dispone di un HBA iSCSI o FC appropriato.
- È stato scaricato un driver di periferica HBA adatto per il server che supporta Windows Server.
- Il server dispone di un'unità CD/DVD o di un supporto virtuale adatto per inserire l'immagine ISO di Windows Server ed è stato scaricato il driver del dispositivo HBA.
- Viene eseguito il provisioning di una LUN iSCSI o FC NetApp sullo storage controller del NetApp.

Implementazione

Per configurare l'avvio da SAN per un host fisico, attenersi alla seguente procedura:

1. Attivare BootBIOS sull'HBA del server.
2. Per gli HBA iSCSI, configurare l'IP iniziatore, il nome del nodo iSCSI e la modalità di avvio della scheda nelle impostazioni del BIOS di avvio.
3. Quando si crea un gruppo iniziatore per iSCSI e/o FC su un controller di storage NetApp, aggiungere l'iniziatore HBA del server al gruppo. L'iniziatore HBA del server è il WWPN per l'HBA FC o il nome del nodo iSCSI per l'HBA iSCSI.
4. Creare un LUN sullo storage controller NetApp con un ID LUN di 0 e associarlo al gruppo iniziatore creato nella fase precedente. Questo LUN serve come LUN di boot.
5. Limitare l'HBA a un singolo percorso verso il LUN di avvio. È possibile aggiungere altri percorsi dopo l'installazione di Windows Server sul LUN di avvio per sfruttare la funzione multipathing.

6. Utilizzare l’utilità BootBIOS dell’HBA per configurare il LUN come dispositivo di avvio.
7. Riavviare l’host e accedere all’utilità BIOS host.
8. Configurare il BIOS host in modo che il LUN di avvio sia il primo dispositivo nell’ordine di avvio.
9. Dall’ISO di Windows Server, avviare il programma di installazione.
10. Quando l’installazione richiede "dove installare Windows?", fare clic su carica driver nella parte inferiore della schermata di installazione per avviare la pagina Seleziona driver da installare. Fornire il percorso del driver di periferica HBA scaricato in precedenza e completare l’installazione del driver.
11. Ora il LUN di avvio creato in precedenza deve essere visibile nella pagina di installazione di Windows. Selezionare il LUN di avvio per l’installazione di Windows Server sul LUN di avvio e terminare l’installazione.

Avvio da SAN per macchina virtuale

Per configurare l’avvio da SAN per una VM, attenersi alla seguente procedura:

Implementazione

1. Quando si crea un gruppo iniziatore per iSCSI o FC su un controller di storage NetApp, aggiungere al controller il codice IQN per iSCSI o il codice WWN per FC del server Hyper-V.
2. Creare LUN o cloni LUN sullo storage controller NetApp e associarli al gruppo iniziatore creato nella fase precedente. Queste LUN fungono da LUN di boot per le macchine virtuali.
3. Rilevare le LUN sul server Hyper-V, portarle online e inizializzarle.
4. Portare i LUN offline.
5. Creare le macchine virtuali con l’opzione Allega un disco rigido virtuale in un secondo momento nella pagina Connetti disco rigido virtuale.
6. Aggiunta di un LUN come disco pass-through a una macchina virtuale.
   a. Aprire le impostazioni VM.
   b. Fare clic su Controller IDE 0, selezionare disco rigido e fare clic su Aggiungi. Selezionando IDE Controller 0 questo disco diventa il primo dispositivo di avvio per la VM.
   c. Selezionare disco rigido fisico nelle opzioni disco rigido e selezionare un disco dall’elenco come disco pass-through. I dischi sono i LUN configurati nelle fasi precedenti.
7. Installare Windows Server sul disco pass-through.

Best practice

- Verificare che i LUN siano offline. In caso contrario, il disco non può essere aggiunto come disco pass-through a una VM.
- Quando esistono più LUN, annotare il numero del disco del LUN nella gestione del disco. Questa operazione è necessaria perché i dischi elencati per la VM sono elencati con il numero del disco. Inoltre, la selezione del disco come disco pass-through per la VM si basa su questo numero di disco.
- NetApp consiglia di evitare il raggruppamento delle schede di rete per le schede di rete iSCSI.
- NetApp consiglia di utilizzare ONTAP MPIO configurato sull’host a scopo di storage.

Provisioning negli ambienti SMB

ONTAP offre storage NAS resiliente e dalle performance elevate per le macchine virtuali

Hyper-V utilizzando il protocollo SMB3.

Al momento della creazione di una SVM con il protocollo CIFS, viene eseguito un server CIFS sopra la SVM che fa parte del dominio Active Directory di Windows. Le condivisioni SMB possono essere utilizzate per una home directory e per ospitare carichi di lavoro Hyper-V e SQL Server. In ONTAP sono supportate le seguenti funzionalità di SMB 3,0:

- Handle persistenti (condivisioni di file sempre disponibili)
- Protocollo testimone
- Failover dei client in cluster
- Consapevolezza in termini di scale-out
- ODX
- VSS remoto

Provisioning delle condivisioni SMB su Windows Server

Prerequisiti

L'utilizzo dello storage NetApp in ambienti NAS in Windows Server presenta i seguenti requisiti:

- Il cluster ONTAP dispone di una licenza CIFS valida.
- Viene creato almeno un aggregato.
- Viene creata una singola interfaccia logica dei dati (LIF) che deve essere configurata per CIFS.
- Sono presenti un server di dominio Windows Active Directory configurato con DNS e credenziali di amministratore di dominio.
- Ogni nodo nel cluster NetApp viene sincronizzato in base all'ora con il controller di dominio Windows.

Controller di dominio Active Directory

È possibile unire e utilizzare uno storage controller NetApp all'interno di un Active Directory simile a un server Windows. Durante la creazione della SVM, è possibile configurare il DNS fornendo i dettagli relativi al nome di dominio e al server dei nomi. La SVM tenta di cercare un controller di dominio Active Directory eseguendo una query al DNS per un server LDAP (Active Directory/Lightweight Directory Access Protocol) in modo simile a Windows Server.

Per il corretto funzionamento della configurazione CIFS, i controller di archiviazione NetApp devono essere sincronizzati a tempo con il controller di dominio Windows. NetApp consiglia di rispettare un intervallo di tempo non superiore a cinque minuti tra il controller di dominio Windows e il controller di storage NetApp. Si consiglia di configurare il server NTP (Network Time Protocol) per il cluster ONTAP in modo che venga sincronizzato con un'origine dell'ora esterna. Per configurare il controller di dominio Windows come server NTP, eseguire il comando seguente sul cluster ONTAP:

```
$domainControllerIP = "<input IP Address of windows domain controller>"
cluster::> system services ntp server create -server $domainControllerIP
```

Implementazione

1. Creazione di una nuova SVM con CIFS del protocollo NAS attivato. È possibile creare una nuova SVM utilizzando uno dei seguenti metodi:

- Comandi CLI su ONTAP
- System Manager
- Il toolkit PowerShell di NetApp

2. Configurare il protocollo CIFS
   a. Fornire il nome del server CIFS.
   b. Fornire l'Active Directory a cui è necessario accedere al server CIFS. È necessario disporre delle credenziali di amministratore del dominio per unirsi al server CIFS in Active Directory.
3. Assegna una SVM con LIF a ciascun nodo del cluster.
4. Avviare il servizio CIFS nell'SVM.
5. Creare un volume con lo stile di protezione NTFS dall'aggregato.
6. Creare un qtree sul volume (opzionale).
7. Creare condivisioni che corrispondono al volume o alla directory del qtree in modo da potervi accedere da Windows Server. Selezionare attiva disponibilità continua per Hyper-V durante la creazione della condivisione, se la condivisione è utilizzata per lo storage Hyper-V. In questo modo, si abilita un'elevata disponibilità per le condivisioni dei file.
8. Modificare la condivisione creata e le autorizzazioni necessarie per accedere alla condivisione. Le autorizzazioni per la condivisione SMB devono essere configurate per consentire l'accesso agli account computer di tutti i server che accedono alla condivisione.

Integrazione host

Il protocollo NAS CIFS è integrato in modo nativo in ONTAP. Pertanto, Windows Server non richiede alcun software client aggiuntivo per accedere ai dati su ONTAP. Uno storage controller NetApp viene visualizzato sulla rete come file server nativo e supporta l'autenticazione Microsoft Active Directory.

Per rilevare la condivisione CIFS creata in precedenza con Windows Server, attenersi alla procedura illustrata di seguito:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Accedere a run.exe e digitare il percorso completo della condivisione CIFS creata per l'accesso alla condivisione.
3. Per mappare in modo permanente la condivisione su Windows Server, fare clic con il pulsante destro del mouse su questo PC, selezionare Connetti unità di rete e specificare il percorso della condivisione CIFS.
4. È possibile eseguire determinate attività di gestione CIFS utilizzando Microsoft Management Console (MMC). Prima di eseguire queste attività, è necessario collegare MMC all'archiviazione ONTAP utilizzando i comandi di menu MMC.
   a. Per aprire MMC in Windows Server, fare clic su Gestione computer nella sezione Strumenti di Gestione server.
   b. Fare clic su altre azioni e su Connetti a un altro computer per aprire la finestra di dialogo Seleziona computer.
   c. Inserisci il nome del server CIFS o l'indirizzo IP del LIF SVM per la connessione al server CIFS.
   d. Espandere Strumenti di sistema e cartelle condivise per visualizzare e gestire file, sessioni e condivisioni aperti.

Best practice

- Per confermare l'assenza di downtime quando un volume viene spostato da un nodo a un altro o in caso di guasto a un nodo, NetApp consiglia di attivare l'opzione di disponibilità continua nella condivisione file.
- Nel provisioning delle macchine virtuali per un ambiente Hyper-V-over-SMB, NetApp consiglia di abilitare l'offload delle copie nel sistema storage. In questo modo si riduce il tempo di provisioning delle VM.
- Se il cluster di storage ospita diversi carichi di lavoro SMB come SQL Server, Hyper-V e server CIFS, NetApp consiglia di ospitare diversi carichi di lavoro SMB su SVM separate in aggregati separati. Questa configurazione è utile perché ciascuno di questi carichi di lavoro garantisce layout unici di volumi e reti di storage.
- NetApp consiglia di collegare gli host Hyper-V e lo storage ONTAP con una rete 10GB, se disponibile. Nel caso della connettività di rete 1GB, NetApp consiglia di creare un gruppo di interfacce composto da più porte 1GB.
- Durante la migrazione di macchine virtuali da una condivisione SMB 3,0 all'altra, NetApp consiglia di attivare la funzionalità di offload delle copie CIFS nel sistema storage, in modo da rendere la migrazione più veloce.

Cose da ricordare

- Quando si eseguono il provisioning di volumi per ambienti SMB, questi volumi devono essere creati con lo stile di protezione NTFS.
- Le impostazioni di tempo sui nodi nel cluster devono essere configurate di conseguenza. Utilizzare il protocollo NTP se il server CIFS NetApp deve far parte del dominio Active Directory di Windows.
- Gli handle persistenti funzionano solo tra nodi in una coppia ha.
- Il protocollo di controllo opera solo tra i nodi in una coppia ha.
- Le condivisioni di file continuamente disponibili sono supportate solo per i workload di Hyper-V e SQL Server.
- Il multicanale SMB è supportato a partire da ONTAP 9,4.
- RDMA non supportato.
- I riferimenti non sono supportati.

Provisioning delle condivisioni SMB su Nano Server

Nano Server non richiede software client aggiuntivo per accedere ai dati della condivisione CIFS su un controller di storage NetApp.

Per copiare file da Nano Server a una condivisione CIFS, eseguire i seguenti cmdlet sul server remoto:

```
$ip = "<input IP Address of the Nano Server>"
```

```
# Create a New PS Session to the Nano Server
$session = New-PSSession -ComputerName $ip -Credential ~\Administrator
```

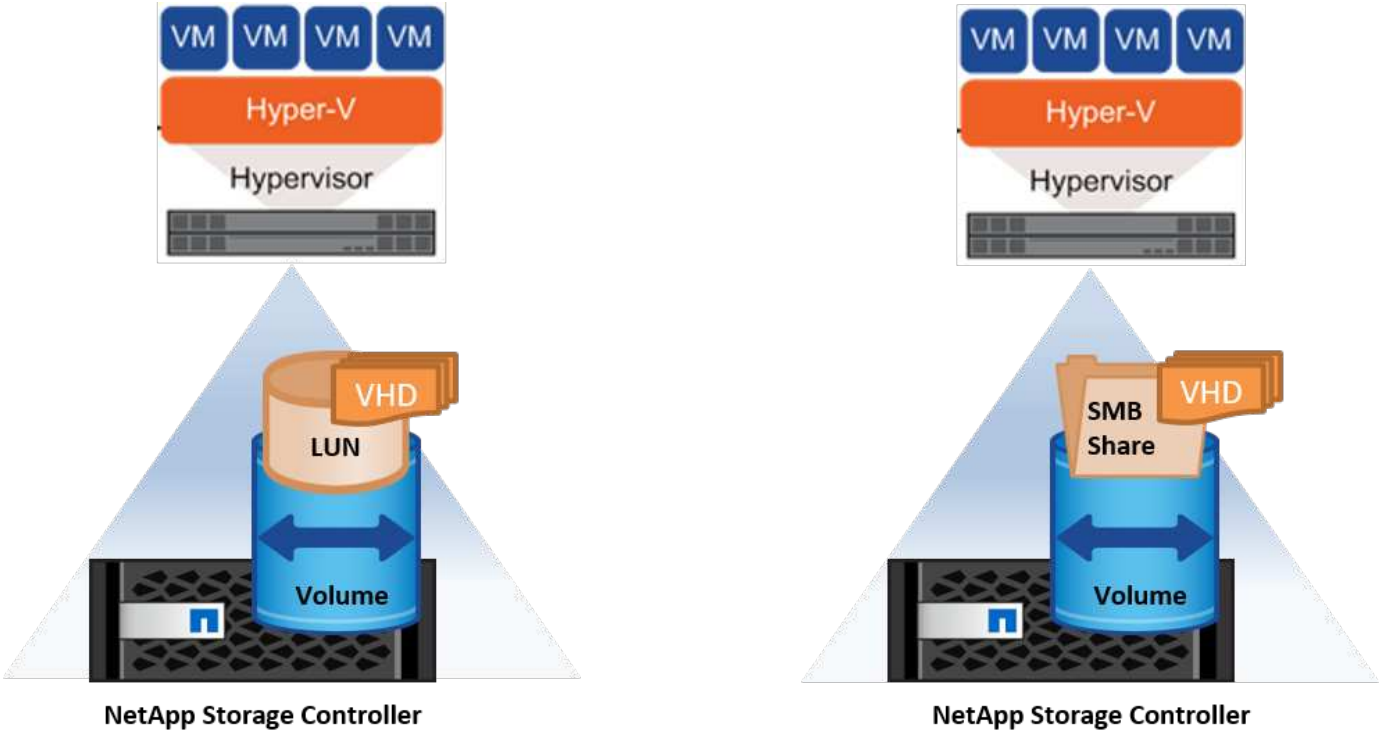
```
 Copy-Item -FromSession $s -Path C:\Windows\Logs\DISM\dism.log
-Destination \\cifsshare
* `cifsshare` È la CIFS share sullo storage controller NetApp.
* Per copiare i file su Nano Server, eseguire il cmdlet seguente:
```

+
Copy-Item -ToSession $s -Path \\cifsshare\<file> -Destination C:\

Per copiare l'intero contenuto di una cartella, specificare il nome della cartella e utilizzare il parametro -Recurse alla fine del cmdlet.

Infrastruttura storage Hyper-V su NetApp

Un'infrastruttura di storage Hyper-V può essere ospitata sui sistemi di storage ONTAP. Lo storage per Hyper-V per la memorizzazione dei file della macchina virtuale e dei relativi dischi può essere fornito utilizzando i LUN di NetApp o le condivisioni CIFS di NetApp, come illustrato nella figura seguente.



Storage Hyper-V su LUN NetApp

- Eseguire il provisioning di una LUN NetApp sulla macchina server Hyper-V. Per ulteriori informazioni, vedere la sezione ""Provisioning negli ambienti SAN"."
- Aprire Hyper-V Manager dalla sezione Strumenti di Server Manager.
- Selezionare il server Hyper-V e fare clic su Impostazioni Hyper-V.
- Specificare la cartella predefinita in cui memorizzare la VM e il relativo disco come LUN. In questo modo, viene impostato il percorso predefinito come LUN per lo storage Hyper-V. Se si desidera specificare esplicitamente il percorso di una VM, è possibile farlo durante la creazione.

Storage Hyper-V su NetApp CIFS

Prima di iniziare i passaggi elencati in questa sezione, rivedere la sezione ""Provisioning negli ambienti SMB"." Per configurare lo storage Hyper-V sulla CIFS share di NetApp, attenersi alla seguente procedura:

1. Aprire Hyper-V Manager dalla sezione Strumenti di Server Manager.
2. Selezionare il server Hyper-V e fare clic su Impostazioni Hyper-V.
3. Specificare la cartella predefinita in cui memorizzare la macchina virtuale e il relativo disco come condivisione CIFS. In questo modo, viene impostato il percorso predefinito come CIFS share per lo storage Hyper-V. Se si desidera specificare esplicitamente il percorso di una VM, è possibile farlo durante la creazione.

A sua volta, ciascuna macchina virtuale di Hyper-V può essere fornita con i LUN di NetApp e le condivisioni CIFS fornite all'host fisico. Questa procedura è la stessa di qualsiasi host fisico. È possibile utilizzare i seguenti metodi per il provisioning dello storage su una macchina virtuale:

- Aggiunta di una LUN di storage tramite FC Initiator all'interno della macchina virtuale
- Aggiunta di una LUN di storage tramite l'iniziatore iSCSI all'interno della macchina virtuale
- Aggiunta di un disco fisico pass-through a una VM
- Aggiunta di VHD/VHDX a una VM dall'host

Best practice

- Quando una macchina virtuale e i relativi dati vengono memorizzati nello storage NetApp, NetApp consiglia di eseguire la deduplica NetApp a livello di volume a intervalli regolari. Questa pratica consente notevoli risparmi di spazio quando macchine virtuali identiche vengono ospitate in una condivisione CSV o SMB. La deduplica viene eseguita sullo storage controller e non influisce sul sistema host e sulle performance delle macchine virtuali.
- Quando si utilizzano LUN iSCSI per Hyper-V, accertarsi di abilitare "Servizio iSCSI (TCP-in) per l'ingresso" e "Servizio iSCSI (TCP-out) per l'uscita" nelle impostazioni del firewall sull'host Hyper-V. In questo modo, si consente il passaggio del traffico iSCSI da e verso l'host Hyper-V e il controller NetApp.
- NetApp consiglia di deselezionare l'opzione Consenti al sistema operativo di gestione di condividere la scheda di rete per lo switch virtuale Hyper-V. In questo modo si crea una rete dedicata per le VM.

Cose da ricordare

- Il provisioning di una macchina virtuale tramite Fibre Channel virtuale richiede un HBA FC abilitato alla virtualizzazione N_Port ID. È supportato un massimo di quattro porte FC.
- Se il sistema host è configurato con più porte FC e presentato alla macchina virtuale, MPIO deve essere installato nella macchina virtuale per consentire il multipathing.
- Non è possibile fornire all'host dischi pass-through se si utilizza MPIO su quell'host, poiché i dischi pass-through non supportano MPIO.
- Il disco utilizzato per i file VHD/VHDx deve utilizzare la formattazione 64K per l'allocazione.

Ulteriori letture

- Per informazioni sugli HBA FC, vedere "Tool di matrice di interoperabilità (IMT)" .
- Per ulteriori informazioni su Fibre Channel virtuale, consultare Microsoft "Panoramica di Hyper-V Virtual Fibre Channel" pagina.

Trasferimento dei dati con offload

Microsoft ODX, noto anche come offload delle copie, permette trasferimenti dei dati diretti all'interno di un dispositivo di storage o tra dispositivi di storage compatibili senza trasferire i dati attraverso il computer host. ONTAP supporta la funzionalità ODX per i protocolli CIFS e SAN. ODX può potenzialmente migliorare le performance se le copie si trovano all'interno dello stesso volume, ridurre l'utilizzo della CPU e della memoria sul client e ridurre l'utilizzo della larghezza di banda i/o di rete.

Con ODX, è più rapido ed efficiente copiare i file all'interno delle condivisioni SMB, nelle LUN e tra le condivisioni SMB e le LUN, se si trovano nello stesso volume. Questo approccio risulta più utile in uno scenario in cui sono necessarie più copie dell'immagine dorata di un sistema operativo (VHD/VHDX) all'interno dello stesso volume. Se le copie si trovano all'interno dello stesso volume, è possibile eseguire più copie della stessa immagine Golden in tempi notevolmente inferiori. ODX viene applicato anche nella migrazione live dello storage Hyper-V per lo spostamento dello storage delle macchine virtuali.

Se la copia è tra i volumi, potrebbe non esserci un aumento significativo delle prestazioni rispetto alle copie basate su host.

Per abilitare la funzionalità ODX su CIFS, esegui i seguenti comandi dell'interfaccia a riga di comando sullo storage controller NetApp:

1. Abilita ODX per CIFS.
   #impostare il livello di privilegio su diagnostico
   cluster::> diagnostica set -privilege

   ```
   #enable the odx feature
   cluster::> vserver cifs options modify -vserver <vserver_name> -copy
   -offload-enabled true
   ```

   ```
   #return to admin privilege level
   cluster::> set privilege admin
   ```

2. Per abilitare la funzionalità ODX su SAN, esegui i seguenti comandi dell'interfaccia a riga di comando sullo storage controller NetApp:
   #impostare il livello di privilegio su diagnostico
   cluster::> diagnostica set -privilege

   ```
   #enable the odx feature
   cluster::> copy-offload modify -vserver <vserver_name> -scsi enabled
   ```

   ```
   #return to admin privilege level
   cluster::> set privilege admin
   ```

Cose da ricordare

- Per CIFS, ODX è disponibile solo se sia il client che il server di storage supportano SMB 3,0 e la

funzionalità ODX.

- Per gli ambienti SAN, l'ODX è disponibile solo se sia il client che il server di storage supportano la funzione ODX.

Ulteriori letture

Per informazioni su ODX, vedere "Miglioramento delle prestazioni di Microsoft Remote Copy" e. "Trasferimenti dati con offload Microsoft" .

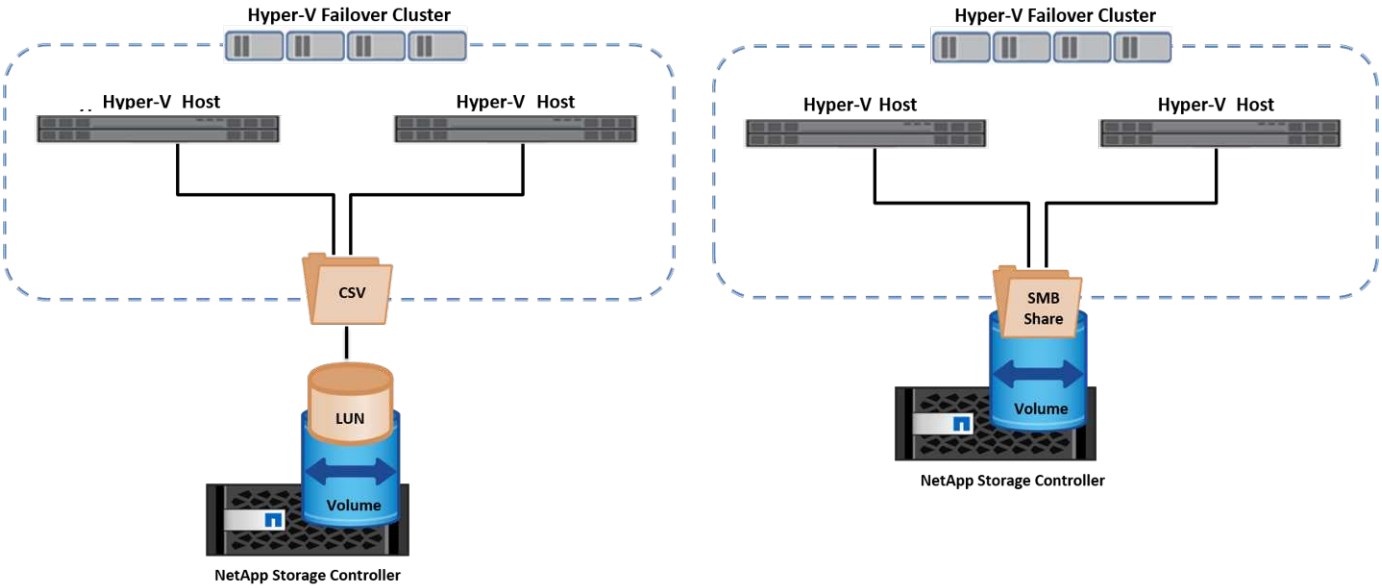
Clustering Hyper-V: Alta disponibilità e scalabilità per le macchine virtuali

I cluster di failover offrono disponibilità e scalabilità elevate per i server Hyper-V. Un cluster di failover è un gruppo di server Hyper-V indipendenti che lavorano insieme per aumentare la disponibilità e la scalabilità delle VM.

I server in cluster Hyper-V (detti nodi) sono connessi dalla rete fisica e da un software cluster. Questi nodi utilizzano lo storage condiviso per archiviare i file delle macchine virtuali, che includono configurazione, file dell'hard disk virtuale (VHD) e snapshot. Lo storage condiviso può essere una condivisione SMB/CIFS di NetApp o un CSV posto sopra una LUN NetApp, come illustrato di seguito. Si tratta di uno storage condiviso che offre un namespace coerente e distribuito, a cui tutti i nodi del cluster possono accedere contemporaneamente. Pertanto, se un nodo si guasta nel cluster, l'altro nodo fornisce il servizio mediante un processo chiamato failover. I cluster di failover possono essere gestiti utilizzando lo snap-in failover Cluster Manager e i cmdlet Windows PowerShell per il clustering di failover.

Volumi condivisi del cluster

I CSV consentono a più nodi in un cluster di failover di avere contemporaneamente l'accesso in lettura/scrittura allo stesso LUN NetApp su cui viene eseguito il provisioning di un volume NTFS o refs. Con i CSV, è possibile eseguire rapidamente il failover di ruoli in cluster da un nodo a un altro senza richiedere una modifica della proprietà delle unità o lo smontaggio e rimontaggio di un volume. I CSV semplificano inoltre la gestione di un numero potenzialmente elevato di LUN in un cluster di failover. I CSV forniscono un file system in cluster per scopi generali, ad esempio superiore a NTFS o Ref.



Best practice

- NetApp consiglia di disattivare la comunicazione del cluster sulla rete iSCSI per impedire il flusso di comunicazioni interne del cluster e del traffico CSV sulla stessa rete.
- NetApp consiglia di disporre di percorsi di rete ridondanti (switch multipli) per garantire resilienza e qualità del servizio.

Cose da ricordare

- I dischi utilizzati per CSV devono essere partizionati con NTFS o Rif. I dischi formattati con FAT o FAT32 non possono essere utilizzati per un CSV.
- I dischi utilizzati per i CSV devono utilizzare la formattazione 64K per l'allocazione.

Ulteriori letture

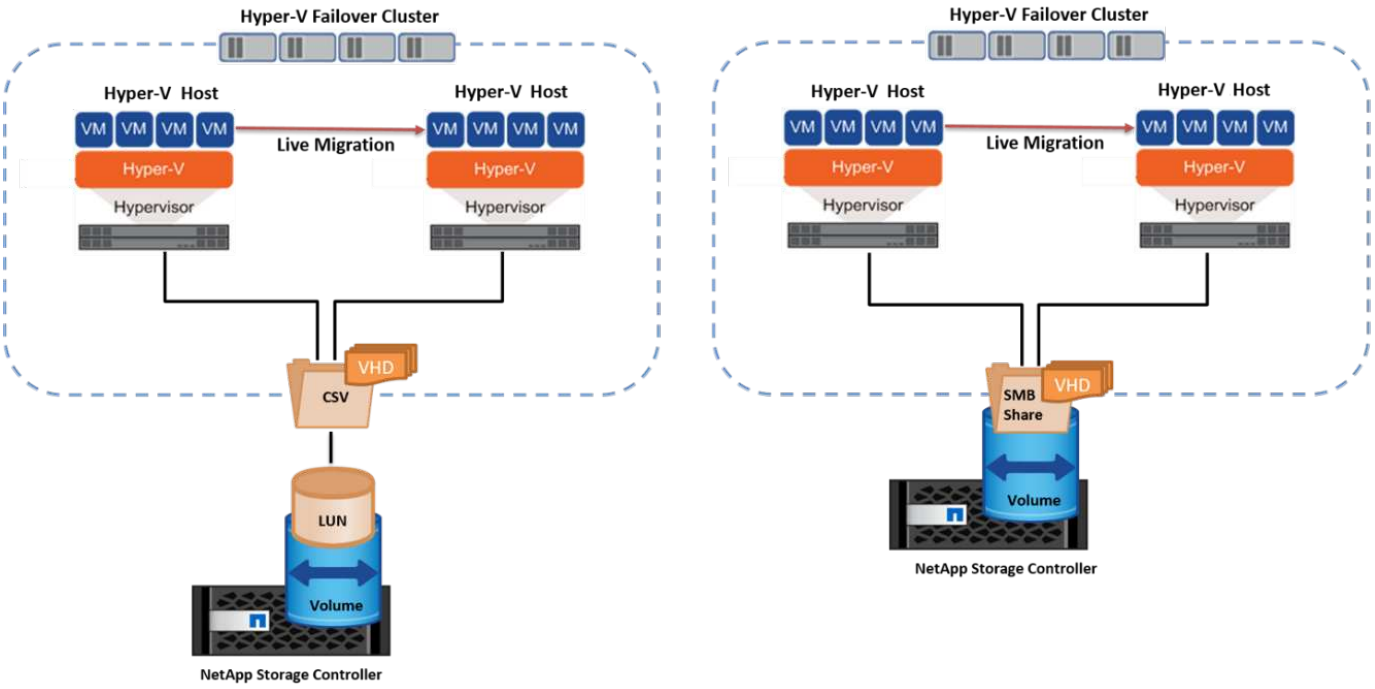
Per informazioni sull'implementazione di un cluster Hyper-V, fare riferimento all'Appendice B: "Distribuire il cluster Hyper-V.".

Hyper-V Live Migration: Migrazione delle VM

A volte è necessario, durante il ciclo di vita delle macchine virtuali, spostarle in un altro host del cluster Windows. Questa operazione potrebbe essere necessaria se l'host sta esaurendo le risorse del sistema o se è necessario riavviare l'host per motivi di manutenzione. Analogamente, potrebbe essere necessario spostare una macchina virtuale in un LUN o una condivisione SMB differente. Ciò potrebbe essere necessario se lo spazio del LUN o della condivisione attuale sta per esaurirsi o sta producendo prestazioni inferiori al previsto. La migrazione live di Hyper-V sposta le macchine virtuali in esecuzione da un server Hyper-V fisico all'altro senza alcun effetto sulla disponibilità delle macchine virtuali per gli utenti. È possibile eseguire in tempo reale la migrazione di macchine virtuali tra server Hyper-V che fanno parte di un cluster di failover o tra server Hyper-V indipendenti che non fanno parte di un cluster.

Live Migration in un ambiente in cluster

È possibile spostare perfettamente le macchine virtuali tra i nodi di un cluster. La migrazione delle macchine virtuali è istantanea perché tutti i nodi del cluster condividono lo stesso storage e hanno accesso alla macchina virtuale e al relativo disco. La figura seguente illustra la migrazione live in un ambiente in cluster.



Best practice

- Disporre di una porta dedicata per il traffico di migrazione live.
- Disporre di una rete host di migrazione live dedicata per evitare problemi relativi alla rete durante la migrazione.

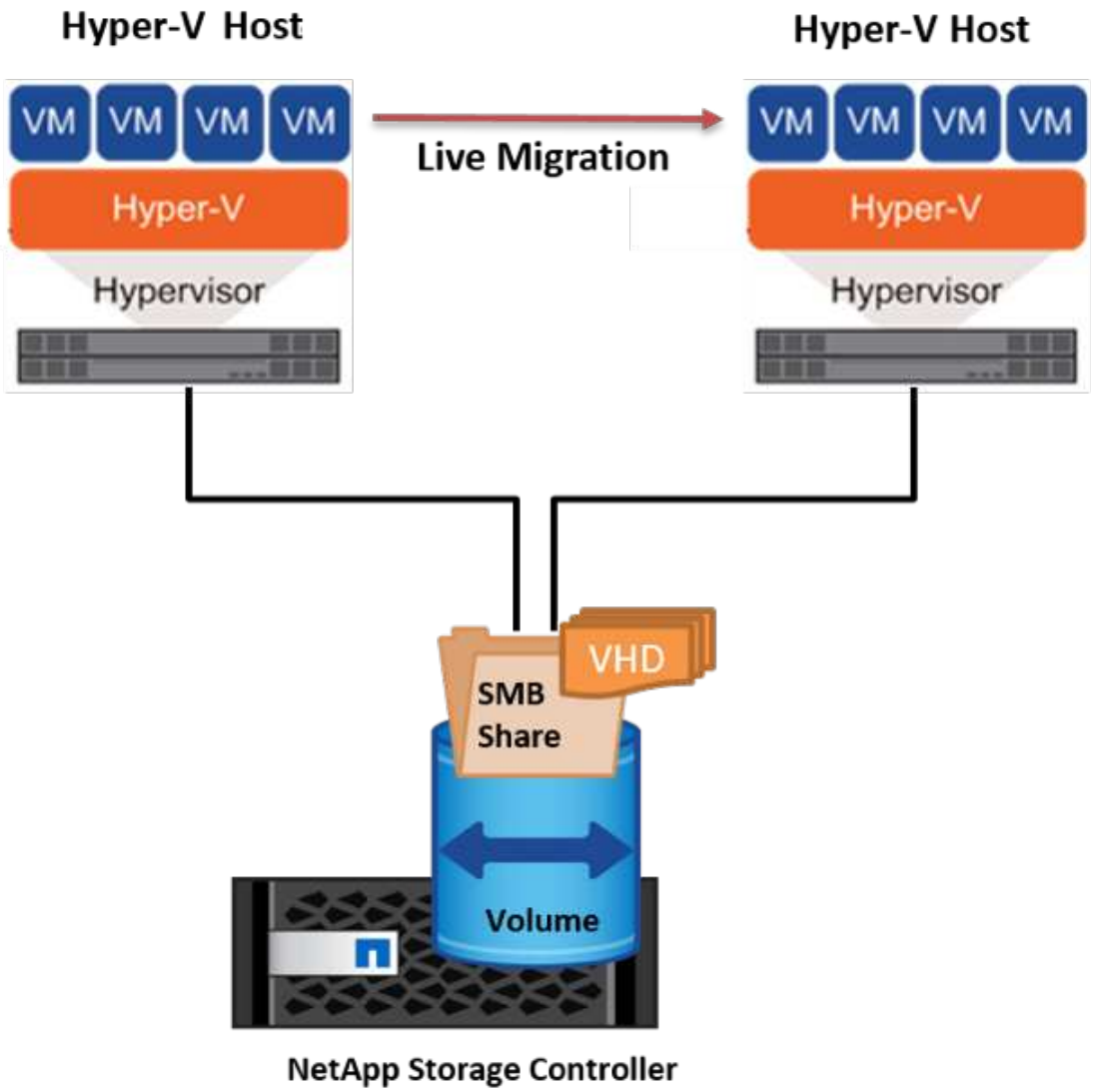
Ulteriori letture

Per informazioni sulla distribuzione della migrazione live in un ambiente in cluster, vedere "Appendice C: Implementare Hyper-V Live Migration in un ambiente cluster".

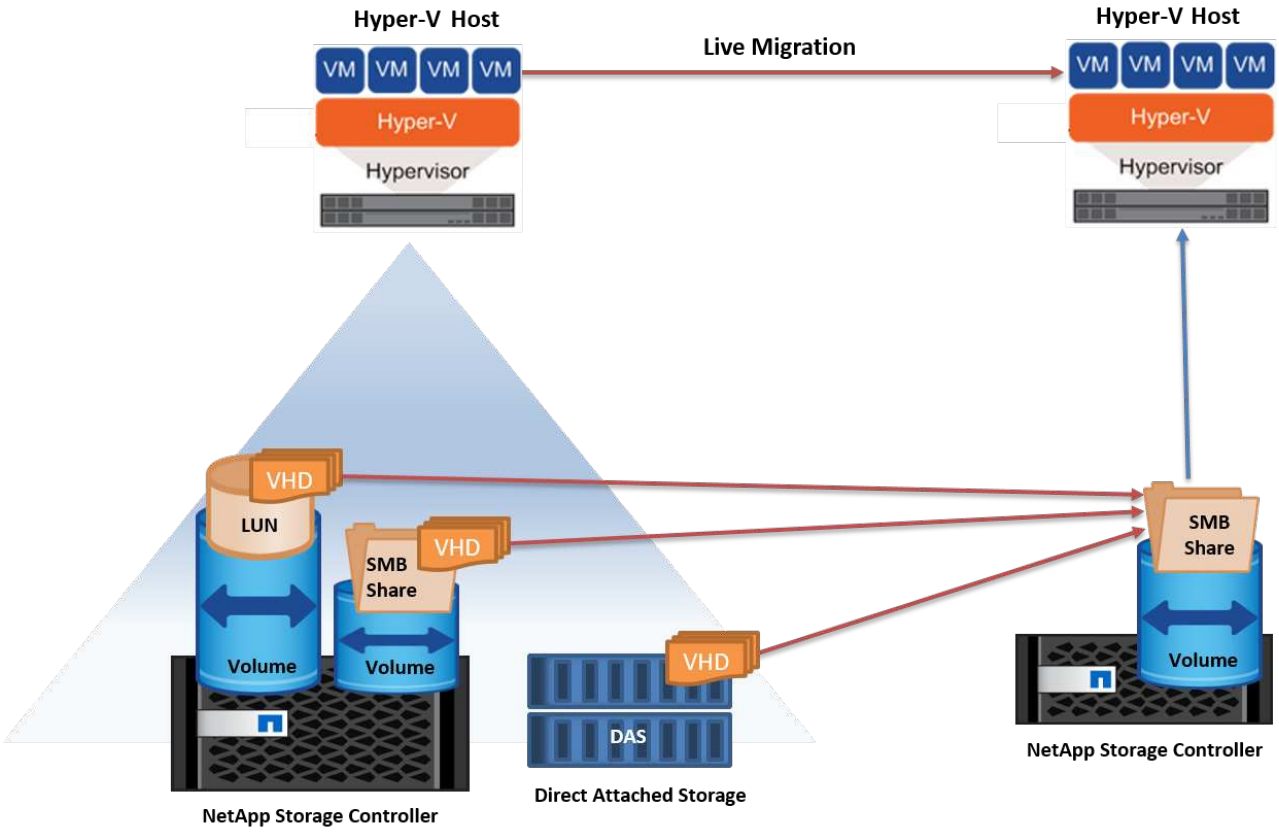
Live Migration all'esterno di un ambiente in cluster

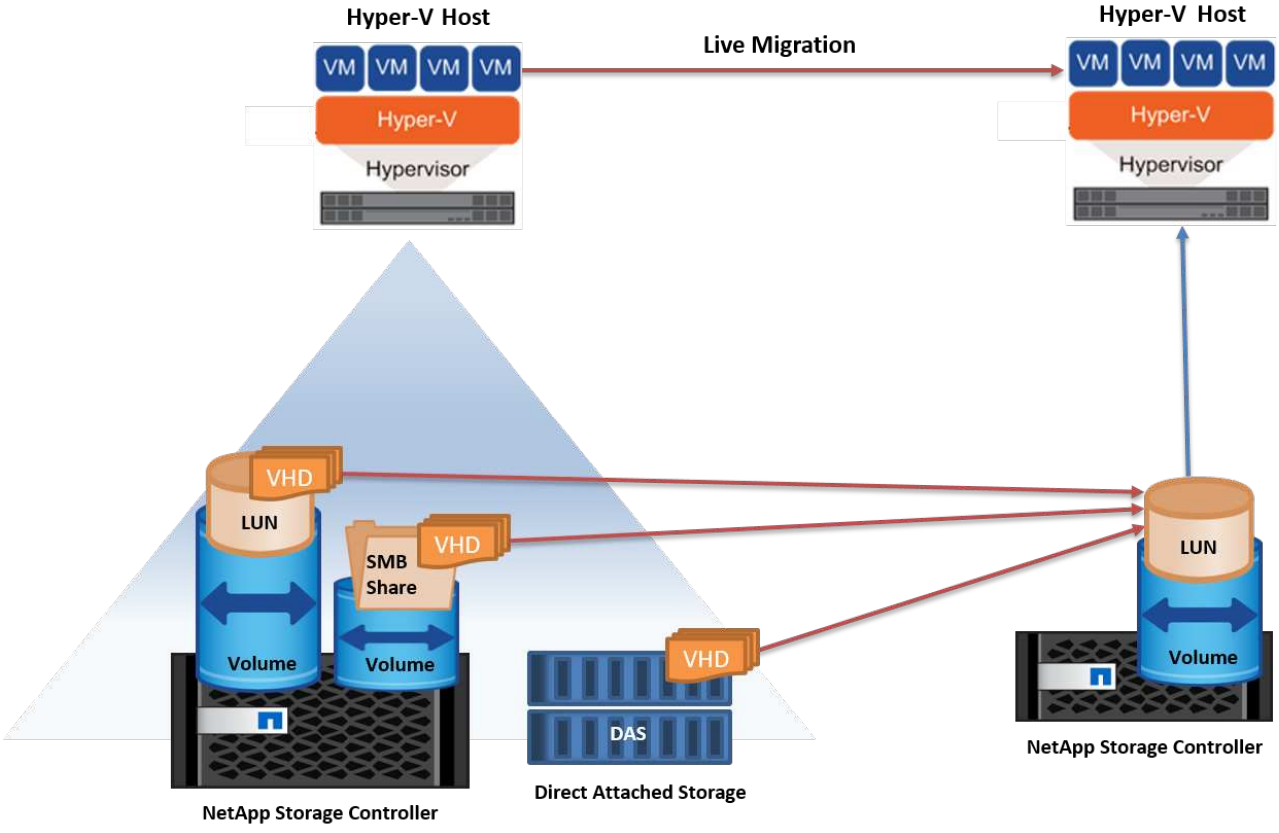
Puoi eseguire la migrazione live di una macchina virtuale tra due server Hyper-V indipendenti e non in cluster. Questo processo può utilizzare la migrazione in tempo reale senza elementi condivisi o condivisi.

- In una migrazione live condivisa, la macchina virtuale viene memorizzata in una condivisione SMB. Pertanto, quando si effettua la migrazione live di una macchina virtuale, lo storage della macchina virtuale rimane sulla condivisione SMB centrale per l'accesso istantaneo da parte dell'altro nodo, come illustrato di seguito.



- Nella migrazione live senza elementi condivisi, ogni server Hyper-V ha il proprio storage locale (può essere una condivisione SMB, un LUN o un DAS) e lo storage della macchina virtuale è locale al proprio server Hyper-V. Quando una VM viene migrata in tempo reale, viene eseguito il mirroring dello spazio di archiviazione della VM sul server di destinazione sulla rete client, quindi viene eseguita la migrazione della VM. La macchina virtuale memorizzata in DAS, un LUN o una condivisione SMB/CIFS può essere spostata in una condivisione SMB/CIFS sull'altro server Hyper-V, come illustrato nella figura seguente. Può anche essere spostata in un LUN, come mostrato nella seconda figura.





Ulteriori letture

Per informazioni sull'implementazione della migrazione live al di fuori di un ambiente in cluster, vedere

"Appendice D: Implementazione di Hyper-V Live Migration al di fuori di un ambiente in cluster".

Migrazione live dello storage Hyper-V.

Durante il ciclo di vita di una macchina virtuale, potrebbe essere necessario spostare lo storage della macchina virtuale (VHD/VHDX) su una diversa condivisione LUN o SMB. Ciò potrebbe essere necessario se lo spazio del LUN o della condivisione attuale sta per esaurirsi o sta producendo prestazioni inferiori al previsto.

Il LUN o la condivisione che attualmente ospita la macchina virtuale possono esaurire lo spazio, essere riutilizzati o fornire prestazioni ridotte. In tali circostanze, è possibile spostare la macchina virtuale senza tempi di inattività su un'altra LUN o condivisione su un volume, aggregato o cluster diverso. Questo processo è più rapido se il sistema storage dispone di funzionalità di offload delle copie. I sistemi di storage NetApp sono abilitati all'offload delle copie per impostazione predefinita per gli ambienti CIFS e SAN.

La funzionalità ODX esegue copie di file completi o di file secondari tra due directory che risiedono su server remoti. Una copia viene creata copiando i dati tra i server (o lo stesso server se entrambi i file di origine e di destinazione si trovano sullo stesso server). La copia viene creata senza che il client legga i dati dall'origine o scriva nella destinazione. Questo processo riduce l'utilizzo di processore e memoria per il client o il server e riduce al minimo la larghezza di banda i/o della rete. La copia è più veloce se è all'interno dello stesso volume. Se la copia è tra i volumi, potrebbe non esserci un aumento significativo delle prestazioni rispetto alle copie basate su host. Prima di procedere con un'operazione di copia sull'host, verificare che le impostazioni di offload delle copie siano configurate sul sistema di storage.

Quando la migrazione live dello storage delle macchine virtuali viene avviata da un host, l'origine e la destinazione vengono identificate e l'attività di copia viene scaricata nel sistema storage. Poiché l'attività viene eseguita dal sistema di archiviazione, l'utilizzo della CPU, della memoria o della rete host è trascurabile.

Gli storage controller NetApp supportano i seguenti scenari ODX:

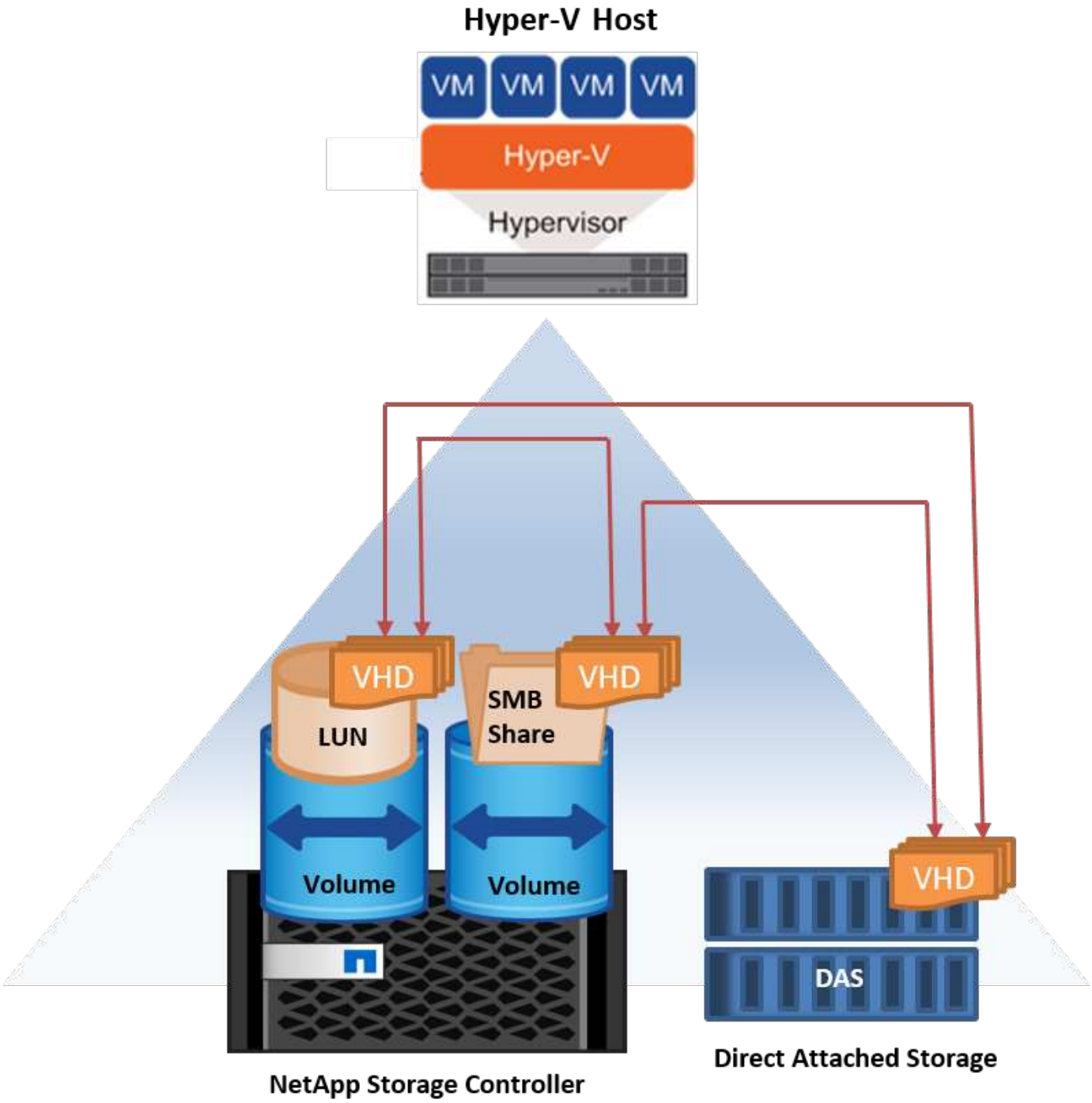
- **IntraSVM.** i dati sono di proprietà della stessa SVM:
- **Intravolume, intranode.** i file o LUN di origine e di destinazione risiedono nello stesso volume. La copia viene eseguita con la tecnologia file FlexClone che offre ulteriori vantaggi in termini di prestazioni delle copie remote.
- **Intervolume, intranode.** i file o LUN di origine e di destinazione si trovano su volumi diversi che si trovano sullo stesso nodo.
- **Intervolume, internodi.** i file o LUN di origine e di destinazione si trovano su volumi diversi che si trovano su nodi diversi.
- **InterSVM.** i dati sono di proprietà di diverse SVM.
- **Intervolume, intranode.** i file o LUN di origine e di destinazione si trovano su volumi diversi che si trovano sullo stesso nodo.
- **Intervolume, internodi.** i file o LUN di origine e di destinazione si trovano su volumi diversi che si trovano su nodi diversi.
- **Intercluster.** a partire da ONTAP 9,0, ODX è supportato anche per i trasferimenti di LUN intercluster in ambienti SAN. Intercluster ODX è supportato solo dai protocolli SAN, non da SMB.

Al termine della migrazione, è necessario riconfigurare i criteri di backup e replica in modo da riflettere il nuovo volume che contiene le VM. Non è possibile utilizzare i backup precedenti eseguiti.

Lo storage delle macchine virtuali (VHD/VHDX) può essere migrato tra i seguenti tipi di storage:

- DAS e la condivisione SMB

- DAS e LUN
- Una condivisione SMB e un LUN
- Tra LUN
- Tra condivisioni SMB

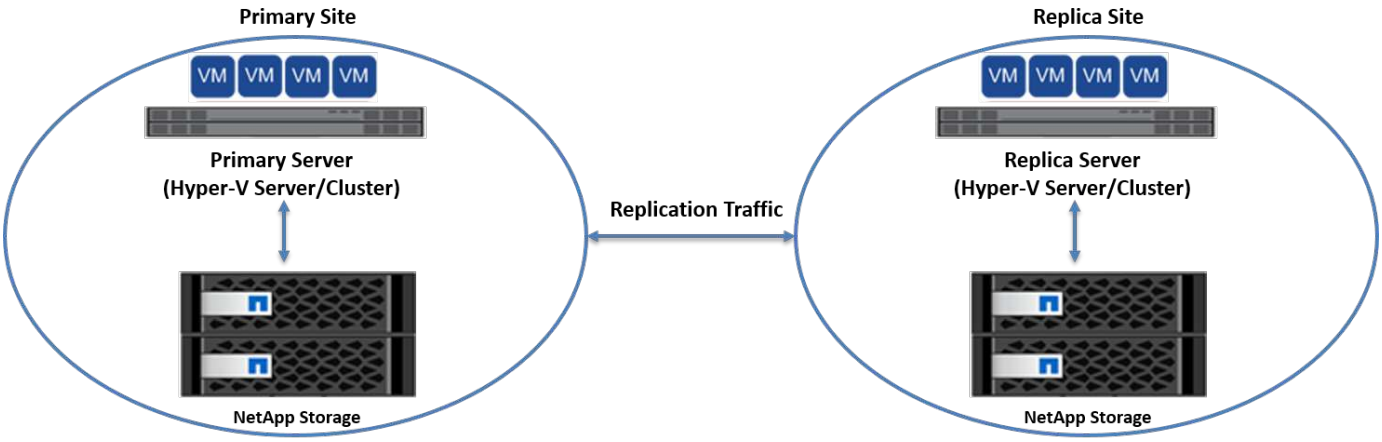


Ulteriori letture

Per informazioni sulla distribuzione della migrazione attiva dello storage, vedere "Appendice e: Implementare Hyper-V Storage Live Migration".

Replica Hyper-V: Disaster recovery per macchine virtuali

Replica di Hyper-V replica le macchine virtuali Hyper-V da un sito primario a una replica delle macchine virtuali su un sito secondario, fornendo in modo asincrono il disaster recovery per le macchine virtuali. Il server Hyper-V nel sito primario che ospita le macchine virtuali è noto come server primario, mentre il server Hyper-V nel sito secondario che riceve le macchine virtuali replicate è noto come server di replica. Nella figura seguente viene mostrato uno scenario di esempio di replica Hyper-V. È possibile utilizzare Hyper-V Replica per macchine virtuali tra server Hyper-V che fanno parte di un cluster di failover o tra server Hyper-V indipendenti che non fanno parte di un cluster.



Replica

Dopo aver abilitato la replica Hyper-V per una macchina virtuale sul server primario, la replica iniziale crea una macchina virtuale identica sul server di replica. Dopo la replica iniziale, Hyper-V Replica mantiene un file di registro per i VHD della VM. Il file di registro viene riprodotto in ordine inverso al VHD di replica secondo la frequenza di replica. Questo registro e l'utilizzo dell'ordine inverso garantiscono che le ultime modifiche vengano memorizzate e replicate in modo asincrono. Se la replica non avviene in linea con la frequenza prevista, viene emesso un avviso.

Replica estesa

Hyper-V Replica supporta la replica estesa in cui è possibile configurare un server di replica secondario per il disaster recovery. È possibile configurare un server di replica secondario affinché il server di replica riceva le modifiche sulle VM di replica. In uno scenario di replica estesa, le modifiche apportate alle macchine virtuali primarie sul server primario vengono replicate sul server di replica. Le modifiche vengono quindi replicate nel server di replica esteso. È possibile eseguire il failover delle macchine virtuali sul server di replica esteso solo quando i server primario e di replica si arrestano.

Failover

Il failover non è automatico; il processo deve essere attivato manualmente. Esistono tre tipi di failover:

- **Test failover.** questo tipo viene utilizzato per verificare che una VM di replica possa avviarsi correttamente sul server di replica e venga avviata sulla VM di replica. Questo processo crea una macchina virtuale di prova duplicata durante il failover e non influisce sulla normale replica di produzione.
- **Failover pianificato.** questo tipo viene utilizzato per eseguire il failover delle macchine virtuali durante tempi di inattività pianificati o interruzioni previste. Questo processo viene avviato sulla macchina virtuale primaria, che deve essere disattivata sul server primario prima di eseguire un failover pianificato. Dopo il failover della macchina, Hyper-V Replica avvia la VM di replica sul server di replica.
- **Failover non pianificato.** questo tipo viene utilizzato quando si verificano interruzioni impreviste. Questo

processo viene avviato sulla macchina virtuale di replica e deve essere utilizzato solo in caso di guasto della macchina principale.

Recovery (recupero)

Quando si configura la replica per una VM, è possibile specificare il numero di punti di ripristino. I punti di ripristino rappresentano i punti nel tempo da cui è possibile ripristinare i dati da un computer replicato.

Ulteriori letture

- Per informazioni sulla distribuzione di replica Hyper-V all'esterno di un ambiente cluster, vedere la sezione ""Implementazione di replica Hyper-V all'esterno di un ambiente cluster"."
- Per informazioni sulla distribuzione di replica Hyper-V in un ambiente cluster, vedere la sezione ""Implementare la replica Hyper-V in un ambiente cluster"."

Efficienza dello storage

ONTAP offre un'efficienza dello storage leader del settore per ambienti virtualizzati, incluso Microsoft Hyper-V. NetApp offre anche programmi di garanzia di efficienza dello storage.

Deduplica NetApp

La deduplica NetApp rimuove i blocchi duplicati a livello di volume di storage, archiviando una sola copia fisica, indipendentemente dal numero di copie logiche presenti. La deduplica si fa quindi illusione nel fatto che esistano numerose copie di quel blocco. La deduplica rimuove automaticamente i blocchi di dati duplicati su un livello di blocco di 4KB KB in un intero volume. Questo processo recupera lo storage per ottenere spazio e potenziali risparmi sulle performance, riducendo il numero di scritture fisiche su disco. La deduplica può garantire un risparmio di spazio superiore al 70% negli ambienti Hyper-V.

Thin provisioning

Il thin provisioning è un modo efficiente per il provisioning dello storage perché lo storage non è preallocato in anticipo. In altre parole, quando un volume o LUN viene creato utilizzando il thin provisioning, lo spazio nel sistema di storage non viene utilizzato. Lo spazio rimane inutilizzato fino a quando i dati non vengono scritti sul LUN o sul volume e viene utilizzato solo lo spazio necessario per memorizzare i dati. NetApp consiglia di attivare il thin provisioning sul volume e di disattivare la prenotazione LUN.

Qualità del servizio

La qualità del servizio di storage in NetApp ONTAP consente di raggruppare gli oggetti di storage e di impostare limiti di throughput per il gruppo. È possibile utilizzare la qualità del servizio di storage per limitare il throughput ai carichi di lavoro e monitorare le performance del carico di lavoro. Grazie a questa possibilità, gli amministratori dello storage possono separare i workload in base all'organizzazione, all'applicazione, alla business unit o agli ambienti di produzione o sviluppo.

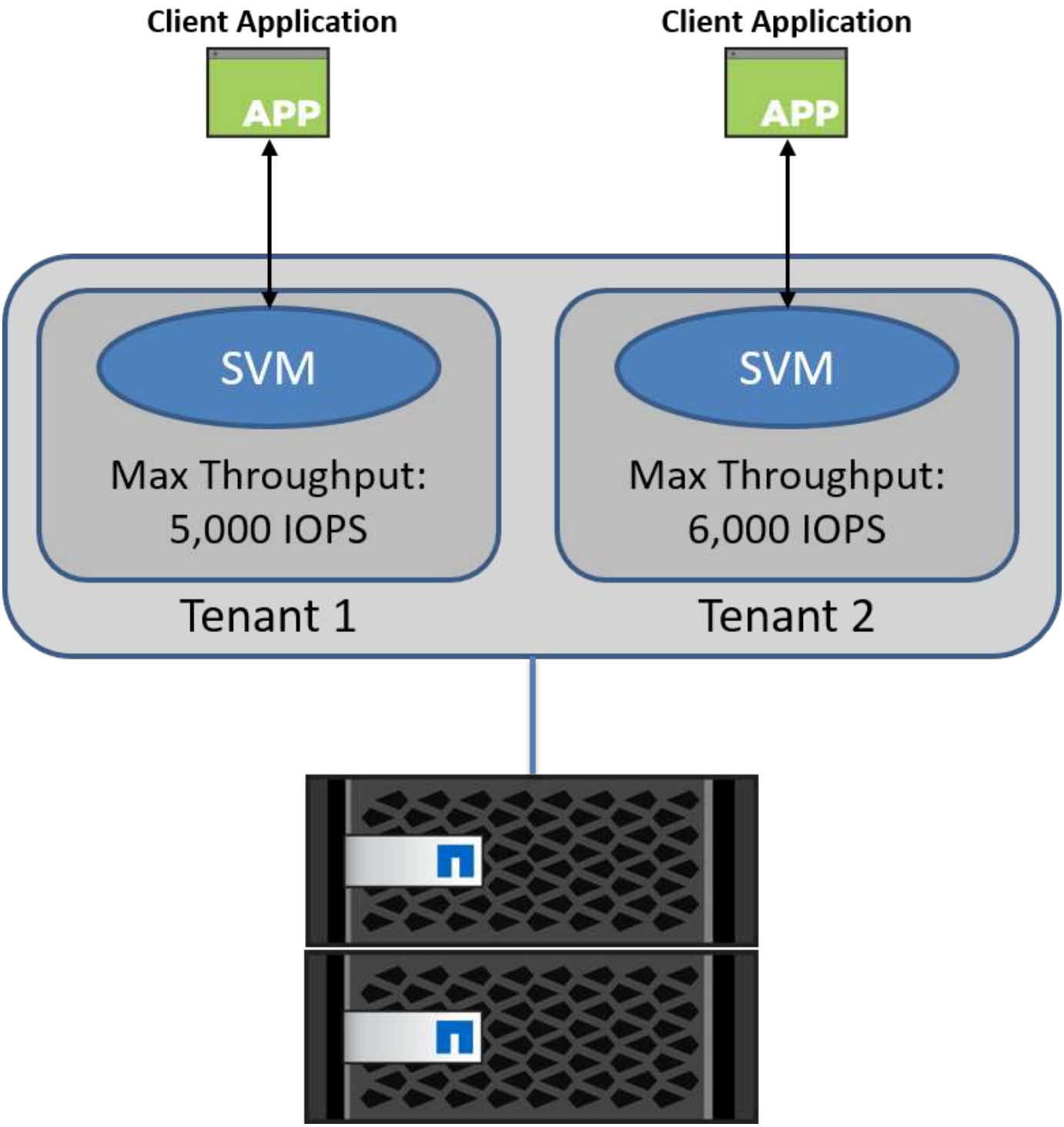
Negli ambienti aziendali, la qualità del servizio di storage aiuta a ottenere quanto segue:

- Impedisce che i carichi di lavoro degli utenti influiscano l'uno sull'altro.
- Protegge le applicazioni critiche con tempi di risposta specifici che è necessario soddisfare in ambienti IT-as-a-service (ITaaS).
- Impedisce che i tenant si influenzino l'uno con l'altro.

- Evita il peggioramento delle performance con l'aggiunta di ogni nuovo tenant.

La qualità del servizio consente di limitare il quantitativo di i/o inviato a una SVM, a un volume flessibile, a una LUN o a un file. L'i/o può essere limitato dal numero di operazioni o dal throughput raw.

La figura seguente illustra l'SVM con una propria policy QoS che applica un limite di throughput massimo.



Per configurare una SVM con la propria policy QoS e monitorare il gruppo di policy, esegui i seguenti comandi sul tuo cluster ONTAP:

```
# create a new policy group pg1 with a maximum throughput of 5,000 IOPS
cluster::> qos policy-group create pg1 -vserver vs1 -max-throughput
5000iops
```

```
# create a new policy group pg2 without a maximum throughput
cluster::> qos policy-group create pg2 -vserver vs2
```

```
# monitor policy group performance
cluster::> qos statistics performance show
```

```
# monitor workload performance
cluster::> qos statistics workload performance show
```

Sicurezza

ONTAP fornisce un sistema di storage sicuro per il sistema operativo Windows.

Windows Defender Antivirus

Windows Defender è un software antimalware installato e attivato in Windows Server per impostazione predefinita. Questo software protegge attivamente Windows Server da malware noti e può aggiornare regolarmente le definizioni antimalware tramite Windows Update. I LUN di NetApp e le condivisioni SMB possono essere sottoposti a scansione utilizzando Windows Defender.

Ulteriori letture

Per ulteriori informazioni, consultare la "Panoramica di Windows Defender".

BitLocker

La crittografia dell'unità BitLocker è una funzione di protezione dei dati continua da Windows Server 2012. Questa crittografia protegge dischi fisici, LUN e CSV.

Best practice

Prima di attivare BitLocker, il file CSV deve essere impostato sulla modalità di manutenzione. Pertanto, NetApp consiglia di prendere decisioni relative alla protezione basata su BitLocker prima di creare le macchine virtuali sul CSV per evitare tempi di inattività.

Distribuire il server Nano

Informazioni sulla distribuzione di Microsoft Windows Nano Server.

Implementazione

Per distribuire un Nano Server come host Hyper-V, attenersi alla seguente procedura:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Copiare la cartella NanoServerImageGenerator dalla cartella \NanoServer nell'ISO di Windows Server sul disco rigido locale.
3. Per creare un Nano Server VHD/VHDX, attenersi alla seguente procedura:
   a. Avviare Windows PowerShell come amministratore, accedere alla cartella NanoServerImageGenerator copiata sul disco rigido locale ed eseguire il seguente cmdlet:

```
Set-ExecutionPolicy RemoteSigned
Import-Module .\NanoServerImageGenerator -Verbose
```

   b. Creare un VHD per Nano Server come host Hyper-V eseguendo il seguente cmdlet PowerShell. Questo comando richiede una password di amministratore per il nuovo VHD.

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath <"input the path to the root of the contents of Windows
Server 2016 ISO"> -TargetPath <"input the path, including the
filename and extension where the resulting VHD/VHDX will be created">
-ComputerName <"input the name of the nano server computer you are
about to create"> -Compute
.. Nel seguente esempio viene creato un Nano Server VHD con la
funzione host Hyper-V con clustering di failover abilitato. Questo
esempio crea un Nano Server VHD da un ISO montato in f:\. Il VHD
appena creato viene inserito in una cartella denominata NanoServer
nella cartella da cui viene eseguito il cmdlet. Il nome del computer
è NanoServer e il VHD risultante contiene l'edizione standard di
Windows Server.
```

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath f:\ -TargetPath .\NanoServer.vhd -ComputerName NanoServer
-Compute -Clustering
.. Con il cmdlet New-NanoServerImage, configurare i parametri che
impostano l'indirizzo IP, la subnet mask, il gateway predefinito, il
server DNS, il nome del dominio, e così via.
```

4. Utilizzare il VHD in una macchina virtuale o in un host fisico per implementare Nano Server come host Hyper-V:
   a. Per l'implementazione su una macchina virtuale, creare una nuova macchina virtuale in Hyper-V Manager e utilizzare il VHD creato al passaggio 3.
   b. Per la distribuzione su un host fisico, copiare il VHD sul computer fisico e configurarlo per l'avvio da questo nuovo VHD. Per prima cosa, montare il VHD, eseguire bcdboot e:\Windows (dove il VHD è

montato sotto e:\), smontare il VHD, riavviare il computer fisico e avviare il Nano Server.

5. Unire il Nano Server a un dominio (opzionale):
   a. Accedere a un qualsiasi computer del dominio e creare un BLOB di dati eseguendo il seguente cmdlet PowerShell:

```
$domain = "<input the domain to which the Nano Server is to be
joined>"
$nanoserver = "<input name of the Nano Server>"
```

```
 djoin.exe /provision /domain $domain /machine $nanoserver /savefile
C:\temp\odjblob /reuse
.. Copiare il file odjblob nel Nano Server eseguendo i seguenti
cmdlet PowerShell su un computer remoto:
```

```
$nanoserver = "<input name of the Nano Server>"
$nanouname = ""<input username of the Nano Server>"
$nanopwd = ""<input password of the Nano Server>"
```

```
$filePath = 'c:\temp\odjblob'
$fileContents = Get-Content -Path $filePath -Encoding Unicode
```

```
$securenanopwd = ConvertTo-SecureString -AsPlainText -Force $nanopwd
$nanosecurecred = new-object management.automation.pscredential
$nanouname, $securenanopwd
```

```
Invoke-Command -VMName $nanoserver -Credential $nanosecurecred
-ArgumentList @($filePath,$fileContents) -ScriptBlock \{
    param($filePath,$data)
    New-Item -ItemType directory -Path c:\temp
    Set-Content -Path $filePath -Value $data -Encoding Unicode
    cd C:\temp
    djoin /requestodj /loadfile c:\temp\odjblob /windowspath
c:\windows /localos
}
```

   b. Riavviare Nano Server.

Connettersi a Nano Server

Per connettersi a Nano Server in remoto utilizzando PowerShell, attenersi alla seguente procedura:

1. Aggiungere Nano Server come host attendibile sul computer remoto eseguendo il seguente cmdlet sul server remoto:

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts "<input IP Address of the
Nano Server>"
. Se l'ambiente è sicuro e si desidera impostare tutti gli host da
aggiungere come host attendibili su un server, eseguire il comando
seguente:
```

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts *
. Avviare la sessione remota eseguendo il seguente cmdlet sul server
remoto. Fornire la password per Nano Server quando richiesto.
```

```
Enter-PSSession -ComputerName "<input IP Address of the Nano Server>"
-Credential ~\Administrator
```

Per connettersi a Nano Server in modalità remota utilizzando gli strumenti di gestione GUI da un Windows Server remoto, completare i seguenti comandi:

1. Accedere a Windows Server come membro del gruppo di amministratori.
2. Avviare Server Manager**.**
3. Per gestire un Nano Server in remoto da Server Manager, fare clic con il pulsante destro del mouse su tutti i server, fare clic su Aggiungi server, fornire le informazioni del Nano Server e aggiungerle. A questo punto è possibile visualizzare il Nano Server nell'elenco dei server. Selezionare il Nano Server, fare clic con il pulsante destro del mouse e iniziare a gestirlo con le varie opzioni fornite.
4. Per gestire i servizi in remoto su un Nano Server, attenersi alla seguente procedura:
   a. Aprire servizi dalla sezione Strumenti di Server Manager.
   b. Fare clic con il pulsante destro del mouse su servizi (locale).
   c. Fare clic su Connetti al server.
   d. Fornire i dettagli di Nano Server per visualizzare e gestire i servizi su Nano Server.
5. Se il ruolo Hyper-V è abilitato su Nano Server, completare i seguenti passaggi per gestirlo in remoto da Hyper-V Manager:
   a. Aprire Hyper-V Manager dalla sezione Strumenti di Server Manager.
   b. Fare clic con il pulsante destro del mouse su Hyper-V Manager.
   c. Fare clic su Connetti al server e fornire i dettagli del Nano Server. Ora Nano Server può essere gestito come server Hyper-V per creare e gestire macchine virtuali.
6. Se il ruolo di clustering di failover è abilitato su Nano Server, completare i seguenti passaggi per gestirlo in remoto dal failover cluster manager:

a. Aprire failover Cluster Manager dalla sezione Strumenti di Server Manager.
b. Eseguire operazioni relative al clustering con Nano Server.

Implementa il cluster Hyper-V.

La presente appendice descrive l'implementazione di un cluster Hyper-V.

Prerequisiti

- Almeno due server Hyper-V sono connessi tra loro.
- Su ciascun server Hyper-V è configurato almeno uno switch virtuale.
- La funzione cluster di failover è abilitata su ogni server Hyper-V.
- Le condivisioni SMB o i CSV vengono utilizzati come storage condiviso per memorizzare macchine virtuali e relativi dischi per il clustering Hyper-V.
- Lo storage non deve essere condiviso tra cluster diversi. È necessaria una sola condivisione CSV/CIFS per cluster.
- Se la condivisione SMB viene utilizzata come storage condiviso, è necessario configurare le autorizzazioni sulla condivisione SMB in modo da consentire l'accesso agli account computer di tutti i server Hyper-V nel cluster.

Implementazione

1. Accedere a uno dei server Windows Hyper-V come membro del gruppo di amministratori.
2. Avviare Server Manager**.**
3. Nella sezione Strumenti, fare clic su failover Cluster Manager.
4. Fare clic sul menu Create Cluster from Actions (Crea cluster da azioni).
5. Fornire dettagli sul server Hyper-V che fa parte di questo cluster.
6. Convalidare la configurazione del cluster. Selezionare Sì quando viene richiesta la convalida della configurazione del cluster e selezionare i test necessari per verificare se i server Hyper-V superano i prerequisiti per far parte del cluster.
7. Una volta completata la convalida, viene avviata la procedura guidata Crea cluster. Nella procedura guidata, specificare il nome del cluster e l'indirizzo IP del cluster per il nuovo cluster. Viene quindi creato un nuovo cluster di failover per il server Hyper-V.
8. Fare clic sul nuovo cluster creato in failover Cluster Manager e gestirlo.
9. Definire lo storage condiviso da utilizzare per il cluster. Può trattarsi di una condivisione SMB o di un CSV.
10. L'utilizzo di una condivisione SMB come storage condiviso non richiede passaggi particolari.
    - Configurazione di una CIFS share su uno storage controller NetApp. A tale scopo, vedere la sezione ""Provisioning negli ambienti SMB"".
11. Per utilizzare un file CSV come archivio condiviso, attenersi alla seguente procedura:
    a. Configurare le LUN su uno storage controller NetApp. A tale scopo, vedere la sezione "Provisioning in ambienti SAN".
    b. Assicurarsi che tutti i server Hyper-V nel cluster di failover siano in grado di vedere i LUN NetApp. A tale scopo, per tutti i server Hyper-V che fanno parte del cluster di failover, assicurarsi che i relativi iniziatori siano aggiunti al gruppo iniziatore sullo storage NetApp. Verificare inoltre che i LUN siano stati rilevati e che MPIO sia attivato.

c. Su uno qualsiasi dei server Hyper-V nel cluster, completare i seguenti passaggi:

   i. Portare il LUN online, inizializzare il disco, creare un nuovo volume semplice e formattarlo utilizzando NTFS o refs.

   ii. In failover Cluster Manager, espandere il cluster, espandere Storage, fare clic con il pulsante destro del mouse su dischi, quindi fare clic su Add Disks (Aggiungi dischi). In questo modo si apre la procedura guidata Aggiungi dischi a un cluster che mostra il LUN come disco. Fare clic su OK per aggiungere il LUN come disco.

   iii. Ora il LUN è chiamato "disco in cluster" e viene mostrato come spazio di archiviazione disponibile in dischi.

d. Fare clic con il pulsante destro del mouse sul LUN ('disco in cluster') e fare clic su Aggiungi a volumi condivisi cluster. Ora il LUN viene visualizzato come CSV.

e. Il CSV è simultaneamente visibile e accessibile da tutti i server Hyper-V del cluster di failover nella sua posizione locale C:\ClusterStorage\.

12. Creare una macchina virtuale altamente disponibile:

    a. In failover Cluster Manager, selezionare ed espandere il cluster creato in precedenza.

    b. Fare clic su ruoli, quindi su macchine virtuali in azioni. Fare clic su Nuova macchina virtuale.

    c. Selezionare il nodo dal cluster in cui deve risiedere la VM.

    d. Nella procedura guidata per la creazione della macchina virtuale, fornire lo storage condiviso (SMB share o CSV) come percorso di archiviazione della macchina virtuale e dei relativi dischi.

    e. Utilizzare Hyper-V Manager per impostare lo storage condiviso (SMB share o CSV) come percorso predefinito per l'archiviazione della VM e dei relativi dischi per un server Hyper-V.

13. Verifica del failover pianificato. Sposta le macchine virtuali su un altro nodo utilizzando migrazione live, migrazione rapida o migrazione dello storage (spostamento). Revisione "Migrazione live in un ambiente cluster" per ulteriori dettagli.

14. Verifica del failover non pianificato. Arrestare il servizio cluster sul server proprietario della VM.

Distribuire Hyper-V Live Migration in un ambiente in cluster

Questa appendice descrive la distribuzione della migrazione live in un ambiente in cluster.

Prerequisiti

Per distribuire la migrazione in tempo reale, è necessario che i server Hyper-V siano configurati in un cluster di failover con storage condiviso. Revisione "Distribuire il cluster Hyper-V." per ulteriori dettagli.

Implementazione

Per utilizzare la migrazione live in un ambiente in cluster, attenersi alla seguente procedura:

1. In failover Cluster Manager, selezionare ed espandere il cluster. Se il cluster non è visibile, fare clic su failover Cluster Manager, fare clic su Connetti al cluster e fornire il nome del cluster.
2. Fare clic su ruoli, che elenca tutte le VM disponibili in un cluster.
3. Fare clic con il pulsante destro del mouse sulla macchina virtuale e fare clic su Sposta. In questo modo, sono disponibili tre opzioni:
   - **Migrazione live.** è possibile selezionare un nodo manualmente o consentire al cluster di selezionare il

nodo migliore. Durante la migrazione live, il cluster copia la memoria utilizzata dalla macchina virtuale dal nodo corrente a un altro nodo. Pertanto, quando la macchina virtuale viene migrata su un altro nodo, la memoria e le informazioni di stato necessarie alla macchina virtuale sono già disponibili per la macchina virtuale. Questo metodo di migrazione è quasi istantaneo, ma è possibile eseguire la migrazione in tempo reale di una sola macchina virtuale alla volta.

- **Migrazione rapida.** è possibile selezionare un nodo manualmente o consentire al cluster di selezionare il nodo migliore. Durante una migrazione rapida, il cluster copia la memoria utilizzata da una macchina virtuale in un disco nello storage. Pertanto, quando la macchina virtuale viene migrata su un altro nodo, le informazioni di memoria e di stato necessarie alla macchina virtuale possono essere lette rapidamente dal disco dall'altro nodo. Con una migrazione rapida, è possibile migrare più macchine virtuali contemporaneamente.
- **Migrazione dell'archiviazione di macchine virtuali.** questo metodo utilizza la procedura guidata Sposta archivio di macchine virtuali. Questa procedura guidata consente di selezionare il disco della macchina virtuale e altri file da spostare in un'altra posizione, ad esempio una condivisione CSV o SMB.

Implementazione di Hyper-V Live Migration all'esterno di un ambiente in cluster

Questa sezione descrive l'implementazione della migrazione live di Hyper-V all'esterno di un ambiente in cluster.

Prerequisiti

- Server Hyper-V standalone con storage indipendente o storage SMB condiviso.
- Il ruolo Hyper-V installato sui server di origine e di destinazione.
- Entrambi i server Hyper-V appartengono allo stesso dominio o a domini che si fidano l'uno dell'altro.

Implementazione

Per eseguire la migrazione live in un ambiente non in cluster, configurare i server Hyper-V di origine e di destinazione in modo che possano inviare e ricevere operazioni di migrazione live. Su entrambi i server Hyper-V, completare la seguente procedura:

1. Aprire Hyper-V Manager dalla sezione Strumenti di Server Manager.
2. In azioni, fare clic su Impostazioni Hyper-V.
3. Fare clic su Live Migration e selezionare Enable Incoming and Outgoing Live Migrations (Abilita migrazioni Live in entrata e in uscita).
4. Scegliere se consentire il traffico di migrazione live su qualsiasi rete disponibile o solo su reti specifiche.
5. In alternativa, è possibile configurare il protocollo di autenticazione e le opzioni relative alle prestazioni dalla sezione Advanced (Avanzate) di Live Migrations (migrazioni in tempo reale).
6. Se si utilizza CredSSP come protocollo di autenticazione, assicurarsi di accedere al server Hyper-V di origine dal server Hyper-V di destinazione prima di spostare la macchina virtuale.
7. Se Kerberos viene utilizzato come protocollo di autenticazione, configurare la delega vincolata. Questa operazione richiede l'accesso al controller di dominio Active Directory. Per configurare la delega, attenersi alla seguente procedura:
   a. Accedere al controller di dominio Active Directory come amministratore.
   b. Avviare Server Manager.
   c. Nella sezione Strumenti, fare clic su utenti e computer di Active Directory.

d. Espandere il dominio e fare clic su computer.

e. Selezionare il server Hyper-V di origine dall'elenco, fare clic con il pulsante destro del mouse su di esso e scegliere Proprietà.

f. Nella scheda delega, selezionare considera attendibile il computer per la delega solo ai servizi specificati.

g. Selezionare utilizza solo Kerberos.

h. Fare clic su Aggiungi per aprire la procedura guidata Aggiungi servizi.

i. In Aggiungi servizi, fare clic su utenti e computer, che apre Seleziona utenti o computer**.**

j. Specificare il nome del server Hyper-V di destinazione e fare clic su OK.

- Per spostare lo storage delle macchine virtuali, selezionare CIFS.
- Per spostare le macchine virtuali, selezionare il servizio Microsoft Virtual System Migration.

k. Nella scheda delega, fare clic su OK.

l. Dalla cartella computer, selezionare il server Hyper-V di destinazione dall'elenco e ripetere il processo. In Seleziona utenti o computer, fornire il nome del server Hyper-V.

8. Spostare la macchina virtuale.

a. Aprire Hyper-V Manager.

b. Fare clic con il pulsante destro del mouse su una macchina virtuale, quindi fare clic su Sposta.

c. Scegliere Sposta macchina virtuale.

d. Specificare il server Hyper-V di destinazione per la macchina virtuale.

e. Scegliere le opzioni di spostamento. Per Shared Live Migration, scegliere Sposta solo la macchina virtuale. Per Shared Nothing Live Migration, scegli una delle altre due opzioni in base alle tue preferenze.

f. Fornire la posizione della macchina virtuale sul server Hyper-V di destinazione in base alle proprie preferenze.

g. Rivedere il riepilogo e fare clic su OK per spostare la VM.

Distribuire Hyper-V Storage Live Migration

Scoprite come configurare la migrazione live dello storage Hyper-V.

Prerequisiti

- È necessario disporre di un server Hyper-V standalone con storage indipendente (DAS o LUN) o storage SMB (locale o condiviso tra altri server Hyper-V).
- Il server Hyper-V deve essere configurato per la migrazione live. Esaminare la sezione relativa alla distribuzione in "Live Migration al di fuori di un ambiente cluster".

Implementazione

1. Aprire Hyper-V Manager.
2. Fare clic con il pulsante destro del mouse su una macchina virtuale, quindi fare clic su Sposta.
3. Selezionare Sposta memoria della macchina virtuale.
4. Selezionare le opzioni per spostare la memoria in base alle proprie preferenze.

5. Fornire la nuova posizione per gli elementi della VM.
6. Rivedere il riepilogo e fare clic su OK per spostare la memoria della VM.

Implementazione di replica Hyper-V all'esterno di un ambiente in cluster

In questa appendice viene descritta la distribuzione di replica Hyper-V all'esterno di un ambiente in cluster.

Prerequisiti

- Sono necessari server Hyper-V standalone ubicati in posizioni geografiche identiche o separate che servono come server primari e di replica.
- Se si utilizzano siti separati, è necessario configurare il firewall di ciascun sito per consentire la comunicazione tra i server primario e di replica.
- Il server di replica deve disporre di spazio sufficiente per archiviare i carichi di lavoro replicati.

Implementazione

1. Configurare il server di replica.
   a. Affinché le regole del firewall in entrata consentano il traffico di replica in entrata, eseguire il seguente cmdlet PowerShell:

```
 Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP Listener
(TCP-In)"
.. Aprire Hyper-V Manager dalla sezione Strumenti di Server Manager.
.. Fare clic su Impostazioni Hyper-V da azioni.
.. Fare clic su Replication Configuration (Configurazione replica) e
selezionare Enable this computer as a Replica Server (Abilita questo
computer come server di replica).
.. Nella sezione Authentication and Ports (autenticazione e porte),
selezionare il metodo e la porta di autenticazione.
.. Nella sezione autorizzazione e archiviazione, specificare la
posizione in cui archiviare le VM e i file replicati.
```

2. Abilitare la replica VM per le VM sul server primario. La replica delle macchine virtuali è abilitata in base alle macchine virtuali e non per l'intero server Hyper-V.
   a. In Hyper-V Manager, fare clic con il pulsante destro del mouse su una macchina virtuale e fare clic su Enable Replication (Abilita replica) per aprire la procedura guidata Enable Replication (Abilita replica).
   b. Fornire il nome del server di replica in cui la VM deve essere replicata.
   c. Fornire il tipo di autenticazione e la porta del server di replica configurata per ricevere il traffico di replica sul server di replica.
   d. Selezionare i VHD da replicare.
   e. Scegliere la frequenza (durata) in cui le modifiche vengono inviate al server di replica.
   f. Configurare i punti di ripristino per specificare il numero di punti di ripristino da mantenere sul server di replica.

g. Scegliere Initial Replication Method (metodo di replica iniziale) per specificare il metodo di trasferimento della copia iniziale dei dati VM al server di replica.

h. Rivedere il riepilogo e fare clic su fine.

i. Questo processo crea una replica VM sul server di replica.

Replica

1. Eseguire un failover di test per assicurarsi che la VM di replica funzioni correttamente sul server di replica. Il test crea una VM temporanea sul server di replica.
   a. Accedere al server di replica.
   b. In Hyper-V Manager, fare clic con il pulsante destro del mouse su una macchina virtuale di replica, fare clic su Replication (Replica) e su Test failover (Test failover).
   c. Scegliere il punto di ripristino da utilizzare.
   d. Questo processo crea una macchina virtuale con lo stesso nome aggiunto a -Test.
   e. Verificare la VM per accertarsi che tutto funzioni correttamente.
   f. Dopo il failover, la macchina virtuale di prova della replica viene eliminata se si seleziona Stop Test failover.
2. Eseguire un failover pianificato per replicare le ultime modifiche sulla macchina virtuale primaria nella macchina virtuale di replica.
   a. Accedere al server primario.
   b. Disattivare la macchina virtuale da sottoporre a failover.
   c. In Hyper-V Manager, fare clic con il pulsante destro del mouse sulla macchina virtuale disattivata, fare clic su Replica, quindi su failover pianificato.
   d. Fare clic su failover per trasferire le ultime modifiche VM al server di replica.
3. Eseguire un failover non pianificato in caso di guasto principale della VM.
   a. Accedere al server di replica.
   b. In Hyper-V Manager, fare clic con il pulsante destro del mouse su una VM di replica, fare clic su Replication (Replica) e su failover.
   c. Scegliere il punto di ripristino da utilizzare.
   d. Fare clic su failover per eseguire il failover della VM.

Distribuire la replica Hyper-V in un ambiente in cluster

Scopri come distribuire e configurare la replica di Hyper-V con il cluster di failover di Windows Server.

Prerequisiti

- Devi avere cluster Hyper-V ubicati nella stessa area geografica o in posizioni separate che fungono da cluster primari e di replica. Revisione "Distribuire il cluster Hyper-V." per ulteriori dettagli.
- Se si utilizzano siti separati, è necessario configurare il firewall in ciascun sito per consentire la comunicazione tra i cluster primario e di replica.
- Il cluster di replica deve disporre di spazio sufficiente per archiviare i workload replicati.

Implementazione

1. Attivare le regole firewall su tutti i nodi di un cluster. Eseguire il seguente cmdlet PowerShell con privilegi di amministratore su tutti i nodi sia nel cluster primario che di replica.

```
# For Kerberos authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP
Listener (TCP-In)"}}
```

```
# For Certificate authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica
HTTPS Listener (TCP-In)"}}
```

2. Configurare il cluster di replica.
   a. Configurare il broker replica Hyper-V con un nome NetBIOS e un indirizzo IP da utilizzare come punto di connessione al cluster utilizzato come cluster di replica.
      i. Aprire failover Cluster Manager.
      ii. Espandere il cluster, fare clic su ruoli e fare clic sul riquadro Configura ruolo dal riquadro azioni.
      iii. Selezionare Broker replica Hyper-V nella pagina Seleziona ruolo.
      iv. Fornire il nome NetBIOS e l'indirizzo IP da utilizzare come punto di connessione al cluster (punto di accesso client).
      v. Questo processo crea un ruolo di broker replica Hyper-V. Verificare che sia online correttamente.
   b. Configurare le impostazioni di replica.
      i. Fare clic con il pulsante destro del mouse sul broker di replica creato nei passaggi precedenti e fare clic su Impostazioni di replica.
      ii. Selezionare attiva questo cluster come server di replica.
      iii. Nella sezione Authentication and Ports (autenticazione e porte), selezionare il metodo e la porta di autenticazione.
      iv. Nella sezione autorizzazione e archiviazione, selezionare i server autorizzati a replicare le VM in questo cluster. Inoltre, specificare la posizione predefinita in cui sono memorizzate le VM replicate.

Replica

La replica è simile al processo descritto nella sezione "Replica al di fuori di un ambiente cluster".

Dove trovare ulteriori informazioni

Ulteriori risorse per Microsoft Windows e Hyper-V.

- Concetti di ONTAP
  https://docs.netapp.com/us-en/ontap/concepts/introducing-data-management-software-concept.html
- Best practice per LE SAN moderne

https://www.netapp.com/media/10680-tr4080.pdf

- Integrità e disponibilità dei dati degli array NetApp All-SAN con NetApp ASA
  https://www.netapp.com/pdf.html?item=/media/85671-tr-4968.pdf
- Documentazione SMB
  https://docs.netapp.com/us-en/ontap/smb-admin/index.html
- Guida introduttiva a Nano Server
  https://technet.microsoft.com/library/mt126167.aspx
- Novità di Hyper-V su Windows Server
  https://technet.microsoft.com/windows-server-docs/compute/hyper-v/what-s-new-in-hyper-v-on-windows

Microsoft SQL Server

Panoramica

ONTAP offre una soluzione per la sicurezza e le prestazioni di livello aziendale per i database Microsoft SQL Server e allo stesso tempo fornisce strumenti di prim'ordine per la gestione dell'ambiente.

Questa documentazione sostituisce il report tecnico precedentemente pubblicato *TR-4590: Best practice guide for Microsoft SQL Server with ONTAP*

NetApp presuppone che il lettore disponga delle seguenti conoscenze operative:

- ONTAP
- Architettura e amministrazione di Microsoft SQL Server
- NetApp SnapCenter come software di backup, che include:
  - Plug-in SnapCenter per Microsoft Windows
  - Plug-in di SnapCenter per SQL Server

L'ambito di questa sezione sulle Best practice è limitato alla progettazione tecnica, basata su principi e standard preferenziali che NetApp consiglia per l'infrastruttura di storage. L'implementazione end-to-end non rientra nell'ambito di applicazione.

Per informazioni sulla compatibilità tra i prodotti NetApp, vedere "Tool di matrice di interoperabilità NetApp (IMT)".

Workload di Microsoft SQL Server

Prima di distribuire SQL Server, è necessario comprendere i requisiti dei carichi di lavoro delle applicazioni supportate dalle istanze dei database di SQL Server. Ogni applicazione ha requisiti differenti in termini di capacità, performance e disponibilità, per cui ogni database dovrebbe essere progettato per supportare al meglio tali requisiti. Molte organizzazioni classificano i database in più Tier di gestione, utilizzando i requisiti delle applicazioni per definire gli SLA. I carichi di lavoro di SQL Server sono spesso classificati come descritto di seguito:

- OLTP, che è spesso i database più critici di un'organizzazione. In genere, questi database supportano le applicazioni rivolte ai clienti e sono considerati essenziali per le operazioni chiave dell'azienda. I database OLTP mission-critical e le applicazioni che questi supportano spesso hanno SLA che richiedono performance elevate, sono sensibili al peggioramento delle performance e richiedono la massima disponibilità. Potrebbero anche essere candidati per cluster di failover sempre attivi o gruppi di disponibilità sempre attivi. La combinazione di i/o di questi tipi di database è in genere caratterizzata da una percentuale compresa tra il 75% e il 90% di random Read e tra il 25% e il 10% di scrittura.
- Database DSS (Decision Support System), talvolta denominati data warehouse. Questi database sono mission-critical per molte organizzazioni che si affidano alle analytics per il loro business. Questi database sono sensibili all'utilizzo della CPU e alle operazioni di lettura dal disco quando vengono eseguite query. In molte organizzazioni, i database DSS sono i più critici durante la fine di mese, trimestre e anno Questo carico di lavoro ha in genere una combinazione di i/o di lettura quasi del 100% e il throughput io è spesso più importante degli IOPS.

Configurazione del database

Configurazione CPU

Le prestazioni di SQL Server dipendono in modo multiplo dalla CPU e dalla configurazione di base.

Hyper-threading

L'Hyper-threading si riferisce all'implementazione simultanea del multithreading (SMT), che migliora la parallelizzazione dei calcoli eseguiti su processori x86. SMT è disponibile sia sui processori Intel che AMD.

L'Hyper-threading produce CPU logiche che appaiono come CPU fisiche nel sistema operativo. SQL Server vede quindi le CPU aggiuntive e le utilizza come se vi fossero più core di quelli fisicamente presenti. Questo può migliorare notevolmente le prestazioni aumentando la parallelizzazione.

Si noti che ogni versione di SQL Server presenta dei limiti specifici sulla potenza di calcolo che può utilizzare. Per ulteriori informazioni, vedere "Limiti di capacità di calcolo per edizione di SQL Server".

Core e licenze

Esistono due opzioni per la licenza di SQL Server. Il primo è noto come modello server + licenza di accesso client (CAL); il secondo è il modello core per processore. Sebbene sia possibile accedere a tutte le funzioni del prodotto disponibili in SQL Server con la strategia server + CAL, esiste un limite hardware di 20 core CPU per socket. Anche se si dispone di SQL Server Enterprise Edition + CAL per un server con più di 20 core di CPU per socket, l'applicazione non può utilizzare tutti questi core alla volta in tale istanza.

L'immagine seguente mostra il messaggio di registro di SQL Server dopo l'avvio che indica l'imposizione del limite principale.

```
2017-01-11 07:16:30.71 Server      Microsoft SQL Server 2016
(RTM) - 13.0.1601.5 (X64)
     Apr 29 2016 23:23:58
     Copyright (c) Microsoft Corporation
     Enterprise Edition (64-bit) on Windows Server 2016
Datacenter 6.3 <X64> (Build 14393: )

2017-01-11 07:16:30.71 Server      UTC adjustment: -8:00
2017-01-11 07:16:30.71 Server      (c) Microsoft Corporation.
2017-01-11 07:16:30.71 Server      All rights reserved.
2017-01-11 07:16:30.71 Server      Server process ID is 10176.
2017-01-11 07:16:30.71 Server      System Manufacturer:
'FUJITSU', System Model: 'PRIMERGY RX2540 M1'.
2017-01-11 07:16:30.71 Server      Authentication mode is MIXED.
2017-01-11 07:16:30.71 Server      Logging SQL Server messages
in file 'C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG'.
2017-01-11 07:16:30.71 Server      The service account is 'SEA-
TM\FUJIA2R30$'. This is an informational message; no user action
is required.
2017-01-11 07:16:30.71 Server      Registry startup parameters:
     -d C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\master.mdf
     -e C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG
     -l C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\mastlog.ldf
     -T 3502
     -T 834
2017-01-11 07:16:30.71 Server      Command Line Startup
Parameters:
     -s "MSSQLSERVER"
2017-01-11 07:16:30.72 Server      SQL Server detected 2 sockets
with 18 cores per socket and 36 logical processors per socket,
72 total logical processors; using 40 logical processors based
on SQL Server licensing. This is an informational message; no
user action is required.
2017-01-11 07:16:30.72 Server      SQL Server is starting at
```

Pertanto, per utilizzare tutte le CPU, è necessario utilizzare la licenza core per processore. Per informazioni dettagliate sulle licenze di SQL Server, vedere "SQL Server 2022: La tua moderna piattaforma per i dati".

Affinità della CPU

È improbabile che sia necessario modificare le impostazioni predefinite di affinità del processore a meno che non si verifichino problemi di prestazioni, ma vale ancora la pena capire cosa sono e come funzionano.

SQL Server supporta l'affinità del processore mediante due opzioni:

- Maschera di affinità della CPU
- Maschera i/o di affinità

SQL Server utilizza tutte le CPU disponibili dal sistema operativo (se si sceglie la licenza core per processore). Inoltre, crea degli scheduler f0r per ogni CPU per utilizzare al meglio le risorse per qualsiasi carico di lavoro. Durante il multitasking, il sistema operativo o altre applicazioni sul server possono passare da un processore all'altro. SQL Server è un'applicazione che richiede molte risorse e in tal caso le prestazioni possono risentirne. Per ridurre al minimo l'impatto, è possibile configurare i processori in modo che tutto il carico di SQL Server venga indirizzato a un gruppo preselezionato di processori. Ciò si ottiene utilizzando la maschera di affinità della CPU.

L'opzione maschera i/o affinità associa l'i/o del disco di SQL Server a un sottoinsieme di CPU. Negli ambienti OLTP di SQL Server, questa estensione può migliorare significativamente le prestazioni dei thread di SQL

Server che emettono operazioni i/O.

Massimo grado di parallelismo (MAXDOP)

Per impostazione predefinita, SQL Server utilizza tutte le CPU disponibili durante l'esecuzione delle query, se si sceglie la licenza core per processore.

Sebbene sia utile per query di grandi dimensioni, può causare problemi di prestazioni e limitare la concorrenza. Un approccio migliore consiste nel limitare il parallelismo al numero di core fisici in un singolo socket CPU. Ad esempio, su un server con due socket CPU fisici con 12 core per socket, indipendentemente dall'Hyper-threading, `MAXDOP` deve essere impostato su 12. `MAXDOP` Non è possibile limitare o dettare quale CPU utilizzare. Limita invece il numero di CPU che possono essere utilizzate da una singola query batch.

NetApp consiglia per DSS, ad esempio data warehouse, iniziare con `MAXDOP` 50 ed esplorare la messa a punto su o giù, se necessario. Assicurarsi di misurare le query critiche nell'applicazione quando si apportano modifiche.

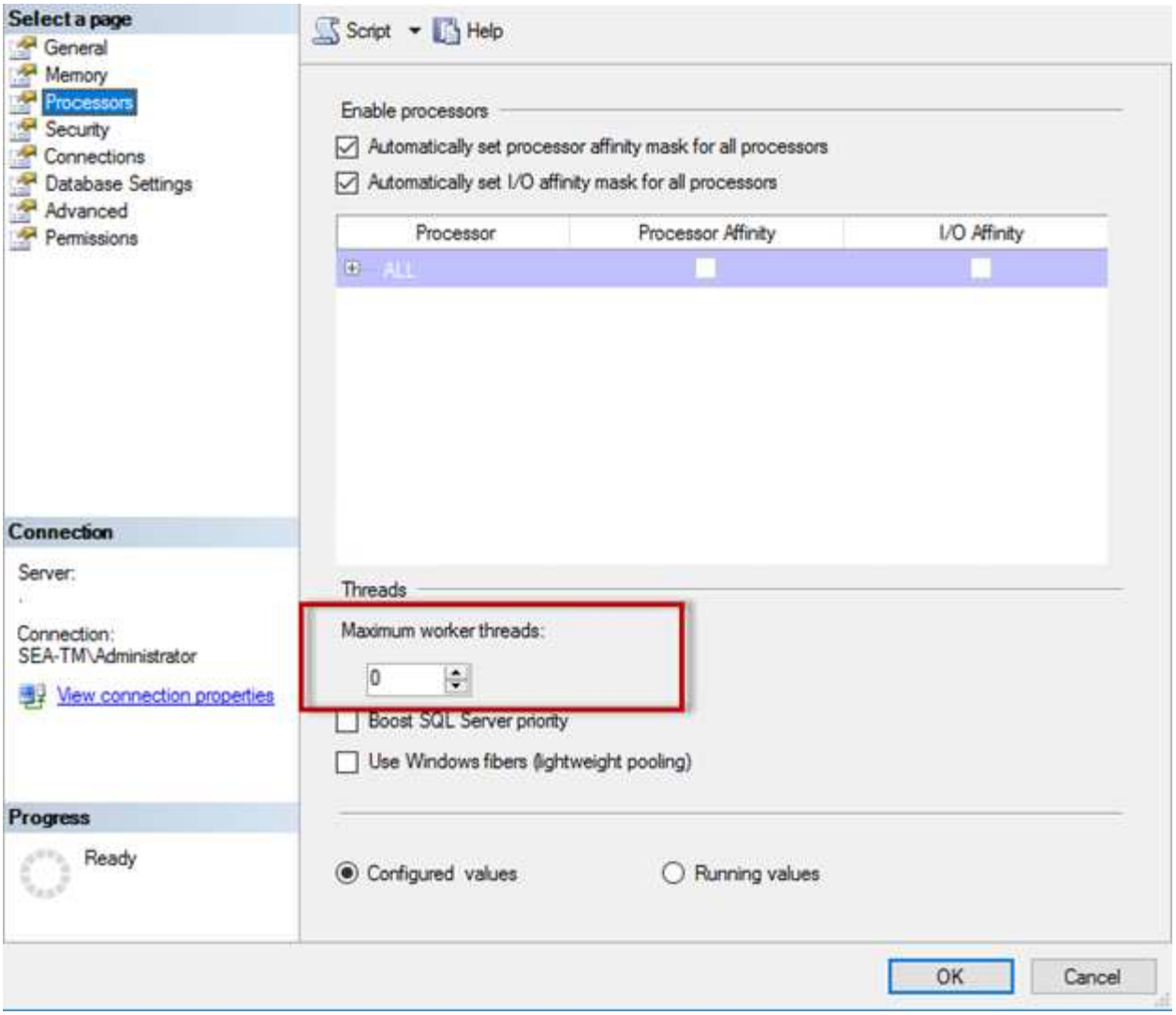
Numero massimo di thread di lavoro

L'opzione numero massimo di thread di lavoro consente di ottimizzare le prestazioni quando un numero elevato di client è connesso a SQL Server.

In genere, per ogni query viene creato un thread del sistema operativo separato. Se vengono effettuate centinaia di connessioni simultanee a SQL Server, la configurazione con un solo thread per query può consumare troppe risorse di sistema. L' `max worker threads`opzione consente di migliorare le prestazioni consentendo a SQL Server di creare un pool di thread di lavoro in grado di gestire collettivamente un maggior numero di richieste di query.

Il valore predefinito è 0, che consente a SQL Server di configurare automaticamente il numero di thread di lavoro all'avvio. Funziona per la maggior parte dei sistemi. Max worker Threads è un'opzione avanzata e non deve essere alterata senza l'assistenza di un amministratore di database esperto (DBA).

Quando è necessario configurare SQL Server per utilizzare più thread di lavoro? Se la lunghezza media della coda di lavoro per ogni pianificatore è superiore a 1, si potrebbe trarre vantaggio dall'aggiunta di più thread al sistema, ma solo se il carico non è legato alla CPU o se si verificano altre attese pesanti. Se si verifica uno di questi due eventi, l'aggiunta di altri thread non aiuta perché sono in attesa di altri colli di bottiglia del sistema. Per ulteriori informazioni sui thread di lavoro max, vedere "Configurare l'opzione di configurazione del server numero massimo di thread di lavoro".



Configurazione di max worker threads con SQL Server Management Studio.

Nell'esempio seguente viene illustrato come configurare l'opzione numero massimo di thread di lavoro utilizzando T-SQL.

```
EXEC sp_configure 'show advanced options', 1;
GO
RECONFIGURE ;
GO
EXEC sp_configure 'max worker threads', 900 ;
GO
RECONFIGURE;
GO
```

Configurazione della memoria

La sezione seguente illustra le impostazioni della memoria di SQL Server necessarie per ottimizzare le prestazioni del database.

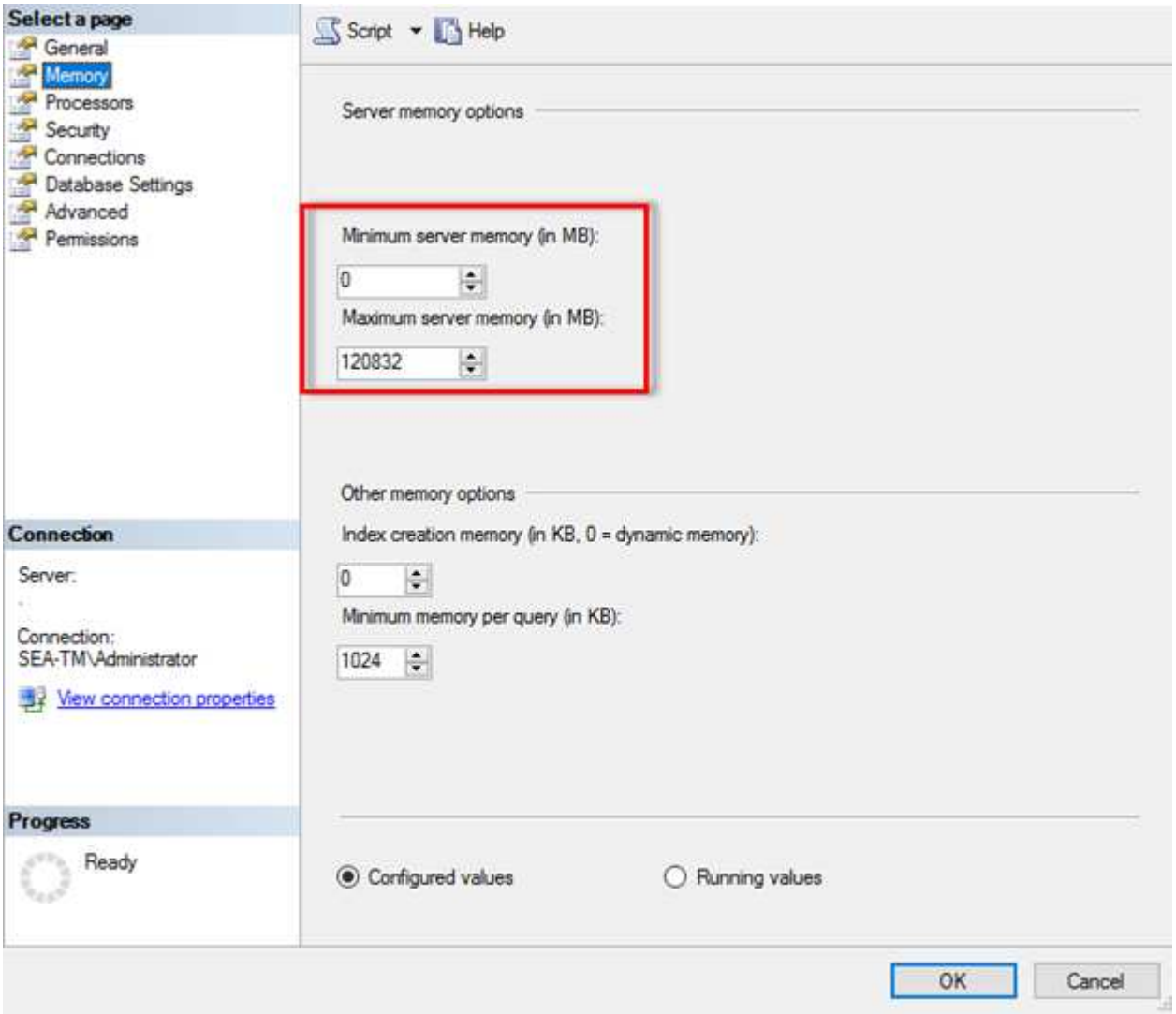
Memoria massima del server

L'opzione memoria massima del server imposta la quantità massima di memoria che l'istanza di SQL Server può utilizzare. Viene generalmente utilizzata se più applicazioni vengono eseguite sullo stesso server in cui SQL Server è in esecuzione e si desidera garantire che queste applicazioni dispongano di memoria sufficiente per funzionare correttamente.

Alcune applicazioni utilizzano solo la memoria disponibile all'avvio e non richiedono memoria aggiuntiva, anche se sono sotto pressione della memoria. È qui che entra in gioco l'impostazione della memoria massima del server.

In un cluster SQL Server con diverse istanze SQL Server, ciascuna istanza potrebbe competere per le risorse. L'impostazione di un limite di memoria per ciascuna istanza di SQL Server può contribuire a garantire le migliori prestazioni per ciascuna istanza.

NetApp consiglia di lasciare almeno 4GB o 6GB GB di RAM per il sistema operativo per evitare problemi di prestazioni.



Regolazione della memoria minima e massima del server mediante SQL Server Management Studio.

L'utilizzo di SQL Server Management Studio per regolare la memoria minima o massima del server richiede il riavvio del servizio SQL Server. È inoltre possibile regolare la memoria del server utilizzando Transact SQL (T-

SQL) utilizzando il seguente codice:

```
EXECUTE sp_configure 'show advanced options', 1
GO
EXECUTE sp_configure 'min server memory (MB)', 2048
GO
EXEC sp_configure 'max server memory (MB)', 120832
GO
RECONFIGURE WITH OVERRIDE
```

Accesso alla memoria non uniforme

L'accesso alla memoria non uniforme (NUMA, non Uniform Memory Access) è una tecnologia di ottimizzazione dell'accesso alla memoria che consente di evitare un carico extra sul bus del processore.

Se NUMA è configurato su un server in cui è installato SQL Server, non è necessaria alcuna configurazione aggiuntiva, in quanto SQL Server è compatibile con NUMA e funziona bene sull'hardware NUMA.

Indice creare memoria

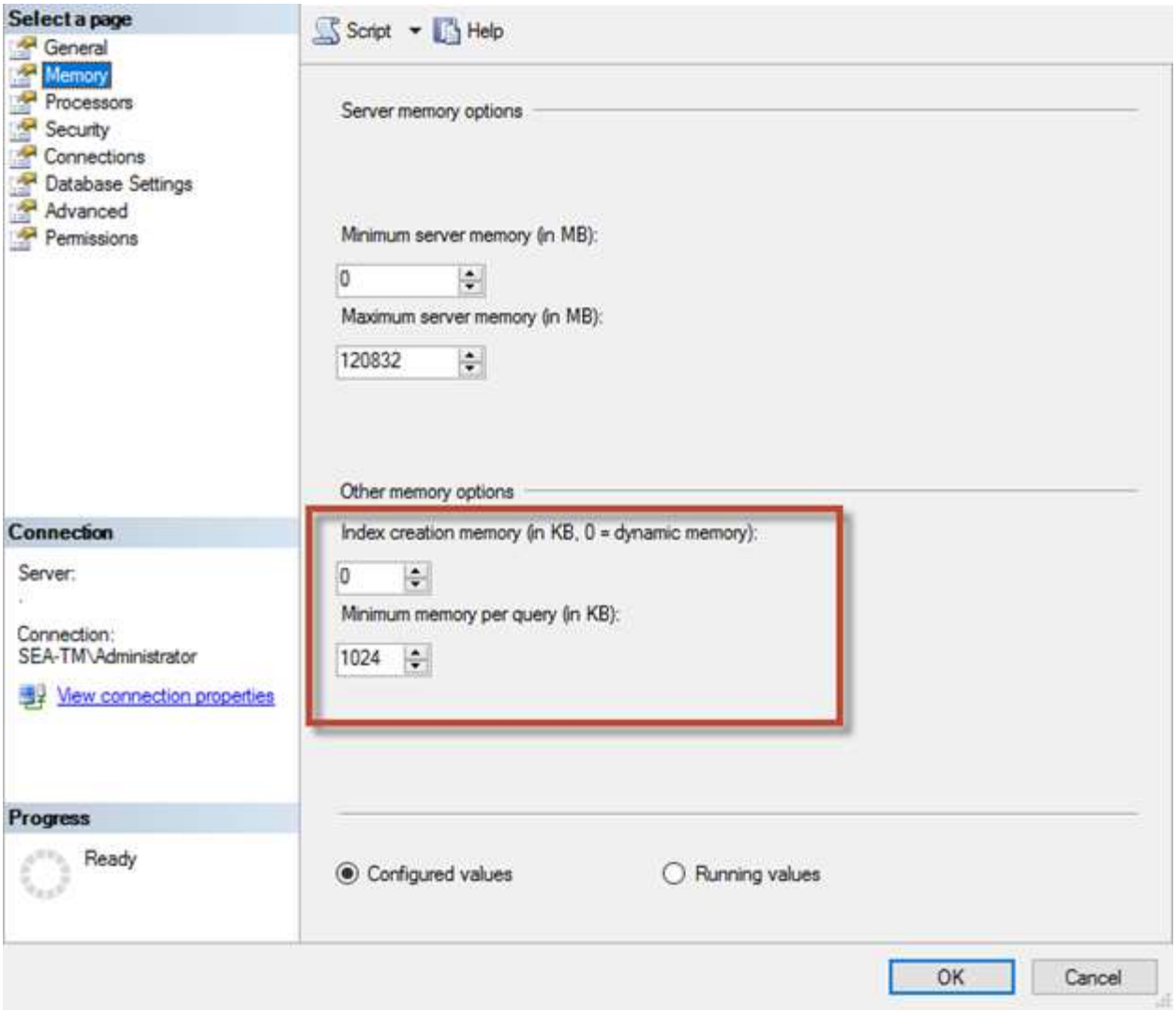
L'opzione di creazione della memoria di indice è un'altra opzione avanzata che non dovrebbe normalmente essere modificata dai valori predefiniti.

Controlla la quantità massima di RAM inizialmente allocata per la creazione degli indici. Il valore predefinito per questa opzione è 0, il che significa che è gestita automaticamente da SQL Server. Tuttavia, se si riscontrano difficoltà nella creazione degli indici, è consigliabile aumentare il valore di questa opzione.

Memoria minima per query

Quando viene eseguita una query, SQL Server tenta di allocare la quantità ottimale di memoria per un'esecuzione efficiente.

Per impostazione predefinita, l'impostazione memoria minima per query assegna >= a 1024KB per ogni query da eseguire. È consigliabile lasciare questa impostazione al valore predefinito per consentire a SQL Server di gestire dinamicamente la quantità di memoria allocata per le operazioni di creazione dell'indice. Tuttavia, se SQL Server dispone di una quantità di RAM superiore a quella necessaria per un'esecuzione efficiente, le prestazioni di alcune query possono essere migliorate se si aumenta questa impostazione. Pertanto, se sul server non viene utilizzata SQL Server, altre applicazioni o il sistema operativo è disponibile memoria, il miglioramento di questa impostazione può contribuire alle prestazioni complessive di SQL Server. Se non è disponibile memoria libera, l'aumento di questa impostazione potrebbe compromettere le prestazioni complessive.



Istanza condivisa contro istanza dedicata

SQL Server può essere configurato come singola istanza per server o come istanze multiple. La decisione giusta dipende in genere da fattori quali l'utilizzo del server per la produzione o lo sviluppo, indipendentemente dal fatto che l'istanza sia considerata di importanza critica per le operazioni aziendali e gli obiettivi prestazionali.

Le configurazioni delle istanze condivise possono essere inizialmente più semplici da configurare, ma possono causare problemi in cui le risorse vengono divise o bloccate, il che a sua volta causa problemi di prestazioni per altre applicazioni che hanno database ospitati nell'istanza condivisa di SQL Server.

La risoluzione dei problemi di prestazioni può essere complicata, perché è necessario capire quale istanza è la causa principale. Questa domanda è valutata rispetto ai costi delle licenze del sistema operativo e delle licenze di SQL Server. Se le performance applicative sono fondamentali, si consiglia vivamente un'istanza dedicata.

Microsoft concede in licenza SQL Server per core a livello di server e non per istanza. Per questo motivo, gli amministratori di database sono tentati di installare tutte le istanze di SQL Server che il server è in grado di gestire per risparmiare sui costi di licenza, il che può portare a gravi problemi di performance in un secondo momento.

NetApp consiglia di scegliere istanze dedicate di SQL Server quando possibile per ottenere prestazioni ottimali.

File tempdb

Il database Tempdb può essere utilizzato in modo intensivo. Oltre al posizionamento ottimale dei file di database degli utenti su ONTAP, anche il posizionamento dei file di dati tempdb è fondamentale per ridurre il conflitto di allocazione. Tempdb deve essere posizionato su un disco separato e non condiviso con i file di dati dell'utente.

Il conflitto di pagina può verificarsi nelle pagine GAM (Global allocation map), SGAM (Shared Global allocation map) o PFS (page free space) quando SQL Server deve scrivere in pagine di sistema speciali per allocare nuovi oggetti. I fermi bloccano queste pagine in memoria. In un'istanza SQL Server occupata, può essere necessario molto tempo per ottenere un blocco in una pagina di sistema in tempdb. Ciò si traduce in tempi di esecuzione delle query più lenti ed è noto come conflitto di latch. Per la creazione di file di dati tempdb, vedere le procedure consigliate riportate di seguito:

- Per < o = a 8 core: File di dati tempdb = numero di core
- Per più di 8 core: 8 file di dati tempdb
- Il file dati tempdb deve essere creato con le stesse dimensioni

Lo script di esempio seguente modifica tempdb creando otto file tempdb di pari dimensioni e spostando tempdb sul punto di montaggio `C:\MSSQL\tempdb` per SQL Server 2012 e versioni successive.

```
use master

go

-- Change logical tempdb file name first since SQL Server shipped with
logical file name called tempdev

alter database tempdb modify file (name = 'tempdev', newname =
'tempdev01');



-- Change location of tempdev01 and log file

alter database tempdb modify file (name = 'tempdev01', filename =
'C:\MSSQL\tempdb\tempdev01.mdf');

alter database tempdb modify file (name = 'templog', filename =
'C:\MSSQL\tempdb\templog.ldf');

GO

-- Assign proper size for tempdev01
```

```
ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'tempdev01', SIZE = 10GB );

ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'templog', SIZE = 10GB );

GO

-- Add more tempdb files

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev02', FILENAME =
N'C:\MSSQL\tempdb\tempdev02.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev03', FILENAME =
N'C:\MSSQL\tempdb\tempdev03.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev04', FILENAME =
N'C:\MSSQL\tempdb\tempdev04.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev05', FILENAME =
N'C:\MSSQL\tempdb\tempdev05.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev06', FILENAME =
N'C:\MSSQL\tempdb\tempdev06.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev07', FILENAME =
N'C:\MSSQL\tempdb\tempdev07.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev08', FILENAME =
N'C:\MSSQL\tempdb\tempdev08.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

GO
```

A partire da SQL Server 2016, il numero di core di CPU visibili al sistema operativo viene rilevato automaticamente durante l'installazione e, in base a tale numero, SQL Server calcola e configura il numero di file tempdb necessari per ottenere prestazioni ottimali.

Configurazione dello storage su sistemi AFF/FAS

Panoramica

La combinazione delle soluzioni storage ONTAP e Microsoft SQL Server genera design di storage per database di livello Enterprise in grado di soddisfare le più esigenti esigenze applicative odierne.

L'ottimizzazione di una soluzione SQL Server su ONTAP richiede la comprensione del modello e delle caratteristiche di i/o di SQL Server. Un layout di storage ben progettato per un database SQL Server deve supportare i requisiti relativi alle performance di SQL Server, offrendo al contempo la massima capacità di

gestione dell'infrastruttura nel suo complesso. Un buon layout dello storage permette inoltre di avere successo nell'implementazione iniziale e di far crescere l'ambiente senza problemi nel tempo, con il crescere dell'azienda.

Progettazione dello storage dei dati

Per i database SQL Server che non utilizzano SnapCenter per eseguire i backup, Microsoft consiglia di posizionare i file di dati e di log su dischi separati. Per le applicazioni che aggiornano e richiedono contemporaneamente i dati, il file di log è intensivo in scrittura e il file di dati (a seconda dell'applicazione) è intensivo in lettura/scrittura. Per il recupero dei dati, il file di log non è necessario. Pertanto, le richieste di dati possono essere soddisfatte dal file di dati posto sul proprio disco.

Quando si crea un nuovo database, Microsoft consiglia di specificare unità separate per i dati e i registri. Per spostare i file dopo la creazione del database, il database deve essere portato offline. Per ulteriori consigli Microsoft, vedere "Posizionare i file di dati e di registro su unità separate".

Aggregati

Gli aggregati sono i container di storage di livello più basso per le configurazioni di storage NetApp. Su Internet esiste una documentazione legacy che consiglia di separare i/o su diversi set di unità sottostanti. Questa operazione non è consigliata con ONTAP. NetApp ha eseguito diverse prove di caratterizzazione dei carichi di lavoro i/o utilizzando aggregati condivisi e dedicati con file di dati e file di log delle transazioni separati. I test dimostrano che un aggregato di grandi dimensioni con più gruppi RAID e dischi ottimizza e migliora le performance dello storage ed è più semplice da gestire per due motivi:

- Un aggregato di grandi dimensioni rende disponibili per tutti i file le funzionalità i/o di tutte le unità.
- Un grande aggregato consente l'utilizzo più efficiente dello spazio su disco.

Per l'high Availability (ha), posiziona la replica sincrona secondaria di SQL Server Always on Availability Group su una Storage Virtual Machine (SVM) separata nell'aggregato. Per scopi di disaster recovery, posizionare la replica asincrona in un aggregato che fa parte di un cluster di storage separato nel sito di disaster recovery, con contenuto replicato utilizzando la tecnologia NetApp SnapMirror. NetApp consiglia di disporre di almeno il 10% di spazio libero in un aggregato per ottenere performance dello storage ottimali.

Volumi

i volumi vengono creati e risiedono all'interno degli aggregati. Questo termine talvolta causa confusione perché un volume ONTAP non è un LUN. Un volume ONTAP è un container di gestione per i dati. Un volume può contenere file, LUN o persino oggetti S3. Un volume non occupa spazio, ma viene utilizzato solo per la gestione dei dati contenuti.

Considerazioni sulla progettazione dei volumi

Prima di creare una progettazione di volumi di database, è importante comprendere in che modo il modello i/o di SQL Server e le relative caratteristiche variano in base al carico di lavoro e ai requisiti di backup e ripristino. Consulta i seguenti consigli NetApp per i volumi flessibili:

- Evitare di condividere i volumi tra gli host. Ad esempio, anche se sarebbe possibile creare 2 LUN in un singolo volume e condividere ogni LUN con un host diverso, questo aspetto dovrebbe essere evitato perché complica la gestione. Se si eseguono più istanze di SQL Server sullo stesso host, a meno che non ci si trovi in prossimità del limite di volume su un nodo, è possibile evitare la condivisione di volumi e utilizzare un volume separato per istanza per host per semplificare la gestione dei dati.
- Utilizzare i punti di montaggio NTFS invece delle lettere dell'unità per superare il limite di 26 lettere di unità in Windows. Quando si utilizzano punti di montaggio del volume, si consiglia di assegnare all'etichetta del

volume lo stesso nome del punto di montaggio.

- Se necessario, configurare un criterio di dimensionamento automatico dei volumi per evitare condizioni di spazio insufficiente.
- Se si installa SQL Server su una condivisione SMB, assicurarsi che Unicode sia attivato sui volumi SMB per la creazione delle cartelle.
- Impostare il valore di riserva snapshot nel volume su zero per semplificare il monitoraggio dal punto di vista operativo.
- Disattivare le pianificazioni delle snapshot e i criteri di conservazione. Utilizzare invece SnapCenter per coordinare le copie Snapshot dei volumi di dati di SQL Server.
- Posizionare i database di sistema di SQL Server su un volume dedicato.
- Tempdb è un database di sistema utilizzato da SQL Server come area di lavoro temporanea, in particolare per operazioni DBCC CHECKDB i/o intensive. Pertanto, collocare questo database su un volume dedicato con un set separato di spindle. In ambienti di grandi dimensioni in cui il numero di volumi rappresenta una sfida, è possibile consolidare il tempdb in un numero inferiore di volumi e memorizzarlo nello stesso volume degli altri database di sistema dopo un'attenta pianificazione. La protezione dei dati per tempdb non è una priorità elevata perché questo database viene ricreato ogni volta che SQL Server viene riavviato.
- Collocare i file di dati utente (`.mdf`) su volumi separati perché si tratta di carichi di lavoro di lettura/scrittura casuali. È comune creare backup del log delle transazioni con maggiore frequenza rispetto ai backup del database. Per questo motivo, inserire i file di log delle transazioni (`.ldf`) in un volume separato o VMDK dai file di dati in modo che sia possibile creare pianificazioni di backup indipendenti per ciascuno di essi. Questa separazione isola inoltre l'i/o di scrittura sequenziale dei file di log dall'i/o di lettura/scrittura casuale dei file di dati e migliora significativamente le prestazioni di SQL Server.

LUN

- Assicurarsi che i file del database utente e la directory di registro per l'archiviazione del backup del registro si trovino su volumi separati per evitare che il criterio di conservazione sovrascriva gli snapshot quando vengono utilizzati con la tecnologia SnapVault.
- Non combinare file di database e non di database, come file di ricerca full-text, sullo stesso LUN.
- L'inserimento di file secondari del database (come parte di un filegroup) in volumi separati migliora le prestazioni del database di SQL Server. Questa separazione è valida solo se il file del database `.mdf` non condivide il LUN con altri `.mdf` file.
- Se si creano LUN con DiskManager o altri strumenti, assicurarsi che la dimensione dell'unità di allocazione sia impostata su 64K per le partizioni durante la formattazione dei LUN.
- Vedere "Microsoft Windows e MPIO nativo nelle Best practice ONTAP per le SAN moderne" Per applicare il supporto multipathing in Windows ai dispositivi iSCSI nelle proprietà MPIO.

File di database e filegroup

Il corretto posizionamento dei file del database SQL Server su ONTAP è fondamentale durante la fase di distribuzione iniziale. Ciò garantisce prestazioni ottimali, gestione dello spazio, tempi di backup e ripristino che possono essere configurati in base alle esigenze aziendali.

In teoria, SQL Server (a 64 bit) supporta 32.767 database per istanza e 524.272TB di dimensioni del database, sebbene l'installazione tipica abbia in genere diversi database. Tuttavia, il numero di database che SQL Server è in grado di gestire dipende dal carico e dall'hardware. Non è insolito vedere le istanze di SQL Server che

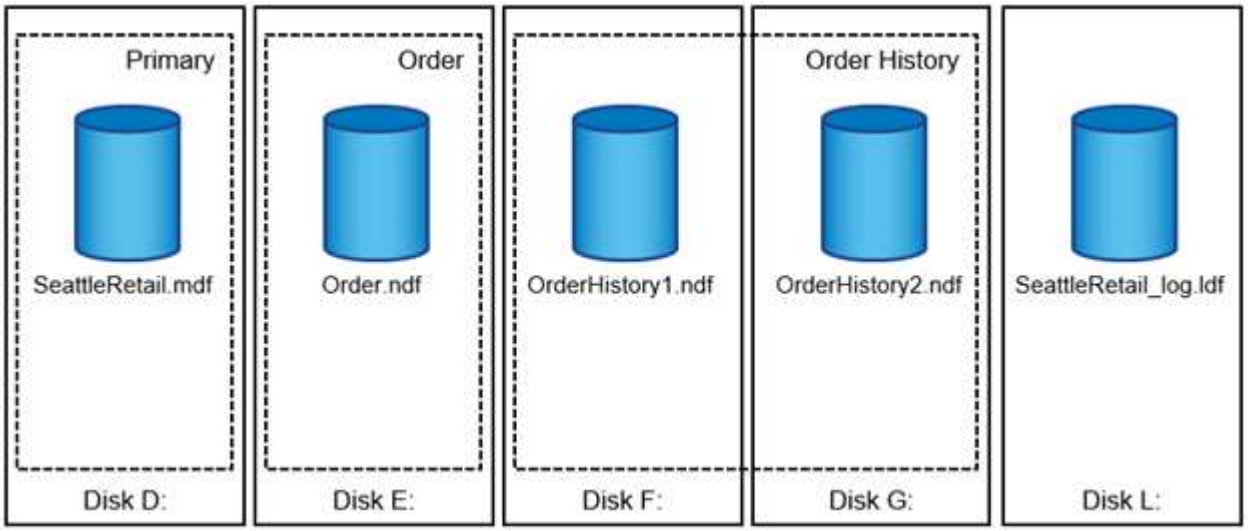
ospitano decine, centinaia o persino migliaia di database di piccole dimensioni.

File di database e filegroup

Ogni database è costituito da uno o più file di dati e da uno o più file di registro delle transazioni. Il registro delle transazioni memorizza le informazioni sulle transazioni del database e tutte le modifiche apportate ai dati da ciascuna sessione. Ogni volta che i dati vengono modificati, SQL Server memorizza informazioni sufficienti nel log delle transazioni per annullare (eseguire il rollback) o ripristinare (riprodurre nuovamente) l'azione. Un log delle transazioni di SQL Server è parte integrante della reputazione di SQL Server in termini di integrità e robustezza dei dati. Il log delle transazioni è fondamentale per le funzionalità di atomicità, coerenza, isolamento e durata (ACID) di SQL Server. SQL Server scrive nel registro delle transazioni non appena si verifica una modifica alla pagina dei dati. Ogni istruzione DML (Data Manipulation Language) (ad esempio, SELECT, INSERT, Update o DELETE) è una transazione completa e il log delle transazioni garantisce che l'intera operazione basata su set abbia luogo, assicurando l'atomicità della transazione.

Ogni database dispone di un file di dati primario che, per impostazione predefinita, ha l'estensione .mdf. Inoltre, ogni database può disporre di file di database secondari. Questi file, per impostazione predefinita, hanno estensioni .ndf.

Tutti i file di database sono raggruppati in filegroup. Un filegroup è l'unità logica, che semplifica l'amministrazione del database. Consentono la separazione tra il posizionamento degli oggetti logici e i file di database fisici. Quando si creano le tabelle degli oggetti del database, si specifica in quale filegroup devono essere posizionati senza preoccuparsi della configurazione del file di dati sottostante.



La possibilità di inserire più file di dati all'interno del filegroup consente di distribuire il carico su diversi dispositivi di archiviazione, migliorando le prestazioni di i/o del sistema. Al contrario, il log delle transazioni non trae vantaggio dai file multipli poiché SQL Server scrive nel log delle transazioni in modo sequenziale.

La separazione tra il posizionamento degli oggetti logici nei filegroup e i file di database fisici consente di ottimizzare il layout dei file di database, ottenendo il massimo dal sottosistema di storage. Il numero di file di dati che supportano un carico di lavoro può essere variato in base alle necessità per supportare i requisiti di i/o e la capacità prevista, senza influire sull'applicazione. Queste variazioni nel layout del database sono trasparenti per gli sviluppatori di applicazioni, che posizionano gli oggetti del database nei filegroup piuttosto che nei file di database.

NetApp recommended evitare l'utilizzo del filegroup primario per oggetti diversi da quelli di sistema. La creazione di un filegroup separato o di un set di filegroup per gli oggetti utente semplifica l'amministrazione del database e il ripristino di emergenza, soprattutto nel caso di database di grandi dimensioni.

Inizializzazione del file di istanza del database

È possibile specificare le dimensioni iniziali del file e i parametri di crescita automatica al momento della creazione del database o dell'aggiunta di nuovi file a un database esistente. SQL Server utilizza un algoritmo di riempimento proporzionale quando sceglie in quale file di dati scrivere i dati. Scrive una quantità di dati proporzionalmente allo spazio libero disponibile nei file. Maggiore è lo spazio libero nel file, maggiore è il numero di scritture gestite.

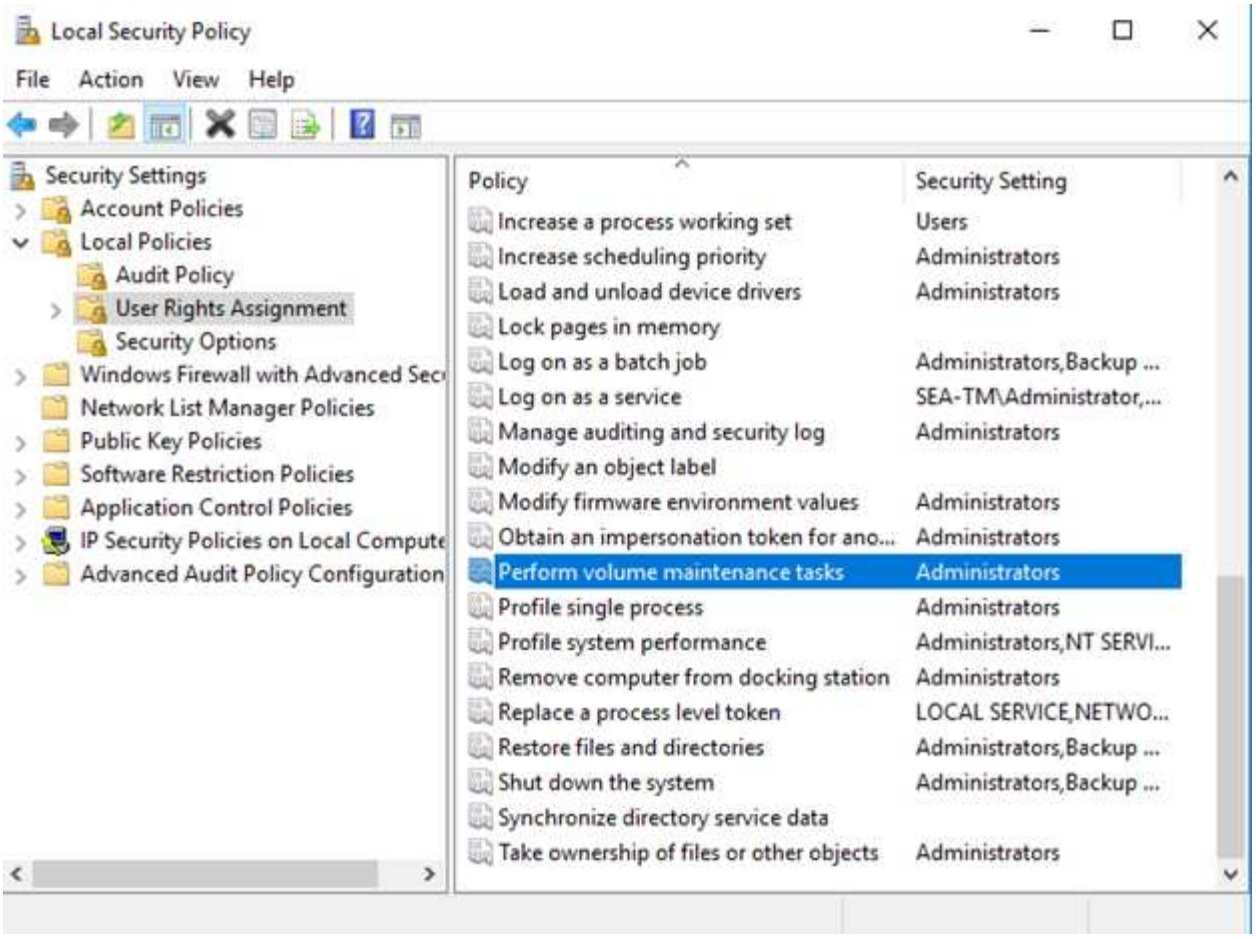
NetApp consiglia che tutti i file nel singolo filegroup abbiano le stesse dimensioni iniziali e parametri di crescita automatica, con la dimensione di crescita definita in megabyte piuttosto che in percentuali. Questo aiuta l'algoritmo di riempimento proporzionale a bilanciare uniformemente le attività di scrittura nei file di dati.

Ogni volta che SQL Server espande i file, riempie di zero lo spazio appena allocato. Questo processo blocca tutte le sessioni che devono scrivere nel file corrispondente o, in caso di crescita del log delle transazioni, genera record di log delle transazioni.

SQL Server azzera sempre il log delle transazioni e questo comportamento non può essere modificato. Tuttavia, è possibile controllare se i file di dati vengono azzerati attivando o disattivando l'inizializzazione istantanea dei file. L'attivazione dell'inizializzazione immediata dei file consente di velocizzare la crescita dei file di dati e di ridurre il tempo necessario per creare o ripristinare il database.

Un piccolo rischio per la sicurezza è associato all'inizializzazione immediata dei file. Quando questa opzione è attivata, le parti non allocate del file di dati possono contenere informazioni provenienti da file del sistema operativo eliminati in precedenza. Gli amministratori di database possono esaminare tali dati.

È possibile attivare l'inizializzazione immediata dei file aggiungendo l'autorizzazione SA_MANAGE_VOLUME_NAME, nota anche come "Esegui attività di manutenzione del volume" all'account di avvio di SQL Server. È possibile eseguire questa operazione nell'applicazione di gestione dei criteri di protezione locale (secpol.msc), come illustrato nella figura seguente. Aprire le proprietà per l'autorizzazione "Esegui attività di manutenzione del volume" e aggiungere l'account di avvio di SQL Server all'elenco degli utenti.



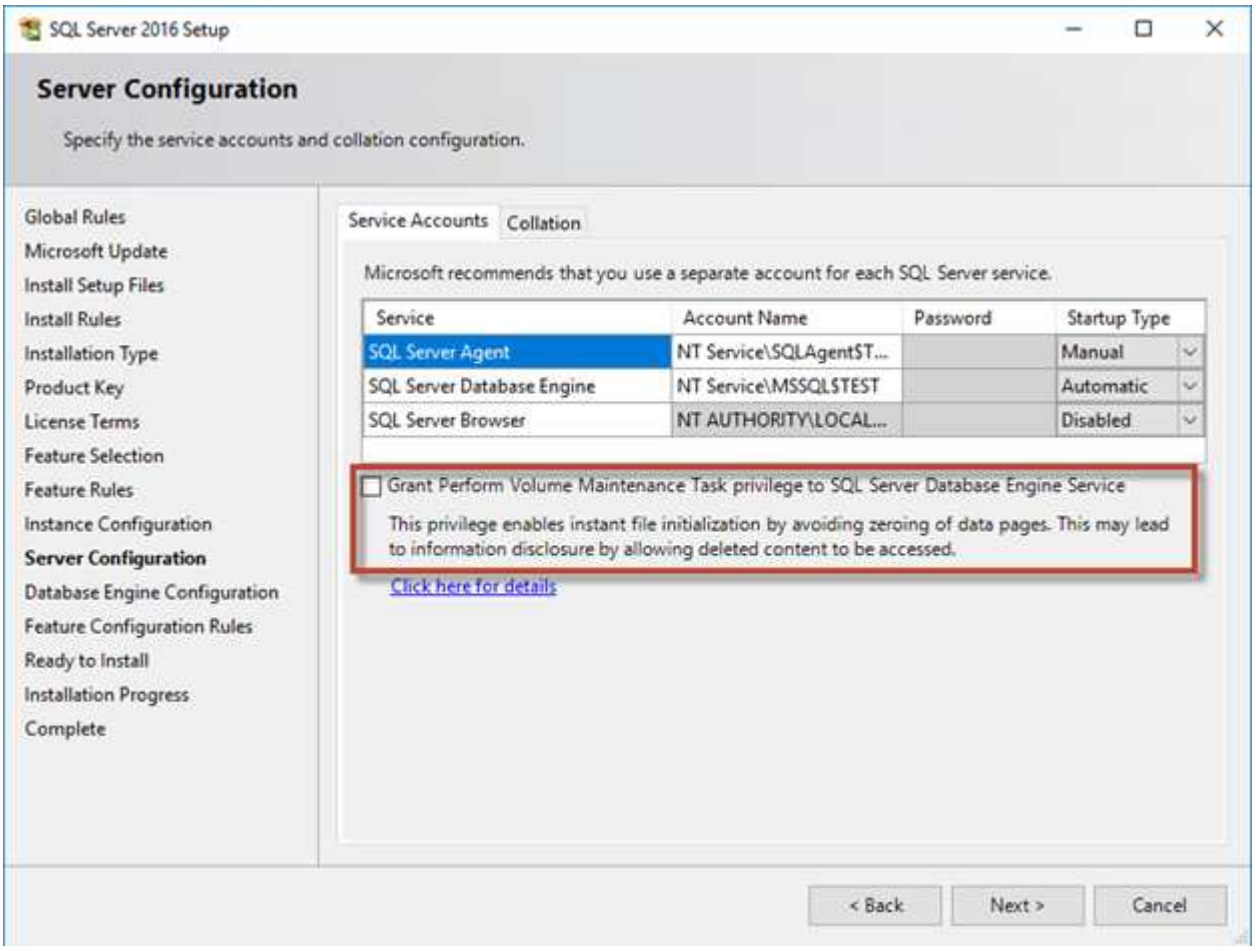
Per verificare se l'autorizzazione è attivata, è possibile utilizzare il codice riportato nell'esempio seguente. Questo codice imposta due flag di traccia che obbligano SQL Server a scrivere informazioni aggiuntive nel registro degli errori, a creare un database di piccole dimensioni e a leggere il contenuto del registro.

```
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO
```

Quando l'inizializzazione immediata del file non è attivata, il registro degli errori di SQL Server mostra che SQL Server sta azzerando il file di dati mdf oltre a azzerare il file di registro ldf, come illustrato nell'esempio seguente. Quando l'inizializzazione immediata del file è attivata, viene visualizzato solo l'azzeramento del file di registro.

	LogDate	ProcessInfo	Text
365	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 flush delta counts.
366	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 logging active xact info.
367	2017-02-09 08:10:07.750	spid53	Ckpt dbid 3 phase 1 ended (8)
368	2017-02-09 08:10:07.750	spid53	About to log Checkpoint end.
369	2017-02-09 08:10:07.880	spid53	Ckpt dbid 3 complete
370	2017-02-09 08:10:08.130	spid53	Starting up database 'DelMe'.
371	2017-02-09 08:10:08.150	spid53	FixupLogTail(progress) zeroing C:\Program Files\Micros
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQ
373	2017-02-09 08:10:08.170	spid53	Zeroing completed on C:\Program Files\Microsoft SQL
374	2017-02-09 08:10:08.710	spid53	Ckpt dbid 6 started
375	2017-02-09 08:10:08.710	spid53	About to log Checkpoint begin.

L'attività Esegui manutenzione volume è semplificata in SQL Server 2016 e viene fornita come opzione durante il processo di installazione. In questa figura viene visualizzata l'opzione per concedere al servizio del motore di database di SQL Server il privilegio di eseguire l'attività di manutenzione del volume.



Un'altra importante opzione del database che controlla le dimensioni dei file di database è l'autohrink. Quando questa opzione è attivata, SQL Server riduce regolarmente i file di database, ne riduce le dimensioni e rilascia spazio al sistema operativo. Questa operazione richiede molte risorse ed è raramente utile perché i file di database crescono di nuovo dopo un certo periodo di tempo quando nuovi dati entrano nel sistema. Il collegamento automatico non deve essere attivato nel database.

Directory di log

La directory di registro è specificata in SQL Server per memorizzare i dati di backup del registro delle transazioni a livello di host. Se si utilizza SnapCenter per eseguire il backup dei file di registro, ciascun host SQL Server utilizzato da SnapCenter deve disporre di una directory di registro host configurata per eseguire i backup dei registri. SnapCenter dispone di un repository di database, pertanto i metadati relativi alle operazioni di backup, ripristino o clonazione vengono memorizzati in un repository di database centrale.

Le dimensioni della directory del registro host vengono calcolate come segue:
Dimensione della directory del log host = ((dimensione massima LDF DB x velocità di modifica giornaliera del log %) x (conservazione snapshot) ÷ (1 - spazio di overhead LUN %)
La formula di dimensionamento della directory del registro host presuppone uno spazio di overhead LUN del 10%

Posizionare la directory di registro su un volume o LUN dedicato. La quantità di dati nella directory del registro host dipende dalle dimensioni dei backup e dal numero di giorni in cui i backup vengono conservati. SnapCenter consente una sola directory di registro host per host SQL Server. È possibile configurare le directory del registro host in SnapCenter -→ host -→ Configura plug-in.

NetApp consiglia quanto segue per una directory del registro host:

- Assicurarsi che la directory del registro host non sia condivisa da altri tipi di dati che potrebbero danneggiare i dati dello snapshot di backup.
- Non posizionare database utente o database di sistema su un LUN che ospita punti di montaggio.
- Creare la directory di registro dell'host su un volume dedicato in cui SnapCenter copia i registri delle transazioni.
- Utilizzare le procedure guidate SnapCenter per migrare i database nello storage NetApp in modo che i database vengano memorizzati in posizioni valide, consentendo operazioni di backup e ripristino SnapCenter corrette. Tenere presente che il processo di migrazione causa interruzioni e può causare la disconnessione dei database mentre è in corso la migrazione.
- Per le istanze di cluster di failover (FCI) di SQL Server devono essere presenti le seguenti condizioni:
  - Se si utilizza un'istanza del cluster di failover, il LUN della directory del log host deve essere una risorsa del disco del cluster nello stesso gruppo di cluster dell'istanza di SQL Server di cui viene eseguito il backup in SnapCenter.
  - Se si utilizza un'istanza cluster di failover, i database utente devono essere collocati su LUN condivisi che sono risorse cluster di dischi fisici assegnate al gruppo di cluster associato all'istanza di SQL Server.

Efficienza dello storage

L'efficienza dello storage di ONTAP è ottimizzata per memorizzare e gestire i dati di SQL Server in modo che utilizzino la minore quantità di spazio di storage senza impatti sulle performance.

Le funzionalità di efficienza in termini di spazio, come compressione, compaction e deduplica, sono progettate per aumentare la quantità di dati logici applicabili a una determinata quantità di storage fisico. Il risultato è una riduzione dei costi e dell'overhead di gestione.

Ad un livello elevato, la compressione è un processo matematico in cui gli schemi nei dati vengono rilevati e codificati in modo da ridurre i requisiti di spazio. La deduplica, invece, rileva i blocchi di dati effettivi e ripetuti e rimuove le copie estranee. La tecnologia di compaction consente a più blocchi logici di dati di condividere lo stesso blocco fisico sui supporti.

Per una spiegazione dell'interazione tra efficienza dello storage e prenotazione frazionata, vedere le sezioni seguenti sul thin provisioning.

Compressione

Prima della disponibilità dei sistemi storage all-flash, la compressione basata su array aveva un valore limitato, perché la maggior parte dei carichi di lavoro con i/o-intensive richiedeva un numero molto elevato di spindle per fornire performance accettabili. I sistemi storage contenevano invariabilmente una capacità superiore rispetto a quella richiesta come effetto collaterale dell'elevato numero di dischi. La situazione è cambiata con l'ascesa dello storage a stato solido. Non è più necessario effettuare un provisioning in eccesso significativo dei dischi solo per ottenere buone prestazioni. Lo spazio su disco di un sistema di storage può essere adattato alle effettive esigenze di capacità.

L'aumento della capacità degli IOPS dei dischi a stato solido (SSD) offre quasi sempre risparmi sui costi rispetto ai dischi rotanti, ma la compressione può ottenere ulteriori risparmi aumentando la capacità effettiva dei supporti a stato solido.

Esistono diversi modi per comprimere i dati. Molti database includono proprie funzionalità di compressione, sebbene raramente queste vengano osservate negli ambienti dei clienti. Il motivo è solitamente la penalizzazione delle prestazioni per una **modifica** dei dati compressi, mentre con alcune applicazioni vi sono elevati costi di licenza per la compressione a livello di database. Infine, ci sono le conseguenze globali delle performance sulle operazioni di database. Ha poco senso pagare un costo elevato di licenza per CPU per una CPU che esegue la compressione e la decompressione dei dati piuttosto che un vero lavoro di database. Un'opzione migliore è trasferire il lavoro di compressione sul sistema storage.

Compressione adattiva

La compressione adattiva è stata testata accuratamente con carichi di lavoro Enterprise senza effetti osservati sulle performance, anche in un ambiente all-flash in cui la latenza viene misurata in microsecondi. Alcuni clienti hanno anche segnalato un aumento delle performance con l'utilizzo della compressione, perché i dati rimangono compressi nella cache, aumentando di fatto la quantità di cache disponibile in un controller.

ONTAP gestisce i blocchi fisici in 4KB unità. La compressione adattiva utilizza dimensioni predefinite dei blocchi di compressione di 8KB KB, il che significa che i dati sono compressi in unità da 8KB KB. Corrisponde alle dimensioni dei blocchi di 8KB KB utilizzate più spesso dai database relazionali. Gli algoritmi di compressione diventano più efficienti con la compressione di un numero maggiore di dati come una singola unità. Una dimensione dei blocchi di compressione da 32KB KB sarebbe più efficiente in termini di spazio rispetto a un'unità dei blocchi di compressione da 8KB KB. Ciò significa che la compressione adattiva che utilizza le dimensioni predefinite dei blocchi di 8KB KB produce tassi di efficienza leggermente inferiori, ma esiste anche un vantaggio significativo nell'utilizzo di dimensioni inferiori dei blocchi di compressione. I carichi di lavoro dei database includono un'elevata attività di sovrascrittura. La sovrascrittura di un 8KB di un blocco di dati 32KB compresso richiede la lettura dell'intero 32KB di dati logici, la decompressione, l'aggiornamento della regione 8KB richiesta, la ricompressione e quindi la riscrittura dell'intero 32KB sui dischi. Si tratta di un'operazione molto costosa per un sistema storage ed è il motivo per cui alcuni storage array concorrenti basati su dimensioni dei blocchi di compressione più grandi implicano anche una significativa penalizzazione delle performance con i carichi di lavoro dei database.

Le dimensioni dei blocchi utilizzate dalla compressione adattiva possono essere aumentate fino a 32KB KB. Questo può migliorare l'efficienza di archiviazione e dovrebbe essere considerato per i file inattivi come i log delle transazioni e i file di backup quando una quantità sostanziale di tali dati è memorizzata nell'array. In alcune situazioni, i database attivi che utilizzano dimensioni blocco 16KB KB o 32KB KB possono anche trarre vantaggio dall'aumento delle dimensioni blocco della compressione adattiva per adeguarsi. Consulta un NetApp o un rappresentante del partner per ottenere indicazioni relative all'adeguatezza del tuo carico di lavoro.

Le dimensioni dei blocchi di compressione superiori a 8KB KB non devono essere utilizzate insieme alla deduplica nelle destinazioni di backup in streaming. Il motivo è che piccole modifiche ai dati di backup influiscono sulla finestra di compressione 32KB. Se la finestra si sposta, i dati compressi risultanti differiscono per l'intero file. La deduplica si verifica dopo la compressione, il che significa che il motore di deduplica vede ogni backup compresso in modo diverso. Se è richiesta la deduplica dei backup in streaming, è consigliabile utilizzare solo la compressione adattiva per blocchi da 8KB KB. La compressione adattiva è preferibile, perché funziona a blocchi di dimensioni inferiori e non interrompe l'efficienza di deduplica. Per motivi simili, la compressione lato host interferisce anche con l'efficienza della deduplica.

Allineamento delle compressioni

La compressione adattiva in un ambiente di database richiede alcune considerazioni sull'allineamento dei blocchi di compressione. Ciò rappresenta solo una preoccupazione per i dati che sono soggetti a sovrascritture casuali di blocchi molto specifici. Questo approccio è simile in teoria all'allineamento complessivo del file system, dove l'inizio di un file system deve essere allineato al limite di un dispositivo 4K e la dimensione di blocco di un file system deve essere un multiplo di 4K.

Ad esempio, una scrittura 8KB in un file viene compressa solo se si allinea con un limite 8KB all'interno del file system stesso. Questo punto significa che deve rientrare nel primo 8KB del file, nel secondo 8KB del file e così via. Il modo più semplice per garantire un corretto allineamento è utilizzare il tipo di LUN corretto, ogni partizione creata dovrebbe avere un offset dall'inizio del dispositivo che è un multiplo di 8K, e utilizzare una dimensione del blocco del file system che è un multiplo della dimensione del blocco del database.

Dati come backup o log delle transazioni sono operazioni scritte in sequenza che coprono più blocchi, tutti compressi. Pertanto, non è necessario considerare l'allineamento. L'unico modello di i/o che desta preoccupazione sono le sovrascritture casuali dei file.

Compaction dei dati

La data compaction è una tecnologia che migliora l'efficienza di compressione. Come indicato in precedenza, la sola compressione adattiva può garantire risparmi 2:1:1 al meglio, perché è limitata alla memorizzazione di un i/o da 8KB KB in un blocco WAFL da 4KB KB. I metodi di compressione con dimensioni dei blocchi maggiori garantiscono una maggiore efficienza. Tuttavia, non sono adatte per i dati che sono soggetti a piccole sovrascritture dei blocchi. La decompressione di 32KB unità di dati, l'aggiornamento di una porzione 8KB, la ricompressione e la riscrittura sui dischi crea overhead.

La data compaction opera consentendo di memorizzare più blocchi logici all'interno dei blocchi fisici. Ad esempio, un database con dati altamente comprimibili come testo o blocchi parzialmente completi può comprimere da 8KB a 1KB. Senza la compaction, quei 1KB PB di dati continuerebbero ad occupare un intero blocco da 4KB KB. Inline data compaction per memorizzare 1KB TB di dati compressi in sole 1KB:1 di spazio fisico insieme ad altri dati compressi. Non si tratta di una tecnologia di compressione, ma semplicemente di un metodo più efficiente per allocare spazio sulle unità e quindi non dovrebbe creare alcun effetto rilevabile sulle prestazioni.

Il grado di risparmio ottenuto varia. I dati già compressi o crittografati non possono in genere essere

ulteriormente compressi, e pertanto tali set di dati non traggono vantaggio dalla compattazione. Al contrario, i file di dati appena inizializzati contenenti poco più dei metadati dei blocchi e la compressione di zeri fino a 80:1.

Efficienza di conservazione sensibile alla temperatura

L'efficienza di stoccaggio sensibile alla temperatura (TSSE) è disponibile in ONTAP 9.8 e versioni successive. Si affida alle mappe termiche di accesso ai blocchi per identificare i blocchi a cui si accede raramente e comprimerli con una maggiore efficienza.

Deduplica

La deduplica consiste nella rimozione di dimensioni dei blocchi duplicate da un set di dati. Ad esempio, se lo stesso blocco 4KB esistesse in 10 file diversi, la deduplica reindirizzerebbe quel blocco 4KB in tutti i file 10 allo stesso blocco fisico da 4KB KB. Il risultato sarebbe un miglioramento di 10:1 volte in efficienza per quei dati.

Dati come i LUN di avvio guest di VMware si deduplicano in genere in modo estremamente efficace poiché sono costituiti da più copie degli stessi file del sistema operativo. Sono state osservate un'efficienza pari o superiore a 100:1.

Alcuni dati non contengono dati duplicati. Ad esempio, un blocco Oracle contiene un'intestazione univoca a livello globale per il database e un trailer quasi univoco. Di conseguenza, la deduplica di un database Oracle raramente offre un risparmio superiore al 1%. La deduplica con i database MS SQL è leggermente migliore, ma i metadati univoci a livello di blocco rimangono un limite.

In pochi casi, sono stati osservati risparmi di spazio fino al 15% nei database con blocchi di dimensioni grandi e 16KB. Il 4KB iniziale di ciascun blocco contiene la testata unica a livello globale, mentre il 4KB finale contiene il rimorchio quasi unico. I blocchi interni sono candidati per la deduplica, sebbene in pratica ciò sia quasi interamente attribuito alla deduplica di dati azzerati.

Molti array della concorrenza rivendicano la capacità di deduplicare i database sulla base del presupposto che un database venga copiato più volte. Anche in questo caso è possibile utilizzare la deduplica NetApp, ma ONTAP offre un'opzione migliore: La tecnologia FlexClone di NetApp. Il risultato finale è lo stesso; vengono create più copie di un database che condividono la maggior parte dei blocchi fisici sottostanti. L'utilizzo di FlexClone è molto più efficiente della necessità di dedicare tempo alla copia e alla deduplica dei file di database. In effetti, non viene effettuata alcuna duplicazione piuttosto che deduplica, poiché al primo posto non viene mai creato un duplicato.

Efficienza e thin provisioning

Le funzionalità di efficienza sono forme di thin provisioning. Ad esempio, una LUN da 100GB GB che occupa un volume da 100GB GB potrebbe comprimere fino a 50GB GB. Non ci sono risparmi effettivi ancora realizzati perché il volume è ancora 100GB. Le dimensioni del volume devono essere innanzitutto ridotte in modo che lo spazio salvato possa essere utilizzato in un'altra posizione del sistema. Se successivamente le modifiche apportate al LUN da 100GB GB rendono i dati meno comprimibili, il LUN aumenta le dimensioni e il volume potrebbe riempirsi.

Il thin provisioning è vivamente consigliato in quanto consente di semplificare la gestione, offrendo al contempo un sostanziale miglioramento della capacità utilizzabile con conseguenti risparmi sui costi. Il motivo è semplice: Gli ambienti di database includono spesso molto spazio vuoto, un elevato numero di volumi e LUN e dati comprimibili. Il thick provisioning crea la riserva di spazio sullo storage per volumi e LUN, nel caso in cui un giorno raggiungano il 100% di riempimento e contengano dati non comprimibili al 100%. È improbabile che ciò accada mai. Il thin provisioning consente di recuperare lo spazio e di utilizzarlo altrove e consente la gestione della capacità basata sul sistema storage stesso piuttosto che su molti volumi e LUN più piccoli.

Alcuni clienti preferiscono utilizzare il thick provisioning, per carichi di lavoro specifici o generalmente basato su pratiche operative e di approvvigionamento consolidate.

> Se un volume viene sottoposto a thick provisioning, è necessario fare attenzione a disattivare completamente tutte le funzionalità di efficienza per quel volume, inclusa la decompressione e la rimozione della deduplica tramite il `sis undo` comando. Il volume non dovrebbe comparire nell' `volume efficiency show`output. In tal caso, il volume è ancora parzialmente configurato per le funzioni di efficienza. Di conseguenza, la sovrascrittura garantisce un funzionamento diverso, aumentando le possibilità che le sovrascritture causino l'esaurimento inaspettato dello spazio del volume, con conseguenti errori di i/o del database.

Best practice di efficienza

NetApp consiglia quanto segue:

Valori predefiniti AFF

I volumi creati su ONTAP in esecuzione su un sistema AFF all-flash vengono sottoposti a thin provisioning con tutte le funzionalità di efficienza inline abilitate. Sebbene in genere i database non beneficino della deduplica e possano includere dati non comprimibili, le impostazioni predefinite sono comunque appropriate per quasi tutti i carichi di lavoro. ONTAP è progettato per elaborare in modo efficiente tutti i tipi di dati e gli schemi i/o, indipendentemente dal fatto che comportino risparmi. Le impostazioni predefinite devono essere modificate solo se le ragioni sono pienamente comprese e se vi è un vantaggio a deviare.

Raccomandazioni generali

- Se i volumi e/o le LUN non sono dotati di thin provisioning, è necessario disabilitare tutte le impostazioni di efficienza perché queste funzionalità non offrono risparmi e la combinazione del thick provisioning con l'efficienza dello spazio può causare comportamenti imprevisti, inclusi errori di spazio esaurito.
- Se i dati non sono soggetti a sovrascritture, ad esempio con i backup o i log delle transazioni dei database, puoi ottenere una maggiore efficienza abilitando TSSE con un periodo di raffreddamento ridotto.
- Alcuni file potrebbero contenere una quantità significativa di dati non comprimibili, ad esempio quando la compressione è già abilitata a livello di applicazione dei file sono crittografati. Se uno di questi scenari è vero, considerare la possibilità di disattivare la compressione per consentire un funzionamento più efficiente su altri volumi che contengono dati comprimibili.
- Non utilizzare sia la compressione 32KB che la deduplica con i backup del database. Vedere la sezione Compressione adattiva per ulteriori informazioni.

Compressione dei database

SQL Server dispone inoltre di funzionalità per comprimere e gestire in modo efficiente i dati. Attualmente SQL Server supporta due tipi di compressione dati: Compressione riga e compressione pagina.

La compressione riga modifica il formato di memorizzazione dei dati. Ad esempio, cambia interi e decimali nel formato a lunghezza variabile invece del formato a lunghezza fissa nativo. Inoltre, le stringhe di caratteri a lunghezza fissa vengono modificate nel formato a lunghezza variabile eliminando gli spazi vuoti. La compressione della pagina implementa la compressione della riga e altre due strategie di compressione (compressione del prefisso e compressione del dizionario). Per ulteriori dettagli sulla compressione delle pagine, consultare "Implementazione della compressione pagina".

La compressione dei dati è attualmente supportata nelle edizioni Enterprise, Developer e Evaluation di SQL Server 2008 e versioni successive. Sebbene la compressione possa essere eseguita dal database stesso, ciò si verifica raramente in un ambiente SQL Server.

Di seguito sono riportati i suggerimenti per la gestione dello spazio per i file di dati di SQL Server

- Utilizzo del thin provisioning negli ambienti SQL Server per migliorare l'utilizzo dello spazio e ridurre i requisiti generali di storage quando viene utilizzata la funzionalità di garanzia di spazio.
  - Utilizza l'espansione automatica per la maggior parte delle configurazioni di implementazione più comuni, perché l'amministratore dello storage deve solo monitorare l'utilizzo dello spazio nell'aggregato.
- Non abilitare la deduplica su alcun volume su FAS contenente file di dati di SQL Server a meno che il volume non contenga più copie degli stessi dati, come ad esempio il ripristino del database da un backup su un singolo volume.

Bonifica dello spazio

Il recupero di spazio può essere avviato periodicamente per recuperare spazio inutilizzato in un LUN. Con SnapCenter, puoi usare il seguente comando PowerShell per iniziare il recupero dello spazio.

```
Invoke-SdHostVolumeSpaceReclaim -Path drive_path
```

Se è necessario eseguire il recupero di spazio, questo processo deve essere eseguito durante i periodi di attività bassa, poiché inizialmente consuma cicli sull'host.

Protezione dei dati

Le strategie di backup dei database devono basarsi su requisiti aziendali identificati, non su capacità teoriche. Combinando la tecnologia Snapshot di ONTAP e sfruttando le API di Microsoft SQL Server, è possibile eseguire rapidamente un backup coerente delle applicazioni indipendentemente dalle dimensioni dei database utente. Per requisiti di gestione dei dati più avanzati o scale-out, NetApp offre SnapCenter.

SnapCenter

SnapCenter è il software di data Protection di NetApp per le applicazioni aziendali. I database di SQL Server possono essere protetti in modo rapido e semplice con il plug-in SnapCenter per SQL Server e con operazioni del sistema operativo gestite dal plug-in SnapCenter per Microsoft Windows.

L'istanza di SQL Server può essere un'installazione autonoma, un'istanza cluster di failover o può essere un gruppo di disponibilità sempre attivo. Il risultato è che, grazie a un singolo pannello di controllo, i database possono essere protetti, clonati e ripristinati da una copia primaria o secondaria. SnapCenter può gestire database SQL Server sia on-premise, nel cloud che in configurazioni ibride. Le copie dei database possono essere create in pochi minuti sull'host originale o alternativo per lo sviluppo o per il reporting.

SQL Server richiede inoltre un coordinamento tra il sistema operativo e lo storage per garantire che i dati corretti siano presenti negli snapshot al momento della creazione. Nella maggior parte dei casi, l'unico metodo sicuro per eseguire questa operazione è SnapCenter o T-SQL. Gli snapshot creati senza questo coordinamento aggiuntivo potrebbero non essere recuperabili in modo affidabile.

Per ulteriori informazioni sul plug-in di SQL Server per SnapCenter, vedere "TR-4714: Guida alle Best practice per SQL Server con NetApp SnapCenter".

Protezione del database mediante snapshot T-SQL

In SQL Server 2022, Microsoft ha introdotto le istantanee T-SQL che offrono un percorso per la creazione di script e l'automazione delle operazioni di backup. Invece di eseguire copie di dimensioni normali, è possibile preparare il database per le snapshot. Una volta che il database è pronto per il backup, è possibile sfruttare le API REST di ONTAP per creare snapshot.

Di seguito è riportato un esempio di flusso di lavoro di backup:

1. Bloccare un database con il comando ALTER. In questo modo il database viene preparato per uno snapshot coerente sullo storage sottostante. Dopo il blocco è possibile scongelare il database e registrare lo snapshot con il comando di BACKUP.
2. Eseguire snapshot di più database sui volumi di storage contemporaneamente con il nuovo GRUPPO DI BACKUP e i comandi DEL SERVER DI BACKUP.
3. Eseguire backup COMPLETI o backup COMPLETI COPY_ONLY. Anche questi backup sono registrati in msdb.
4. Eseguire il recovery point-in-time utilizzando i backup di log eseguiti con il normale approccio di streaming dopo il backup COMPLETO delle snapshot. Se lo si desidera, sono supportati anche i backup differenziali in streaming.

Per ulteriori informazioni, vedere "Documentazione Microsoft per conoscere le istantanee T-SQL".

NetApp recommended Using SnapCenter to create Snapshot copy. Anche il metodo T-SQL descritto sopra funziona, ma SnapCenter offre un'automazione completa sul processo di backup, ripristino e cloning. Esegue inoltre il rilevamento per garantire che vengano creati gli snapshot corretti. Non è necessaria alcuna pre-configurazione.

Gruppo di disponibilità di SQL Server con SnapCenter

SnapCenter supporta il backup del database del gruppo di disponibilità SQL Server configurato con il cluster di failover Windows.

Il plugin SnapCenter per Microsoft SQL Server deve essere installato su tutti i nodi del cluster di failover del server Windows. Consultare i "documentazione"prerequisiti di ON e la procedura per configurare i plug-in di SnapCenter.

SnapCenter rileva che tutti i database, le istanze e i gruppi di disponibilità negli host e nelle risorse di Windows sono elencati nella pagina delle risorse di SnapCenter.

Protezione dei database nel gruppo di disponibilità always-on

I database del gruppo di disponibilità possono essere protetti in più modi.

- Backup a livello di database: Selezionare il database di disponibilità per la pagina delle risorse del database, aggiungere il criterio costituito dal backup completo/registro, pianificare il backup. SnapCenter utilizza il backup indipendentemente dal ruolo del database, sia che si tratti di una replica primaria che di una replica secondaria. La protezione può anche essere configurata aggiungendo database al gruppo di risorse.
- Backup a livello di istanza: Selezionare l'istanza per proteggere tutti i database in esecuzione sull'istanza in base al criterio selezionato. Viene eseguito il backup di tutti i database, incluso il database di disponibilità in esecuzione come replica primaria o secondaria, utilizzando SnapCenter. La protezione può anche essere configurata aggiungendo un'istanza al gruppo di risorse.

- Backup a livello di gruppo di disponibilità: Durante la configurazione del criterio, SnapCenter dispone di un'opzione avanzata per il backup a livello di gruppo di disponibilità. L'impostazione del gruppo di disponibilità nel criterio consente agli utenti di selezionare la preferenza di replica per il backup. È possibile selezionare la replica primaria, secondaria o tutte. L'opzione predefinita si basa sulla replica di backup impostata nella configurazione del gruppo di disponibilità di SQL Server.

L'impostazione del gruppo di disponibilità nel criterio SnapCenter verrà applicata solo se il backup a livello di gruppo di disponibilità viene utilizzato per proteggere i database del gruppo di disponibilità e non viene applicato per il backup a livello di database o istanza.

NetApp consiglia di utilizzare il backup a livello di disponibilità per il backup di tutta la replica in esecuzione sullo storage NetApp ONTAP.

Configurazione del backup del registro in SnapCenter

Se il gruppo di disponibilità è impostato su un'installazione autonoma di SQL Server, è necessario montare un disco dedicato su ciascun nodo di un cluster di failover del server Windows. Utilizzare un disco dedicato per configurare la directory del registro per salvare i backup del registro delle transazioni.

Se il gruppo di disponibilità è impostato sul cluster di failover di SQL Server, è necessario creare un disco cluster nell'istanza del cluster di failover di SQL Server per la directory del registro host.

Ripristino del database nella configurazione del gruppo di disponibilità con SnapCenter

- SnapCenter fornisce l'opzione di reseeding per ripristinare automaticamente il database dall'ultimo snapshot disponibile nella replica secondaria. L'operazione di risigillatura ripristinerà automaticamente e unirà il backup del database al gruppo di disponibilità.
- Un modo alternativo per ripristinare il database di replica nel gruppo di disponibilità consiste nell'interrompere il gruppo di disponibilità ed eseguire il ripristino completo e completo del registro. Utilizzare SnapCenter per ripristinare il database in modalità norecovery, quindi utilizzare SQL Server Management Studio o T-SQL per accedere nuovamente al gruppo di disponibilità del database.
- Per ripristinare solo il sottoinsieme di dati, è possibile utilizzare le funzionalità di cloning di SnapCenter per creare una copia clone del database. La copia del database viene creata in pochi minuti utilizzando SnapCenter, quindi esportare i dati nella replica primaria utilizzando gli strumenti nativi di SQL Server.

Per le procedure consigliate per configurare il layout dello storage del database in modo da soddisfare i requisiti RTO e RPO, vedere "TR-4714 Best practice per Microsoft SQL Server che utilizza NetApp SnapCenter".

SnapCenter non supporta il gruppo di disponibilità distribuita e il gruppo di disponibilità contenuta.

Disaster recovery

Disaster recovery

I database e le infrastrutture applicative aziendali spesso richiedono la replica per proteggersi da disastri naturali o interruzioni impreviste del business, con tempi di inattività minimi.

La funzionalità di replica del gruppo di disponibilità always-on di SQL Server può essere un'opzione eccellente e NetApp offre opzioni per integrare la protezione dei dati con Always-on. In alcuni casi, tuttavia, è consigliabile

prendere in considerazione la tecnologia di replica ONTAP. Sono disponibili tre opzioni di base.

SnapMirror

La tecnologia SnapMirror offre una soluzione aziendale rapida e flessibile per la replica dei dati su LAN e WAN. La tecnologia SnapMirror trasferisce solo i blocchi di dati modificati a destinazione dopo la creazione del mirror iniziale, riducendo in modo significativo i requisiti di larghezza di banda della rete. Può essere configurato in modalità sincrona o asincrona.

Sincronizzazione attiva di NetApp MetroCluster e SnapMirror

Per molti clienti, il disaster recovery non richiede solo una copia remota dei dati, ma anche la capacità di sfruttarli in maniera rapida. NetApp offre due tecnologie che soddisfano questa esigenza: MetroCluster e SnapMirror Active Sync

MetroCluster fa riferimento a ONTAP in una configurazione hardware che include storage con mirroring sincrono di basso livello e numerose funzionalità aggiuntive. Le soluzioni integrate come MetroCluster semplificano le complesse e scalabili infrastrutture di database, applicazioni e virtualizzazione. Sostituisce diversi prodotti e strategie di protezione dati esterni con un unico semplice storage array centrale. Fornisce inoltre backup, recovery, disaster recovery e alta disponibilità (ha) integrati in un singolo sistema storage in cluster.

SnapMirror Active Sync si basa su SnapMirror Synchronous. Con MetroCluster, ogni controller ONTAP è responsabile della replica dei dati dell'unità in una posizione remota. Con la sincronizzazione attiva di SnapMirror, avrai essenzialmente due sistemi ONTAP diversi che mantengono copie indipendenti dei dati LUN, ma cooperano per presentare una singola istanza di tale LUN. Dal punto di vista dell'host, si tratta di una singola entità LUN.

Confronto SM-AS e MCC

SM-AS e MetroCluster sono simili per quanto riguarda le funzionalità generali, ma esistono importanti differenze nel modo in cui è stata implementata la replica RPO=0 e nel modo in cui viene gestita. Anche se è possibile utilizzare la modalità asincrona e sincrona di SnapMirror come parte di un piano di disaster recovery, non sono progettate come tecnologie di replica ha.

- Una configurazione MetroCluster è più simile a un cluster integrato con nodi distribuiti tra i siti. SM-AS si comporta come due cluster altrimenti indipendenti che stanno cooperando nel servire un RPO selezionato=0 LUN replicati in modo sincrono.
- I dati in una configurazione MetroCluster sono accessibili solo da un determinato sito alla volta. Una seconda copia dei dati è presente sul sito opposto, ma i dati sono passivi. Non è possibile accedervi senza un failover del sistema storage.
- MetroCluster e SM-as eseguono il mirroring a diversi livelli. Il mirroring MetroCluster viene eseguito al livello RAID. I dati di basso livello sono memorizzati in un formato di mirroring utilizzando SyncMirror. L'utilizzo del mirroring è praticamente invisibile ai livelli di LUN, volume e protocollo.
- Al contrario, il mirroring SM-AS avviene a livello di protocollo. I due cluster sono complessivamente cluster indipendenti. Una volta sincronizzate le due copie di dati, i due cluster devono solo eseguire il mirroring delle scritture. Quando si verifica una scrittura su un cluster, questa viene replicata nell'altro cluster. La scrittura viene riconosciuta all'host solo quando la scrittura è stata completata su entrambi i siti. A parte questo comportamento di suddivisione del protocollo, i due cluster sono altrimenti normali cluster ONTAP.
- Il ruolo principale di MetroCluster è la replica su larga scala. Puoi replicare un intero array con RPO=0 e RTO prossimo allo zero. Questo semplifica il processo di failover perché esiste un solo "problema" da eseguire e consente una scalabilità perfetta in termini di capacità e IOPS.

- Un caso d'utilizzo chiave per SM-AS è la replica granulare. A volte non vuoi replicare tutti i dati come una singola unità oppure devi eseguire il failover selettivo su alcuni carichi di lavoro.
- Un altro caso d'utilizzo chiave per SM-AS è per operazioni Active-Active, dove desideri che siano disponibili copie dei dati completamente utilizzabili su due cluster diversi situati in due posizioni diverse con caratteristiche di performance identiche e, se desiderato, non richiedere l'estensione della SAN tra i siti. Le applicazioni possono essere già in esecuzione su entrambi i siti, riducendo così l'RTO complessivo durante le operazioni di failover.

SnapMirror

Di seguito sono riportati alcuni consigli su SnapMirror per SQL Server:

- In caso di utilizzo di SMB, la SVM di destinazione deve appartenere allo stesso dominio Active Directory del quale fa parte la SVM di origine, in modo da non interrompere le liste per il controllo degli accessi (ACL) archiviate nei file NAS durante il ripristino in caso di disastro.
- L'utilizzo di nomi di volumi di destinazione identici ai nomi di volumi di origine non è necessario, ma può semplificare la gestione del processo di montaggio dei volumi di destinazione nella destinazione. Se viene utilizzato SMB, occorre rendere identico il namespace NAS di destinazione nei percorsi e nella struttura delle directory al namespace di origine.
- Per motivi di coerenza, non pianificare gli update SnapMirror dai controller. Attiva invece gli update di SnapMirror da SnapCenter per aggiornare SnapMirror al termine del backup completo o del log.
- Distribuire volumi che contengono dati SQL Server tra diversi nodi nel cluster per consentire a tutti i nodi del cluster di condividere l'attività di replica di SnapMirror. Questa distribuzione ottimizza l'utilizzo delle risorse dei nodi.
- Utilizza la replica sincrona in cui la richiesta di un rapido data recovery è maggiore e soluzioni asincrone per la flessibilità negli RPO.

Per ulteriori informazioni su SnapMirror, vedere "TR-4015: Guida alle Best practice e alla configurazione di SnapMirror per ONTAP 9".

MetroCluster

Architettura

La distribuzione di Microsoft SQL Server con ambiente MetroCluster richiede alcune spiegazioni sulla progettazione fisica di un sistema MetroCluster.

MetroCluster esegue il mirroring sincrono di dati e configurazione tra due cluster ONTAP in posizioni separate o domini di errore. MetroCluster offre storage sempre disponibile per le applicazioni gestendo automaticamente due obiettivi:

- RPO (Zero Recovery Point Objective) eseguendo il mirroring sincrono dei dati scritti nel cluster.
- RTO (Recovery Time Objective) quasi nullo attraverso la configurazione del mirroring e l'automazione dell'accesso ai dati del secondo sito.

MetroCluster offre semplicità con mirroring automatico dei dati e configurazione tra i due cluster indipendenti situati nei due siti. Poiché lo storage viene fornito all'interno di un cluster, viene automaticamente eseguito il mirroring nel secondo cluster del secondo sito. NetApp SyncMirror® fornisce una copia completa di tutti i dati con RPO pari a zero. Ciò significa che i carichi di lavoro da un sito possono passare in qualsiasi momento al sito opposto e continuare a fornire i dati senza perdita di dati. MetroCluster gestisce il processo di switchover per fornire accesso ai dati forniti da NAS e SAN al secondo sito. La progettazione di MetroCluster come

soluzione validata contiene dimensionamento e configurazione che consente di eseguire uno switchover entro i periodi di timeout del protocollo (generalmente inferiori a 120 secondi). Questo consente un RPO prossimo allo zero e le applicazioni possono continuare ad accedere ai dati senza incorrere in guasti. Il MetroCluster è disponibile in diverse variazioni definite dal fabric dello storage back-end.

MetroCluster è disponibile in 3 diverse configurazioni

- Coppie HA con connettività IP
- Coppie HA con connettività FC
- Controller singolo con connettività FC

Il termine connettività si riferisce alla connessione cluster utilizzata per la replica tra siti. Non si riferisce ai protocolli host. Tutti i protocolli lato host sono supportati come di consueto in una configurazione MetroCluster indipendentemente dal tipo di connessione utilizzata per la comunicazione tra cluster.

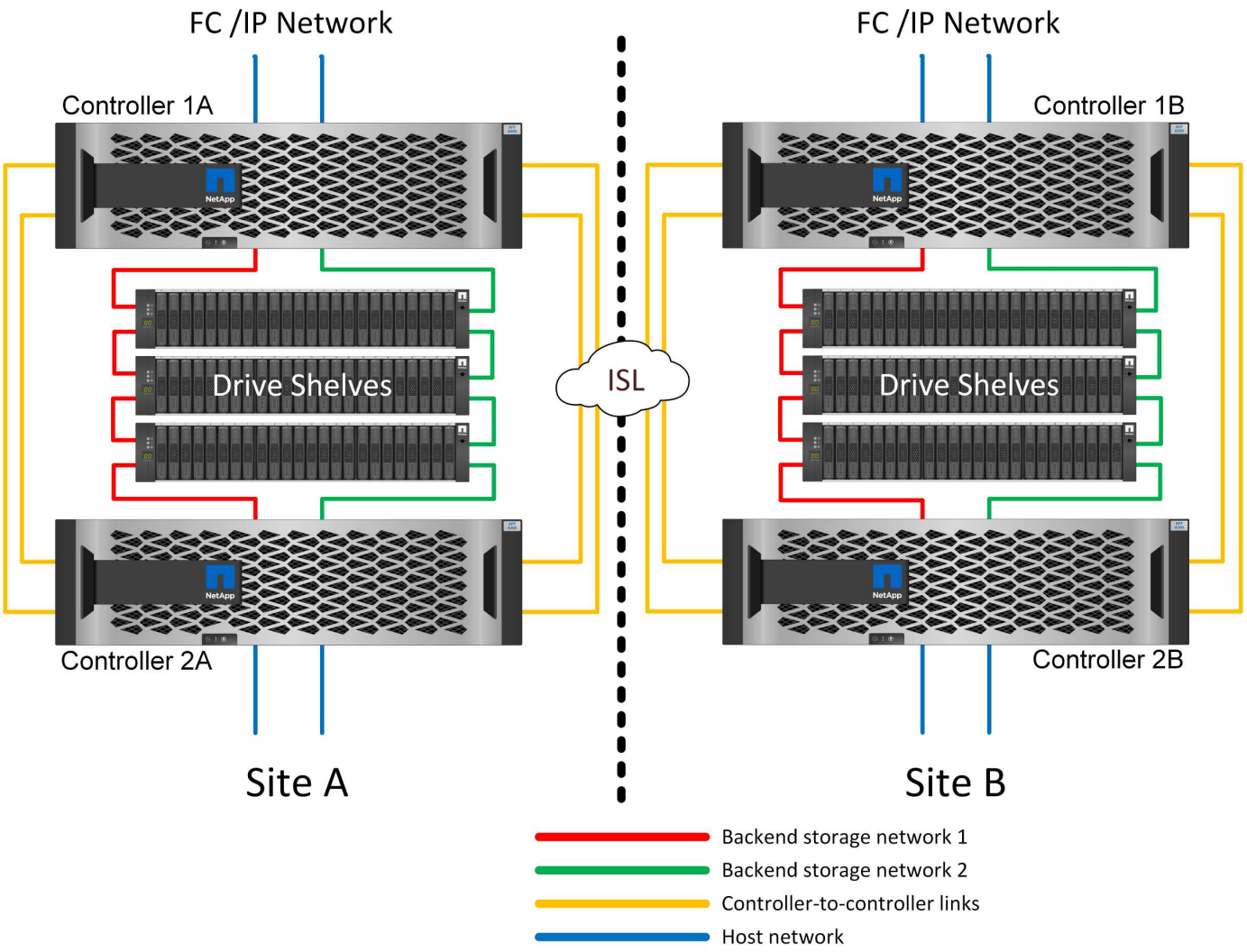
IP MetroCluster

La configurazione MetroCluster IP ha-Pair utilizza due o quattro nodi per sito. Questa opzione di configurazione aumenta la complessità e i costi rispetto all'opzione a due nodi, ma offre un vantaggio importante: La ridondanza intrasite. Un semplice errore del controller non richiede l'accesso ai dati nella WAN. L'accesso ai dati rimane locale attraverso il controller locale alternativo.

La maggior parte dei clienti sceglie la connettività IP perché i requisiti dell'infrastruttura sono più semplici. In passato, la connettività cross-site ad alta velocità era generalmente più semplice da fornire utilizzando gli switch FC e in fibra scura, ma oggi i circuiti IP ad alta velocità e a bassa latenza sono più prontamente disponibili.

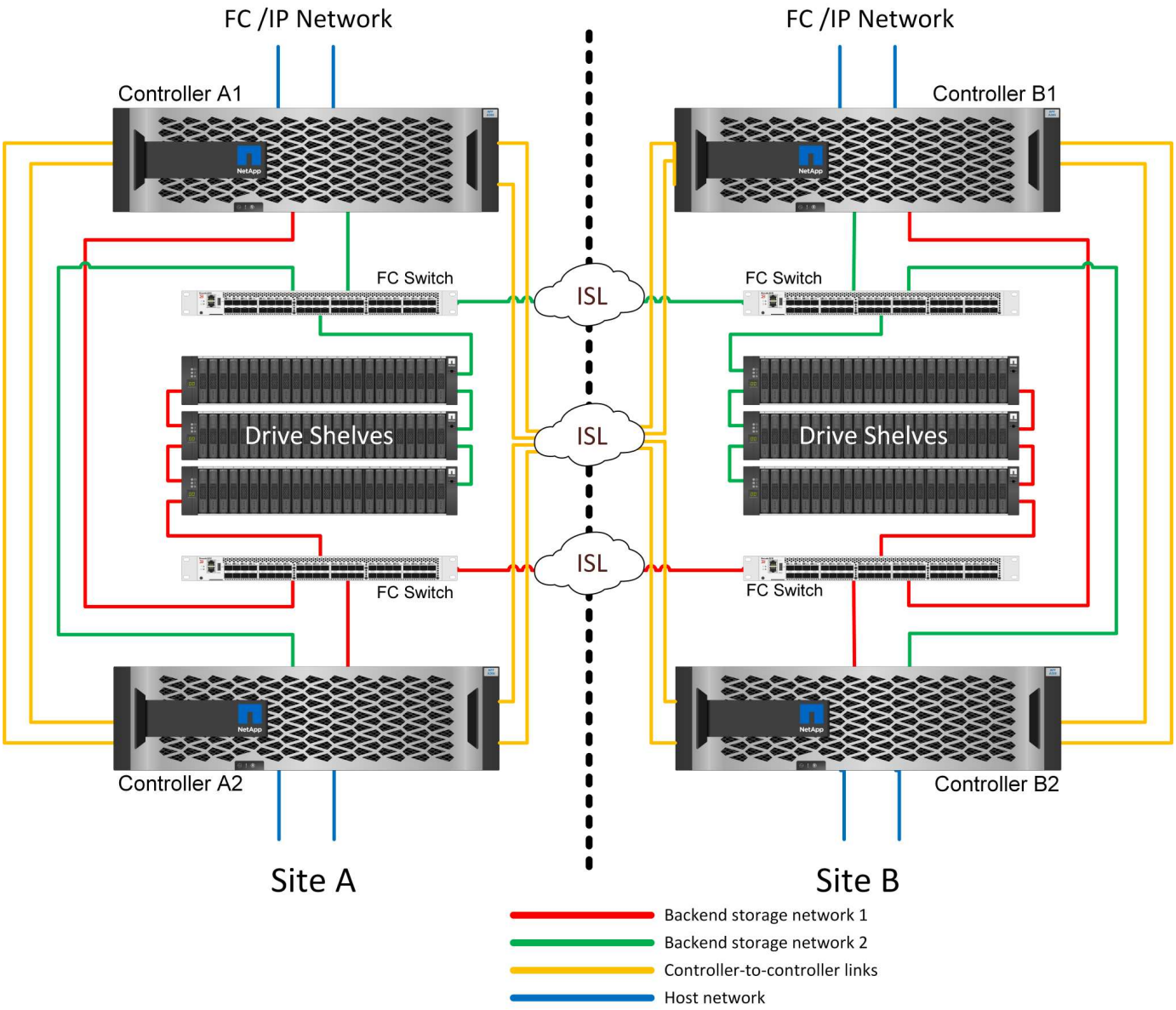
L'architettura è anche più semplice perché le uniche connessioni cross-site sono per i controller. Nei MetroClusters collegati a FC SAN, un controller scrive direttamente sulle unità del sito opposto e quindi richiede connessioni SAN, switch e bridge aggiuntivi. Al contrario, un controller in una configurazione IP scrive sulle unità opposte tramite il controller.

Per ulteriori informazioni, consultare la documentazione ufficiale di ONTAP e. "Architettura e progettazione della soluzione IP di MetroCluster".



MetroCluster HA-Pair FC SAN-Attached

La configurazione ha-Pair MetroCluster FC utilizza due o quattro nodi per sito. Questa opzione di configurazione aumenta la complessità e i costi rispetto all'opzione a due nodi, ma offre un vantaggio importante: La ridondanza intrasite. Un semplice errore del controller non richiede l'accesso ai dati nella WAN. L'accesso ai dati rimane locale attraverso il controller locale alternativo.

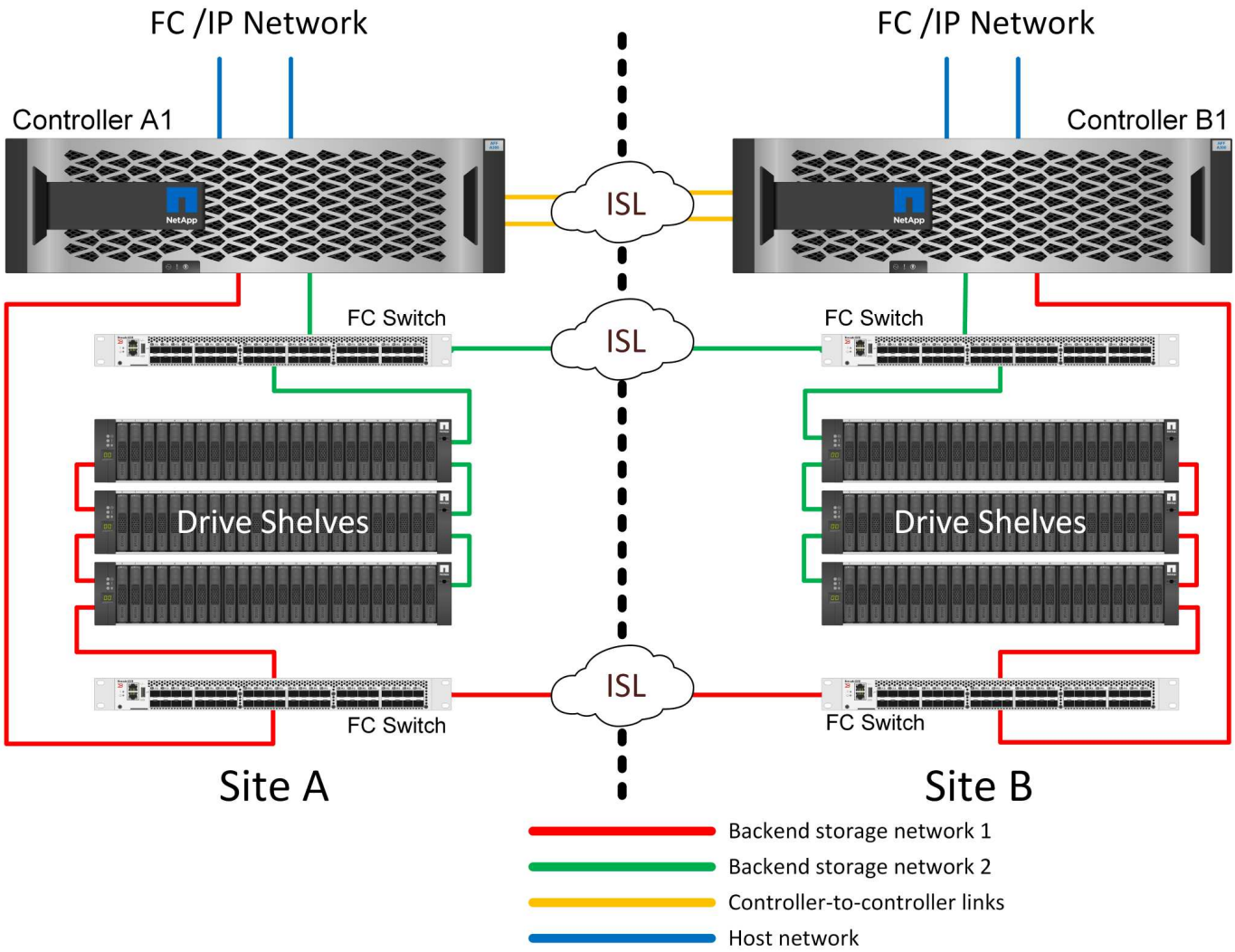


Alcune infrastrutture multisito non sono progettate per le operazioni Active-Active, ma vengono utilizzate maggiormente come sito primario e sito di disaster recovery. In questa situazione, è generalmente preferibile un'opzione ha-Pair MetroCluster per i seguenti motivi:

- Anche se un cluster MetroCluster a due nodi è un sistema ha, un guasto imprevisto di un controller o una manutenzione pianificata richiedono che i servizi dati vengano online sul sito opposto. Se la connettività di rete tra i siti non supporta la larghezza di banda richiesta, le prestazioni ne risentono. L'unica opzione sarebbe anche eseguire il failover dei vari sistemi operativi host e dei servizi associati al sito alternativo. Il cluster MetroCluster ha-Pair elimina questo problema grazie alla perdita di un controller che consente di eseguire un semplice failover all'interno dello stesso sito.
- Alcune topologie di rete non sono progettate per l'accesso tra siti, ma utilizzano sottoreti o SAN FC isolate. In questi casi, il cluster MetroCluster a due nodi non funziona più come sistema ha, perché il controller alternativo non può fornire dati ai server del sito opposto. L'opzione ha-Pair MetroCluster è necessaria per garantire ridondanza completa.
- Se un'infrastruttura a due siti viene vista come una singola infrastruttura ad alta disponibilità, la configurazione MetroCluster a due nodi è adatta. Tuttavia, se il sistema deve funzionare per un periodo di tempo prolungato dopo il guasto del sito, è preferibile una coppia ha perché continua a fornire ha all'interno di un singolo sito.

MetroCluster FC SAN-attached a due nodi

La configurazione MetroCluster a due nodi utilizza un solo nodo per sito. Questo design è più semplice rispetto all'opzione ha-Pair perché richiede meno componenti da configurare e gestire. Inoltre, ha ridotto le richieste di infrastruttura in termini di cablaggio e switch FC. Infine, riduce i costi.



L'evidente impatto di questa progettazione è che un guasto del controller su un singolo sito implica che i dati sono disponibili dal sito opposto. Questa restrizione non è necessariamente un problema. Molte aziende hanno operazioni di data center multisito con reti estese, ad alta velocità e a bassa latenza che funzionano essenzialmente come una singola infrastruttura. In questi casi, la configurazione preferita è la versione a due nodi di MetroCluster. Diversi service provider utilizzano attualmente sistemi a due nodi con scalabilità di petabyte.

Funzionalità di resilienza di MetroCluster

Non esistono single point of failure in una soluzione MetroCluster:

- Ogni controller dispone di due percorsi indipendenti verso gli shelf di dischi sul sito locale.
- Ogni controller dispone di due percorsi indipendenti verso gli shelf di dischi sul sito remoto.
- Ciascun controller dispone di due percorsi indipendenti verso i controller sul sito opposto.
- Nella configurazione ha-Pair, ogni controller ha due percorsi verso il partner locale.

Riassumendo, qualsiasi componente della configurazione può essere rimosso senza compromettere la capacità di MetroCluster di fornire dati. L'unica differenza in termini di resilienza tra le due opzioni è che la versione ha-Pair è ancora un sistema storage ha generale dopo un guasto del sito.

SyncMirror

La protezione per SQL Server con MetroCluster si basa su SyncMirror, che offre una tecnologia di mirroring sincrono scale-out dalle performance massime.

Data Protection con SyncMirror

Al livello più semplice, la replica sincrona significa che qualsiasi modifica deve essere apportata a entrambi i lati dello storage con mirroring prima che venga riconosciuta. Ad esempio, se un database sta scrivendo un registro o un guest VMware viene aggiornato, non deve mai andare persa una scrittura. Come livello di protocollo, il sistema di storage non deve riconoscere la scrittura fino a quando non è stato assegnato a un supporto non volatile in entrambi i siti. Solo allora è sicuro procedere senza il rischio di perdita dei dati.

L'utilizzo di una tecnologia di replica sincrona è il primo passo nella progettazione e nella gestione di una soluzione di replica sincrona. La considerazione più importante è capire cosa potrebbe accadere durante i vari scenari di guasto pianificati e non pianificati. Non tutte le soluzioni di replica sincrona offrono le stesse funzionalità. Se hai bisogno di una soluzione che offra un recovery point objective (RPO) pari a zero, ovvero zero data loss, devi prendere in considerazione tutti gli scenari di guasto. In particolare, qual è il risultato previsto quando la replica è impossibile a causa della perdita di connettività tra i siti?

Disponibilità dei dati SyncMirror

La replica MetroCluster si basa sulla tecnologia NetApp SyncMirror, che è progettata per passare in modo efficiente dalla modalità sincrona alla modalità sincrona e viceversa. Questa funzionalità soddisfa i requisiti dei clienti che richiedono una replica sincrona, ma che hanno bisogno anche di un'alta disponibilità per i propri servizi dati. Ad esempio, se la connettività a un sito remoto viene interrotta, è generalmente preferibile che il sistema di archiviazione continui a funzionare in uno stato non replicato.

Molte soluzioni di replica sincrona sono in grado di funzionare solo in modalità sincrona. Questo tipo di replica "tutto o niente" viene talvolta chiamato modalità domino. Tali sistemi storage smettono di fornire i dati piuttosto che permettere che le copie locali e remote dei dati diventino non sincronizzate. Se la replica viene forzata, la risincronizzazione può richiedere molto tempo e lasciare un cliente esposto a una perdita di dati completa durante il tempo in cui il mirroring viene ristabilita.

Non solo SyncMirror può passare alla modalità sincrona senza problemi se il sito remoto non è raggiungibile, ma può anche risincronizzare rapidamente uno stato RPO = 0 al ripristino della connettività. La copia obsoleta dei dati nel sito remoto può anche essere preservata in uno stato utilizzabile durante la risincronizzazione, garantendo l'esistenza in ogni momento di copie locali e remote dei dati.

Quando è richiesta la modalità domino, NetApp offre SnapMirror Synchronous (SM-S). Esistono anche opzioni a livello di applicazione, come Oracle DataGuard o SQL Server Always on Availability Groups. Il mirroring del disco a livello del sistema operativo può essere opzionale. Per ulteriori informazioni e opzioni, consulta il tuo NetApp o il partner account team.

SQL Server con MetroCluster

Un'opzione per la protezione dei database di SQL Server con RPO pari a zero è MetroCluster. MetroCluster è una semplice tecnologia di replica RPO=0 dalle performance elevate che consente di replicare un'intera infrastruttura tra i siti in modo semplice.

SQL Server è in grado di scalare fino a migliaia di database su un unico sistema MetroCluster. Potrebbero esistere istanze standalone di SQL Server o istanze cluster di failover, MetroCluster System non aggiunge o modifica necessariamente le Best practice per la gestione di un database.

Una spiegazione completa di MetroCluster esula dall'ambito di questo documento, ma i principi sono semplici. MetroCluster può offrire una soluzione di replica con RPO=0 e failover rapido. Ciò che costruite su questa base dipende dalle vostre esigenze.

Ad esempio, una procedura di DR rapida di base dopo una perdita improvvisa del sito potrebbe utilizzare la seguente procedura di base:

- Forzare uno switchover MetroCluster
- Rilevamento di LUN FC/iSCSI (solo SAN)
- Montare i file system
- Avviare SQL Services

Il requisito principale di questo approccio è rappresentato da un sistema operativo in esecuzione sul sito remoto. Deve essere preconfigurato con l'installazione di SQL Server e deve essere aggiornato con una versione di build equivalente. È inoltre possibile eseguire il mirroring dei database di sistema di SQL Server nel sito remoto e montarli se viene dichiarata un'emergenza.

Se i volumi, i file system e il datastore che ospitano database virtualizzati non vengono utilizzati nel sito di disaster recovery prima dello switchover, non è necessario impostare i `dr-force-nvfail` volumi associati.

Sincronizzazione attiva di SnapMirror

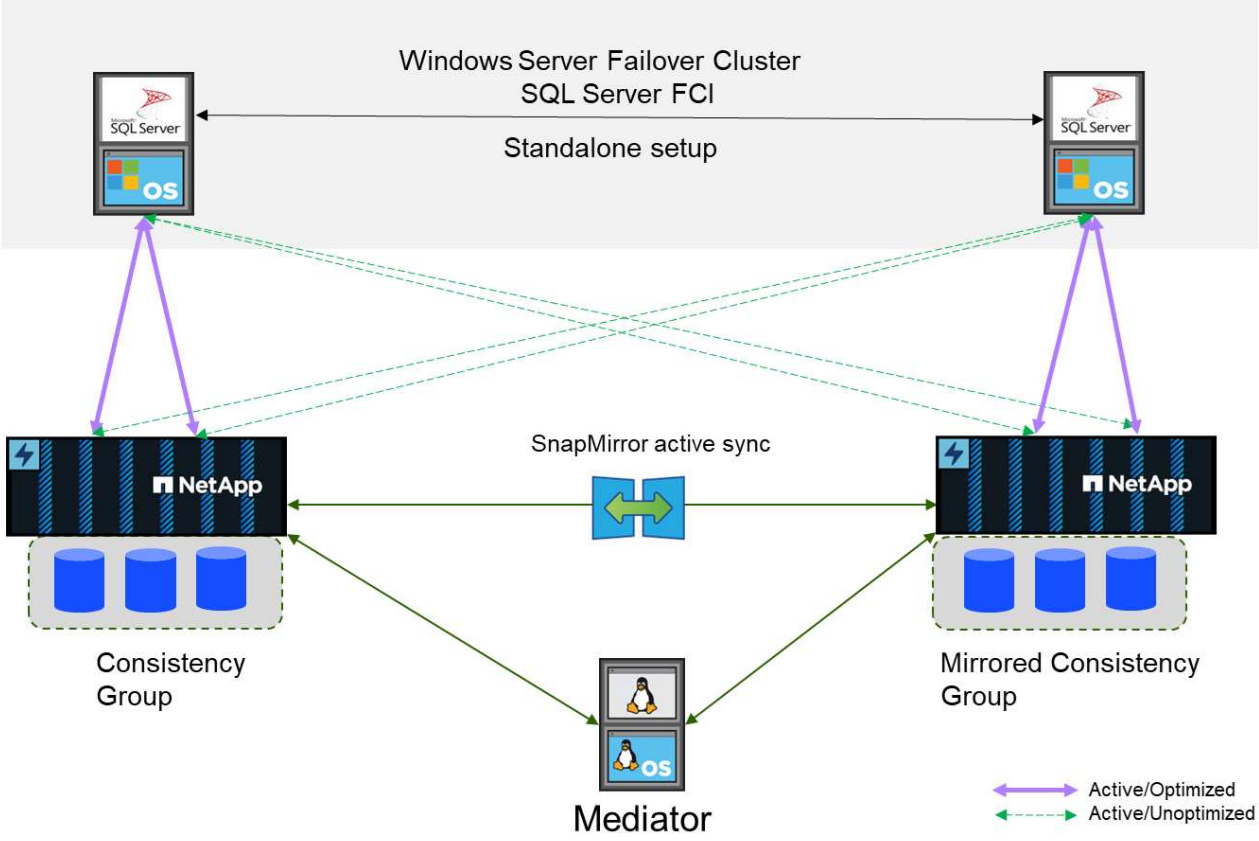
Panoramica

La sincronizzazione attiva di SnapMirror consente a singoli database e applicazioni SQL Server di continuare le operazioni durante interruzioni di rete e storage, con failover trasparente dello storage senza interventi manuali.

A partire da ONTAP 9.15,1, SnapMirror Active Sync supporta l'architettura Active/Active simmetrica oltre alla configurazione asimmetrica esistente. La funzionalità Active/Active simmetrica offre replica bidirezionale sincrona per la business continuity e il disaster recovery. Protegge l'accesso ai dati per carichi di lavoro critici SAN con accesso simultaneo in lettura e scrittura ai dati in diversi domini di errore, garantendo operazioni senza interruzioni e riducendo al minimo i downtime in caso di disastri o errori di sistema.

Gli host SQL Server accedono allo storage utilizzando LUN Fiber Channel (FC) o iSCSI. Replica tra ciascun cluster in cui risiede una copia dei dati replicati. Poiché questa funzionalità è la replica a livello di storage, le istanze di SQL Server in esecuzione su istanze di cluster di failover o host standalone possono eseguire operazioni di lettura/scrittura in entrambi i cluster. Per le fasi di pianificazione e configurazione, fare riferimento a "Documentazione di ONTAP sulla sincronizzazione attiva di SnapMirror" .

Architettura di sincronizzazione attiva SnapMirror con Active/Active simmetrico



Replica sincrona

Durante l'utilizzo normale, ciascuna copia è una replica sincrona con RPO=0/7, con un'unica eccezione. Se i dati non possono essere replicati, ONTAP rilascerà il requisito di replicare i dati e riprendere la distribuzione io su un sito mentre le LUN dell'altro sito vengono portate offline.

Hardware di archiviazione

A differenza di altre soluzioni di disaster recovery per lo storage, SnapMirror Active Sync offre una flessibilità asimmetrica della piattaforma. Non è necessario che l'hardware di ciascun sito sia identico. Questa funzionalità consente di dimensionare correttamente l'hardware utilizzato per supportare la sincronizzazione attiva di SnapMirror. Il sistema di storage remoto può essere identico al sito primario se deve supportare un carico di lavoro di produzione completo, ma se un disastro determina una riduzione dell'i/o, rispetto a un sistema più piccolo nel sito remoto potrebbe risultare più conveniente.

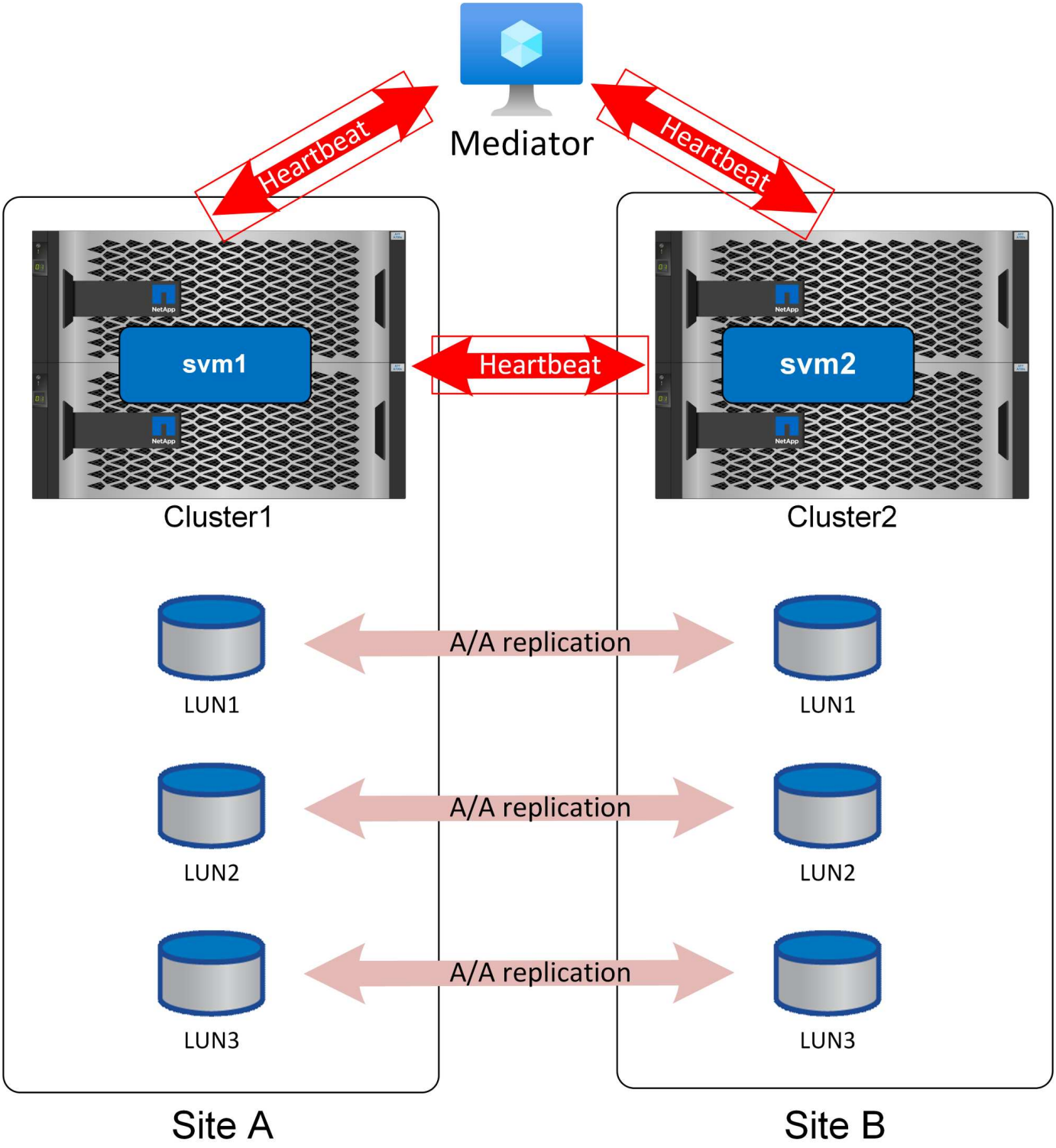
ONTAP mediatore

ONTAP Mediator è un'applicazione software che viene scaricata dal supporto NetApp e che viene in genere distribuita su una piccola macchina virtuale. Il ONTAP Mediator non è un tiebreaker. È un canale di comunicazione alternativo per i due cluster che partecipano alla replica sincrona attiva SnapMirror. Le operazioni automatizzate sono gestite da ONTAP in base alle risposte ricevute dal partner tramite connessioni dirette e tramite il mediatore.

Mediatore ONTAP

Mediator è necessario per automatizzare in modo sicuro il failover. Idealmente, sarebbe posizionato su un sito 3rd indipendente, ma può comunque funzionare per la maggior parte delle esigenze se colocated con uno dei cluster che partecipano alla replica.

Il mediatore non è realmente un tiebreaker, anche se quella è effettivamente la funzione che fornisce. Non esegue alcuna azione, ma fornisce un canale di comunicazione alternativo per la comunicazione tra cluster e cluster.



La sfida #1 con il failover automatizzato è il problema split-brain, e questo problema sorge se i due siti perdono la connettività tra loro. Che cosa dovrebbe accadere? Non si desidera che due siti diversi si designino come copie sopravvissute dei dati, ma in che modo un singolo sito può distinguere tra la perdita effettiva del sito opposto e l'impossibilità di comunicare con il sito opposto?

Qui entra il mediatore nell'immagine. Se si trova in un sito 3rd e ciascun sito dispone di una connessione di rete separata per tale sito, è disponibile un percorso aggiuntivo per ciascun sito per convalidare lo stato dell'altro. Esaminare nuovamente l'immagine sopra riportata e considerare i seguenti scenari.

- Cosa succede se il mediatore non riesce o è irraggiungibile da uno o entrambi i siti?
  - I due cluster possono ancora comunicare tra loro sullo stesso link utilizzato per i servizi di replica.
  - I dati sono ancora serviti con protezione RPO=0/7
- Cosa succede se il sito A non funziona?
  - Il sito B vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito B sostituirà i servizi dati, ma senza RPO = mirroring 0:1
- Cosa succede se il sito B non funziona?
  - Il sito A vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito A sostituirà i servizi dati, ma senza RPO = mirroring 0:1

Esiste un altro scenario da considerare: La perdita del collegamento di replica dei dati. In caso di perdita del link di replica tra i siti, RPO=0 Mirroring sarà ovviamente impossibile. Che cosa dovrebbe accadere allora?

Questo è controllato dallo stato del sito preferito. In una relazione SM-AS, uno dei siti è secondario all'altro. Questo non ha alcun effetto sulle normali operazioni e tutto l'accesso ai dati è simmetrico, ma se la replica viene interrotta, il legame dovrà essere interrotto per riprendere le operazioni. Ne risulta che il sito preferito continuerà le operazioni senza mirroring e il sito secondario interromperà l'elaborazione io fino al ripristino della comunicazione di replica.

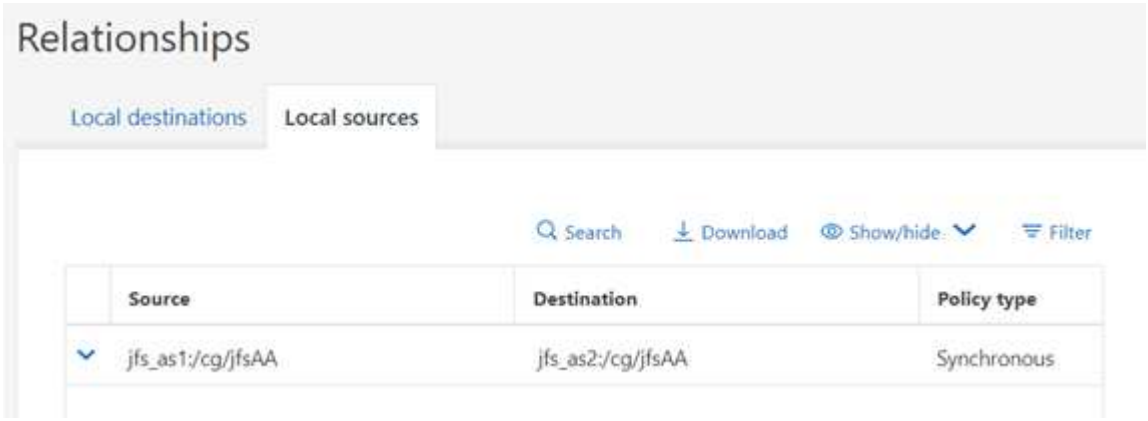
Sito preferito

Il comportamento di sincronizzazione attiva di SnapMirror è simmetrico, con una eccezione importante: Configurazione del sito preferita.

La sincronizzazione attiva di SnapMirror considererà un sito la "fonte" e l'altro la "destinazione". Ciò implica una relazione di replica unidirezionale, ma ciò non si applica al comportamento io. La replica è bidirezionale e simmetrica, mentre i tempi di risposta io sono identici su entrambi i lati del mirror.

La `source` designazione è controlla il sito preferito. In caso di perdita del link di replica, i percorsi delle LUN nella copia di origine continueranno a fornire i dati mentre i percorsi delle LUN nella copia di destinazione non saranno disponibili finché la replica non viene ristabilita e SnapMirror ritorna allo stato sincrono. I percorsi riprenderanno a fornire i dati.

La configurazione di origine/destinazione può essere visualizzata tramite SystemManager:



O all’interfaccia CLI:

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                      Destination Path: jfs_as2:/cg/jfsAA
                     Relationship Type: XDP
                Relationship Group Type: consistencygroup
                    SnapMirror Schedule: -
                 SnapMirror Policy Type: automated-failover-duplex
                      SnapMirror Policy: AutomatedFailOverDuplex
                            Tries Limit: -
                      Throttle (KB/sec): -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

La chiave è che source è l’SVM su cluster1. Come menzionato sopra, i termini "origine" e "destinazione" non descrivono il flusso di dati replicati. Entrambi i siti possono elaborare una scrittura e replicarla nel sito opposto. In effetti, entrambi i cluster sono origini e destinazioni. L’effetto della designazione di un cluster come origine controlla semplicemente quale cluster sopravvive come sistema di storage in lettura e scrittura in caso di perdita del link di replica.

Topologia di rete

Accesso uniforme

Una rete ad accesso uniforme significa che gli host sono in grado di accedere ai percorsi su entrambi i siti (o ai domini di errore all’interno dello stesso sito).

Una caratteristica importante di SM-AS è la capacità di configurare i sistemi storage per sapere dove si trovano gli host. Quando si mappano i LUN a un determinato host, è possibile indicare se sono prossimali o meno a un determinato sistema di archiviazione.

Impostazioni di prossimità

La prossimità si riferisce a una configurazione per cluster che indica che un determinato WWN host o ID iniziatore iSCSI appartiene a un host locale. Si tratta di un secondo passo opzionale per la configurazione dell’accesso LUN.

Il primo passo è la normale configurazione di igroup. Ogni LUN deve essere mappato a un igroup che contiene gli ID WWN/iSCSI degli host che devono accedere a quel LUN. Controlla quale host ha *accesso* a un LUN.

Il secondo passo, opzionale, consiste nel configurare la prossimità all’host. Questo non controlla l’accesso, controlla *priority*.

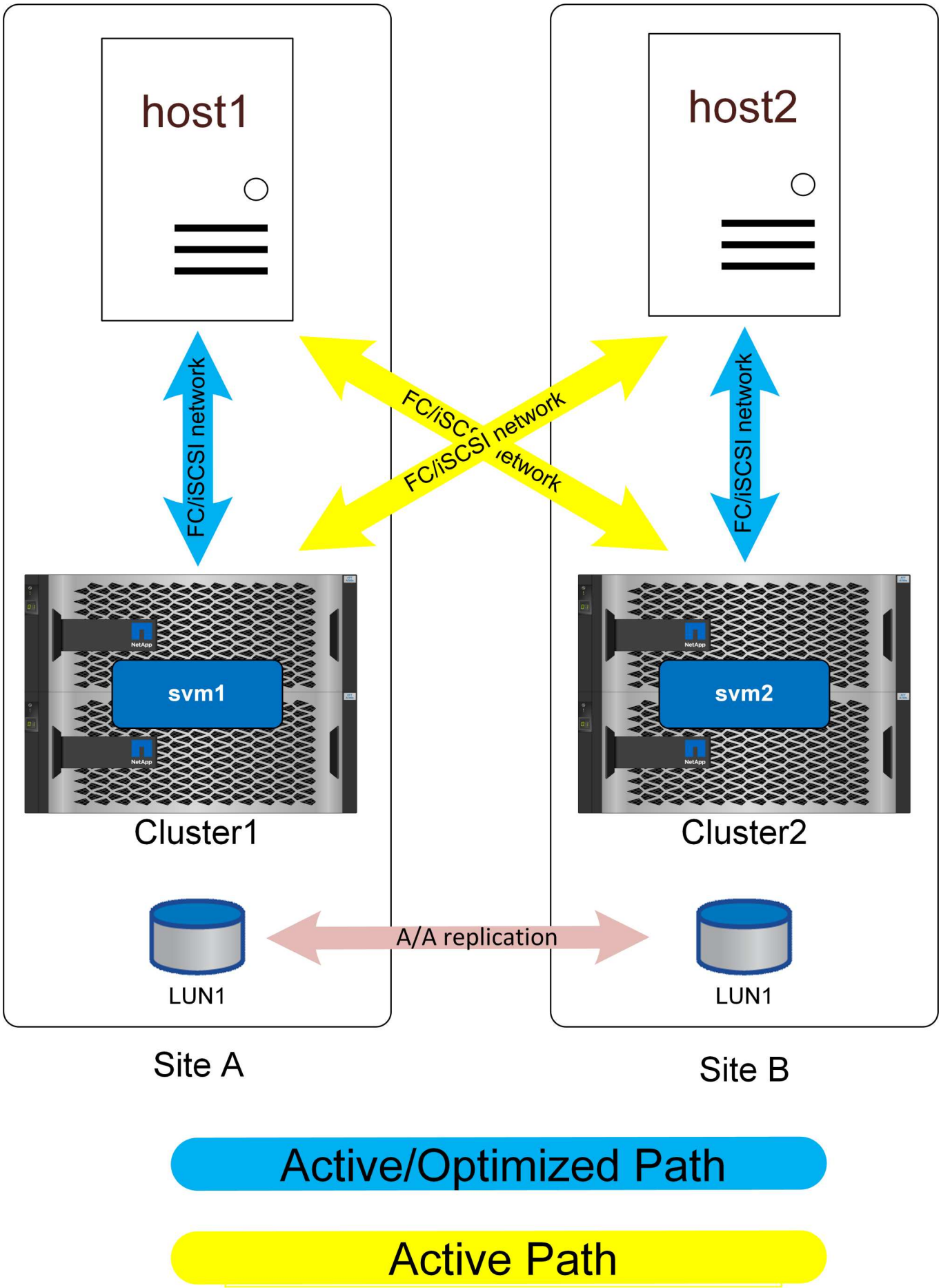
Ad esempio, un host del sito A potrebbe essere configurato in modo da accedere a una LUN protetta dalla sincronizzazione attiva di SnapMirror e, poiché la SAN è estesa tra i siti, i percorsi sono disponibili per tale LUN utilizzando lo storage sul sito A o lo storage sul sito B.

Senza impostazioni di prossimità, l’host utilizzerà entrambi i sistemi storage allo stesso modo perché entrambi

i sistemi storage pubblicizzeranno i percorsi attivi/ottimizzati. Se la latenza SAN e/o la larghezza di banda tra i siti è limitata, questa operazione potrebbe non essere disattivabile e potrebbe essere necessario assicurarsi che durante il normale funzionamento ogni host utilizzi preferenzialmente i percorsi verso il sistema di storage locale. Viene configurato aggiungendo l'ID WWN/iSCSI dell'host al cluster locale come host prossimale. Questa operazione può essere eseguita dalla CLI o da SystemManager.

AFF

Con un sistema AFF, i percorsi vengono visualizzati come mostrato di seguito quando è stata configurata la prossimità dell'host.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
svm2
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

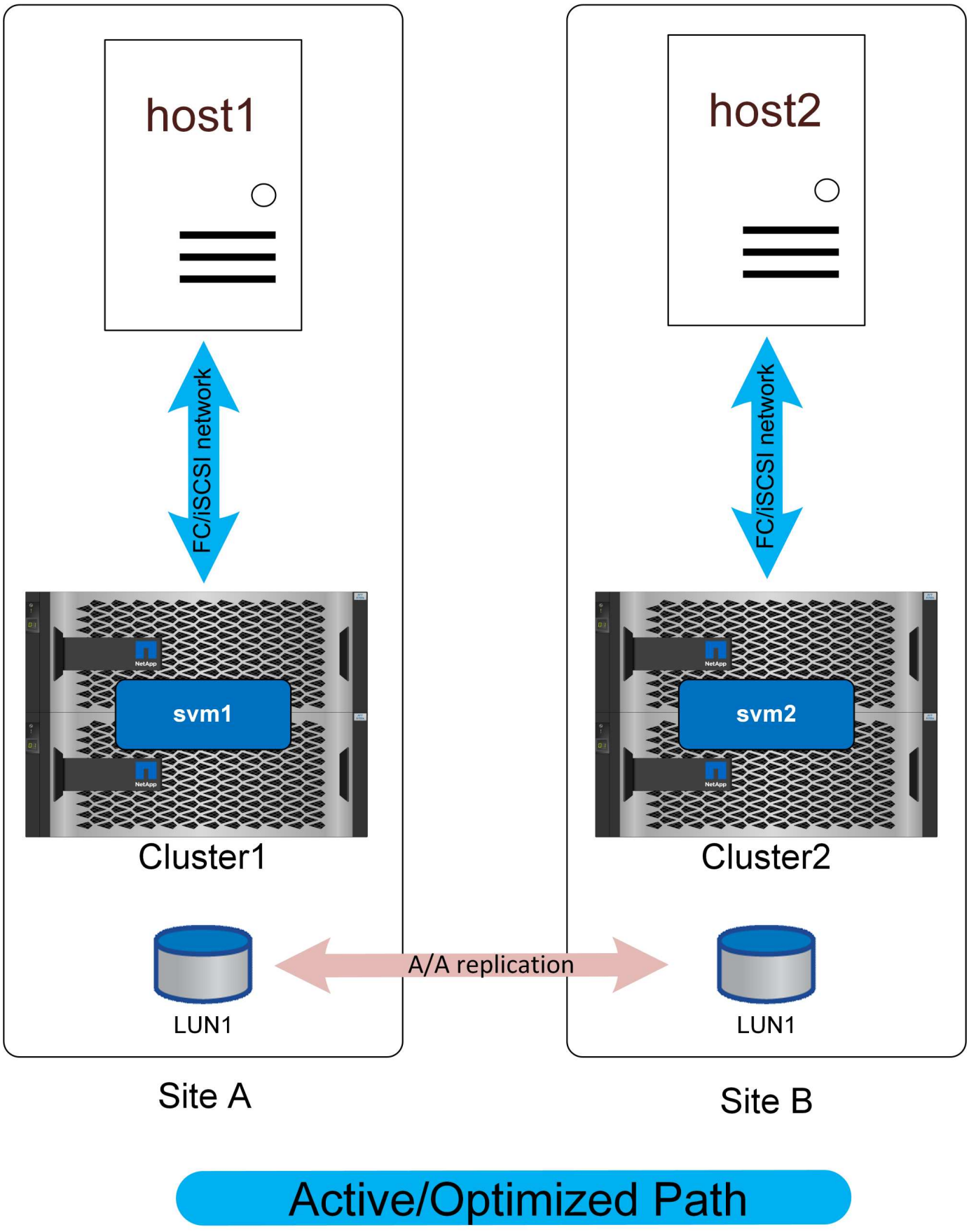
Durante le normali operazioni, tutti i io sono i/o locali. Le letture e le scritture sono gestite dallo storage array locale. Gli io in scrittura, naturalmente, dovranno anche essere replicati dal controller locale sul sistema remoto prima di essere riconosciuti, ma tutti gli io in lettura saranno serviti localmente e non subiranno una latenza aggiuntiva attraverso il collegamento SAN tra i siti.

L'unica volta in cui verranno utilizzati i percorsi non ottimizzati è quando tutti i percorsi attivi/ottimizzati vengono persi. Ad esempio, se l'intero array sul sito A perde energia, gli host sul sito A sarebbero comunque in grado di accedere ai percorsi dell'array sul sito B e di rimanere quindi operativi, anche se sperimenterebbero una latenza più elevata.

Esistono percorsi ridondanti attraverso il cluster locale non mostrati in questi diagrammi per semplicità. I sistemi di storage ONTAP sono ad alta disponibilità, pertanto un guasto a un controller non dovrebbe causare guasti nel sito. Ciò dovrebbe comportare solo una modifica dei percorsi locali utilizzati nel sito interessato.

Accesso non uniforme

Una rete di accesso non uniforme significa che ogni host ha solo accesso alle porte sul sistema di storage locale. LA SAN non è estesa a più siti (o presenta errori nei domini dello stesso sito).



Il vantaggio principale di questo approccio è la semplicità DELLE SAN, eliminando l'esigenza di stretching di una SAN via rete. Alcuni clienti non dispongono di connettività a latenza sufficientemente bassa tra i siti o non

dispongono dell'infrastruttura per il tunnel del traffico FC SAN su una rete intersito.

Lo svantaggio legato all'accesso non uniforme è che alcuni scenari di errore, inclusa la perdita del collegamento di replica, provocheranno la perdita dell'accesso allo storage da parte di alcuni host. In caso di interruzione della connettività dello storage locale, le applicazioni eseguite come istanze singole, come ad esempio i database non in cluster, eseguiti in maniera intrinseca solo su un singolo host in uno qualsiasi dei supporti di montaggio, si guasterebbero. I dati sarebbero comunque protetti, ma il server di database non avrebbe più accesso. Dovrebbe essere riavviato su un sito remoto, preferibilmente tramite un processo automatizzato. Ad esempio, VMware ha è in grado di rilevare una situazione di tutti i percorsi verso l'esterno su un server e di riavviare una macchina virtuale su un altro server in cui sono disponibili i percorsi.

Al contrario, un'applicazione in cluster come Oracle RAC può offrire un servizio disponibile contemporaneamente in due siti diversi. Perdere un sito non significa perdere il servizio dell'applicazione nel suo complesso. Le istanze sono ancora disponibili e in esecuzione nel sito sopravvissuto.

In molti casi, l'overhead della latenza aggiuntivo di un'applicazione che accede allo storage attraverso un collegamento da sito a sito sarebbe inaccettabile. Ciò significa che la migliore disponibilità di una rete uniforme è minima, poiché la perdita di storage su un sito comporterebbe comunque la necessità di arrestare i servizi sul sito in cui si è verificato l'errore.

Esistono percorsi ridondanti attraverso il cluster locale non mostrati in questi diagrammi per semplicità. I sistemi di storage ONTAP sono ad alta disponibilità, pertanto un guasto a un controller non dovrebbe causare guasti nel sito. Ciò dovrebbe comportare solo una modifica dei percorsi locali utilizzati nel sito interessato.

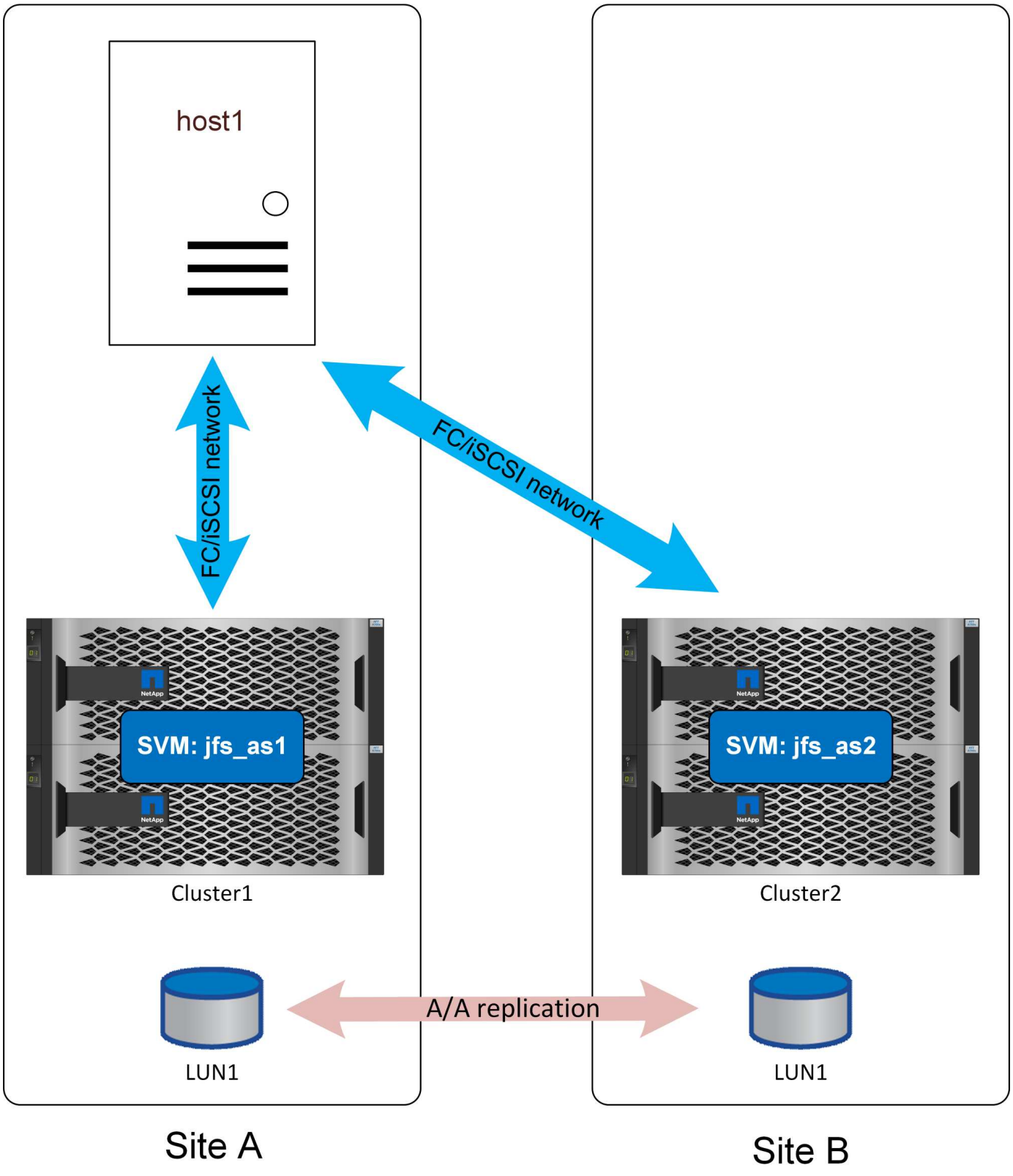
Panoramica

SQL Server può essere configurato per funzionare con la sincronizzazione attiva di SnapMirror in diversi modi. La risposta giusta dipende dalla connettività di rete disponibile, dai requisiti RPO e dai requisiti di disponibilità.

Istanza standalone di SQL Server

Le procedure consigliate per il layout dei file e la configurazione del server sono identiche a quelle consigliate nella "Server SQL su ONTAP" documentazione.

Con una configurazione standalone, SQL Server potrebbe essere eseguito solo in un sito. Presumibilmente "uniforme"si userebbe l'accesso.



Grazie a un accesso uniforme, un guasto dello storage in uno dei siti non interromperebbe le operazioni di database. Un errore completo del sito che includeva il server di database comporterebbe, ovviamente, un'interruzione del servizio.

Alcuni clienti potrebbero configurare un sistema operativo in esecuzione nel sito remoto con un'installazione di SQL Server preconfigurata, aggiornata con una versione di build equivalente a quella dell'istanza di

produzione. Il failover richiede l'attivazione dell'istanza standalone di SQL Server nel sito alternativo, il rilevamento dei LUN e l'avvio del database. È possibile automatizzare il processo completo con il cmdlet di Windows PowerShell, in quanto non sono richieste operazioni dal lato storage.

"Non uniforme" è possibile utilizzare anche l'accesso, ma il risultato sarebbe un'interruzione del database se il sistema storage in cui si trovava il server del database non avesse avuto esito positivo, perché il database non avrebbe percorsi disponibili per lo storage. In alcuni casi ciò può ancora essere accettabile. La sincronizzazione attiva di SnapMirror garantirebbe comunque una data Protection con RPO=0 e, in caso di guasto del sito, la copia restante sarebbe attiva e pronta a riprendere le operazioni seguendo la stessa procedura utilizzata con un accesso uniforme descritta sopra.

Un processo di failover semplice e automatizzato può essere configurato più facilmente con l'uso di un host virtualizzato. Ad esempio, se i file di dati di SQL Server vengono replicati in modo sincrono su storage secondario insieme a un VMDK di avvio, in caso di disastro è possibile attivare l'ambiente completo nel sito alternativo. Un amministratore può attivare manualmente l'host nel sito rimasto o automatizzare il processo tramite un servizio come VMware ha.

Istanza del cluster di failover di SQL Server

Le istanze di failover di SQL Server possono essere ospitate anche in un cluster di failover di Windows in esecuzione su un server fisico o virtuale come sistema operativo guest. Questa architettura multi-host offre un'istanza di SQL Server e la resilienza dello storage. Tale implementazione è utile in ambienti a domanda elevata che richiedono solidi processi di failover senza rinunciare a prestazioni avanzate. In una configurazione del cluster di failover, quando un host o uno storage primario viene colpito, SQL Services eseguirà il failover sull'host secondario e, allo stesso tempo, lo storage secondario sarà disponibile per servire io. Non è richiesto alcuno script di automazione o intervento dell'amministratore.

Scenari di errore

La pianificazione di un'architettura completa dell'applicazione SnapMirror Active Sync richiede la comprensione del modo in cui SM-AS risponderà in vari scenari di failover pianificati e non pianificati.

Per gli esempi seguenti, si supponga che il sito A sia configurato come sito preferito.

Interruzione della connettività di replica

Se la replica SM-AS viene interrotta, non è possibile completare la scrittura io perché sarebbe impossibile per un cluster replicare le modifiche al sito opposto.

Sito A (sito preferito)

Il risultato dell'errore del collegamento di replica sul sito preferito sarà una pausa di circa 15 secondi nell'elaborazione io in scrittura, poiché ONTAP ritenta le operazioni di scrittura replicate prima di determinare che il collegamento di replica è veramente irraggiungibile. Trascorsi i 15 secondi, il sistema del sito A riprende l'elaborazione io in lettura e scrittura. I percorsi SAN non vengono modificati e i LUN rimangono online.

Sito B

Poiché il sito B non è il sito preferito di sincronizzazione attiva SnapMirror, i relativi percorsi LUN non saranno più disponibili dopo circa 15 secondi.

Errore del sistema di storage

Il risultato di un errore del sistema di storage è quasi identico al risultato della perdita del collegamento di replica. Il sito sopravvissuto dovrebbe subire una pausa io di circa 15 secondi. Trascorso questo periodo di 15 secondi, io riprenderà sul sito come di consueto.

Perdita del mediatore

Il servizio di mediazione non controlla direttamente le operazioni di storage. Funziona come un percorso di controllo alternativo tra cluster. Esiste principalmente per automatizzare il failover senza il rischio di uno scenario split-brain. Durante l'utilizzo normale, ogni cluster replica le modifiche al partner e pertanto ogni cluster può verificare che il cluster partner sia online e fornisca i dati. Se il collegamento di replica non è riuscito, la replica viene interrotta.

Il motivo per cui è necessario un mediatore per il failover automatizzato sicuro è perché altrimenti sarebbe impossibile per un cluster di storage determinare se la perdita di comunicazione bidirezionale fosse il risultato di un'interruzione della rete o di un errore effettivo dello storage.

Il mediatore fornisce un percorso alternativo per ciascun cluster per verificare lo stato di salute del partner. Gli scenari sono i seguenti:

- Se un cluster può contattare direttamente il partner, i servizi di replica sono operativi. Non è richiesta alcuna azione.
- Se un sito preferito non può contattare direttamente il proprio partner o tramite il mediatore, presuppone che il partner sia effettivamente non disponibile oppure è stato isolato e ha portato i percorsi LUN offline. Il sito preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'io in lettura e in scrittura.
- Se un sito non preferito non può contattare direttamente il proprio partner, ma può contattarlo tramite il mediatore, prenderà i suoi percorsi offline e attenderà il ritorno della connessione di replica.
- Se un sito non preferito non può contattare direttamente il proprio partner o tramite un mediatore operativo, supporterà che il partner sia effettivamente non disponibile, oppure che sia stato isolato e che abbia portato i percorsi LUN offline. Il sito non preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'i/o in lettura e scrittura. Assumerà il ruolo dell'origine della replica e diventerà il nuovo sito preferito.

Se il mediatore non è completamente disponibile:

- In caso di guasto dei servizi di replica per qualsiasi motivo, incluso un guasto del sito o del sistema storage non preferito, il sito preferito rilascerà lo stato RPO=0 e riprenderà l'elaborazione i/o in lettura e scrittura. Il sito non preferito prenderà i suoi percorsi offline.
- Il guasto del sito preferito causerà un'interruzione poiché il sito non preferito non sarà in grado di verificare che il sito opposto sia effettivamente offline e quindi non sarebbe sicuro per il sito non preferito riprendere i servizi.

Ripristino dei servizi in corso

Dopo aver risolto un errore, come il ripristino della connettività da sito a sito o l'accensione di un sistema guasto, gli endpoint di sincronizzazione attivi di SnapMirror rilevano automaticamente la presenza di una relazione di replica difettosa e la riportano allo stato RPO=0. Una volta ristabilita la replica sincrona, i percorsi non riusciti torneranno in linea.

In molti casi, le applicazioni in cluster rilevano automaticamente la restituzione dei percorsi non riusciti e tali applicazioni tornano online. In altri casi, potrebbe essere necessaria una scansione SAN a livello di host

oppure potrebbe essere necessario riportare le applicazioni online manualmente. Dipende dall'applicazione e dal modo in cui è configurata, e in generale tali attività possono essere facilmente automatizzate. ONTAP si sta auto-riparando e non deve richiedere alcun intervento da parte dell'utente per riprendere le operazioni di storage RPO = 0 KB.

Failover manuale

La modifica del sito preferito richiede un'operazione semplice. I/o si fermeranno per un secondo o due come autorità sugli switch del comportamento di replica tra i cluster, ma in caso contrario i/o non vengono influenzati.

Configurazione dello storage sui sistemi ASA R2

Panoramica

NetApp ASA R2 è una soluzione potente e semplificata per i clienti solo SAN che eseguono carichi di lavoro mission-critical. La combinazione della piattaforma ASA R2 con soluzioni di storage ONTAP e Microsoft SQL Server offre design di storage dei database di livello Enterprise in grado di soddisfare le esigenze applicative più esigenti di oggi.

Le seguenti piattaforme ASA sono classificate come sistemi ASA R2 che supportano tutti i protocolli SAN (iSCSI, FC, NVMe/FC, NVMe/TCP). I protocolli iSCSI, FC, NVMe/FC e NVMe/TCP supportano l'architettura Active-Active simmetrica per il multipathing, in modo che tutti i percorsi tra gli host e lo storage siano attivi/ottimizzati:

- ASA A1K
- ASA A90
- ASA A70
- ASA A50
- ASA A30
- ASA A20

Per ulteriori informazioni, vedere "NetApp ASA"

L'ottimizzazione di una soluzione SQL Server su ONTAP richiede la comprensione del modello e delle caratteristiche di i/o di SQL Server. Un layout di storage ben progettato per un database SQL Server deve supportare i requisiti relativi alle performance di SQL Server, offrendo al contempo la massima capacità di gestione dell'infrastruttura nel suo complesso. Un buon layout dello storage permette inoltre di avere successo nell'implementazione iniziale e di far crescere l'ambiente senza problemi nel tempo, con il crescere dell'azienda.

Progettazione dello storage dei dati

Microsoft consiglia di collocare i dati e i file di registro su unità separate. Per le applicazioni che aggiornano e richiedono contemporaneamente i dati, il file di log è intensivo in scrittura e il file di dati (a seconda dell'applicazione) è intensivo in lettura/scrittura. Per il recupero dei dati, il file di log non è necessario. Pertanto, le richieste di dati possono essere soddisfatte dal file di dati posto sul proprio disco.

Quando si crea un nuovo database, Microsoft consiglia di specificare unità separate per i dati e i registri. Per spostare i file dopo la creazione del database, il database deve essere portato offline. Per ulteriori consigli

Microsoft, vedere "Posizionare i file di dati e di registro su unità separate".

Considerazioni sull'unità di conservazione

L'unità storage in ASA si riferisce a LUN per host SCSI/FC o a un namespace NVMe per host NVMe. In base al protocollo supportato, ti verrà richiesto di creare LUN, namespace NVMe o entrambi. Prima di creare un'unità di storage per la distribuzione del database, è importante comprendere in che modo il modello i/o di SQL Server e le relative caratteristiche variano in base al carico di lavoro e ai requisiti di backup e ripristino. Vedere le seguenti raccomandazioni NetApp per l'unità di conservazione:

- Evitare la condivisione della stessa unità di storage tra più SQL Server in esecuzione sullo stesso host, per evitare una gestione complicata. Nel caso di esecuzione di più istanze di SQL Server sullo stesso host, a meno che non si sia vicini al limite dell'unità di archiviazione su un nodo, evitare la condivisione e disporre invece di un'unità di archiviazione separata per istanza per host per semplificare la gestione dei dati.
- Utilizzare i punti di montaggio NTFS invece delle lettere dell'unità per superare il limite di 26 lettere di unità in Windows.
- Disattivare le pianificazioni delle snapshot e i criteri di conservazione. Utilizzare invece SnapCenter per coordinare le copie Snapshot dell'unità di archiviazione dati di SQL Server.
- Posizionare i database del sistema SQL Server su un'unità di archiviazione dedicata.
- Tempdb è un database di sistema utilizzato da SQL Server come area di lavoro temporanea, in particolare per operazioni DBCC CHECKDB i/o intensive. Pertanto, collocare questo database su un'unità di archiviazione dedicata. In ambienti di grandi dimensioni in cui il numero di unità di storage rappresenta una sfida, è possibile consolidare tempdb con database di sistema nella stessa unità di storage dopo un'attenta pianificazione. La protezione dei dati per tempdb non è una priorità elevata perché questo database viene ricreato ogni volta che SQL Server viene riavviato.
- Collocare i file di dati utente (`.mdf`) su un'unità di archiviazione separata perché si tratta di carichi di lavoro di lettura/scrittura casuali. È comune creare backup del log delle transazioni con maggiore frequenza rispetto ai backup del database. Per questo motivo, inserire i file di log delle transazioni (`.ldf`) in un'unità di archiviazione separata o VMDK dai file di dati in modo che sia possibile creare pianificazioni di backup indipendenti per ciascuno di essi. Questa separazione isola inoltre l'i/o di scrittura sequenziale dei file di log dall'i/o di lettura/scrittura casuale dei file di dati e migliora significativamente le prestazioni di SQL Server.
- Assicurarsi che i file del database utente e la directory di log per l'archiviazione del backup del log si trovino su un'unità di archiviazione separata per evitare che il criterio di conservazione sovrascriva gli snapshot quando vengono utilizzati con la funzione SnapMirror con i criteri del vault.
- Non mischiare file di database e non di database, come file di ricerca full-text, sulla stessa unità di archiviazione.
- L'inserimento dei file secondari del database (come parte di un filegroup) in un'unità di archiviazione separata migliora le prestazioni del database di SQL Server. Questa separazione è valida solo se il file del database `.mdf` non condivide l'unità di archiviazione con altri `.mdf` file.
- Durante la formattazione del disco utilizzando il gestore dischi nel server Windows, assicurarsi che la dimensione dell'unità di allocazione sia impostata su 64K per la partizione.
- Non posizionare database utente o database di sistema su un'unità di archiviazione che ospita punti di montaggio.
- Vedere "Microsoft Windows e MPIO nativo nelle Best practice ONTAP per le SAN moderne" Per applicare il supporto multipathing in Windows ai dispositivi iSCSI nelle proprietà MPIO.
- Se si utilizza un'istanza cluster di failover sempre attiva, i database utente devono essere collocati nell'unità di storage condivisa tra i nodi cluster di failover del server Windows e le risorse del cluster di dischi fisici vengono assegnate al gruppo di cluster associato all'istanza di SQL Server.

File di database e filegroup

Il corretto posizionamento dei file del database SQL Server su ONTAP è fondamentale durante la fase di distribuzione iniziale. Ciò garantisce prestazioni ottimali, gestione dello spazio, tempi di backup e ripristino che possono essere configurati in base alle esigenze aziendali.

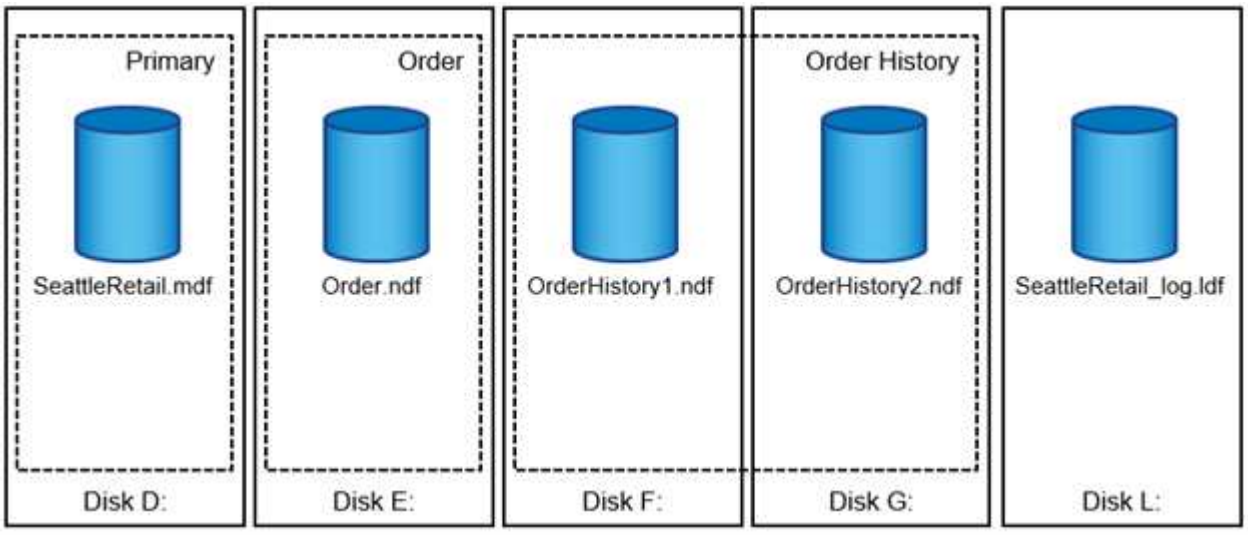
In teoria, SQL Server (a 64 bit) supporta 32.767 database per istanza e 524.272TB di dimensioni del database, sebbene l'installazione tipica abbia in genere diversi database. Tuttavia, il numero di database che SQL Server è in grado di gestire dipende dal carico e dall'hardware. Non è insolito vedere le istanze di SQL Server che ospitano decine, centinaia o persino migliaia di database di piccole dimensioni.

File di database e filegroup

Ogni database è costituito da uno o più file di dati e da uno o più file di registro delle transazioni. Il registro delle transazioni memorizza le informazioni sulle transazioni del database e tutte le modifiche apportate ai dati da ciascuna sessione. Ogni volta che i dati vengono modificati, SQL Server memorizza informazioni sufficienti nel log delle transazioni per annullare (eseguire il rollback) o ripristinare (riprodurre nuovamente) l'azione. Un log delle transazioni di SQL Server è parte integrante della reputazione di SQL Server in termini di integrità e robustezza dei dati. Il log delle transazioni è fondamentale per le funzionalità di atomicità, coerenza, isolamento e durata (ACID) di SQL Server. SQL Server scrive nel registro delle transazioni non appena si verifica una modifica alla pagina dei dati. Ogni istruzione DML (Data Manipulation Language) (ad esempio, SELECT, INSERT, Update o DELETE) è una transazione completa e il log delle transazioni garantisce che l'intera operazione basata su set abbia luogo, assicurando l'atomicità della transazione.

Ogni database dispone di un file di dati primario che, per impostazione predefinita, ha l'estensione .mdf. Inoltre, ogni database può disporre di file di database secondari. Questi file, per impostazione predefinita, hanno estensioni .ndf.

Tutti i file di database sono raggruppati in filegroup. Un filegroup è l'unità logica, che semplifica l'amministrazione del database. Consentono la separazione tra il posizionamento degli oggetti logici e i file di database fisici. Quando si creano le tabelle degli oggetti del database, si specifica in quale filegroup devono essere posizionati senza preoccuparsi della configurazione del file di dati sottostante.



La possibilità di inserire più file di dati all'interno del filegroup consente di distribuire il carico su diversi dispositivi di archiviazione, migliorando le prestazioni di i/o del sistema. Al contrario, il log delle transazioni non trae vantaggio dai file multipli poiché SQL Server scrive nel log delle transazioni in modo sequenziale.

La separazione tra il posizionamento degli oggetti logici nei filegroup e i file di database fisici consente di ottimizzare il layout dei file di database, ottenendo il massimo dal sottosistema di storage. Il numero di file di dati che supportano un carico di lavoro può essere variato in base alle necessità per supportare i requisiti di i/o e la capacità prevista, senza influire sull'applicazione. Queste variazioni nel layout del database sono trasparenti per gli sviluppatori di applicazioni, che posizionano gli oggetti del database nei filegroup piuttosto che nei file di database.

NetApp recommended evitare l'utilizzo del filegroup primario per oggetti diversi da quelli di sistema. La creazione di un filegroup separato o di un set di filegroup per gli oggetti utente semplifica l'amministrazione del database e il ripristino di emergenza, soprattutto nel caso di database di grandi dimensioni.

Inizializzazione del file di istanza del database

È possibile specificare le dimensioni iniziali del file e i parametri di crescita automatica al momento della creazione del database o dell'aggiunta di nuovi file a un database esistente. SQL Server utilizza un algoritmo di riempimento proporzionale quando sceglie in quale file di dati scrivere i dati. Scrive una quantità di dati proporzionalmente allo spazio libero disponibile nei file. Maggiore è lo spazio libero nel file, maggiore è il numero di scritture gestite.

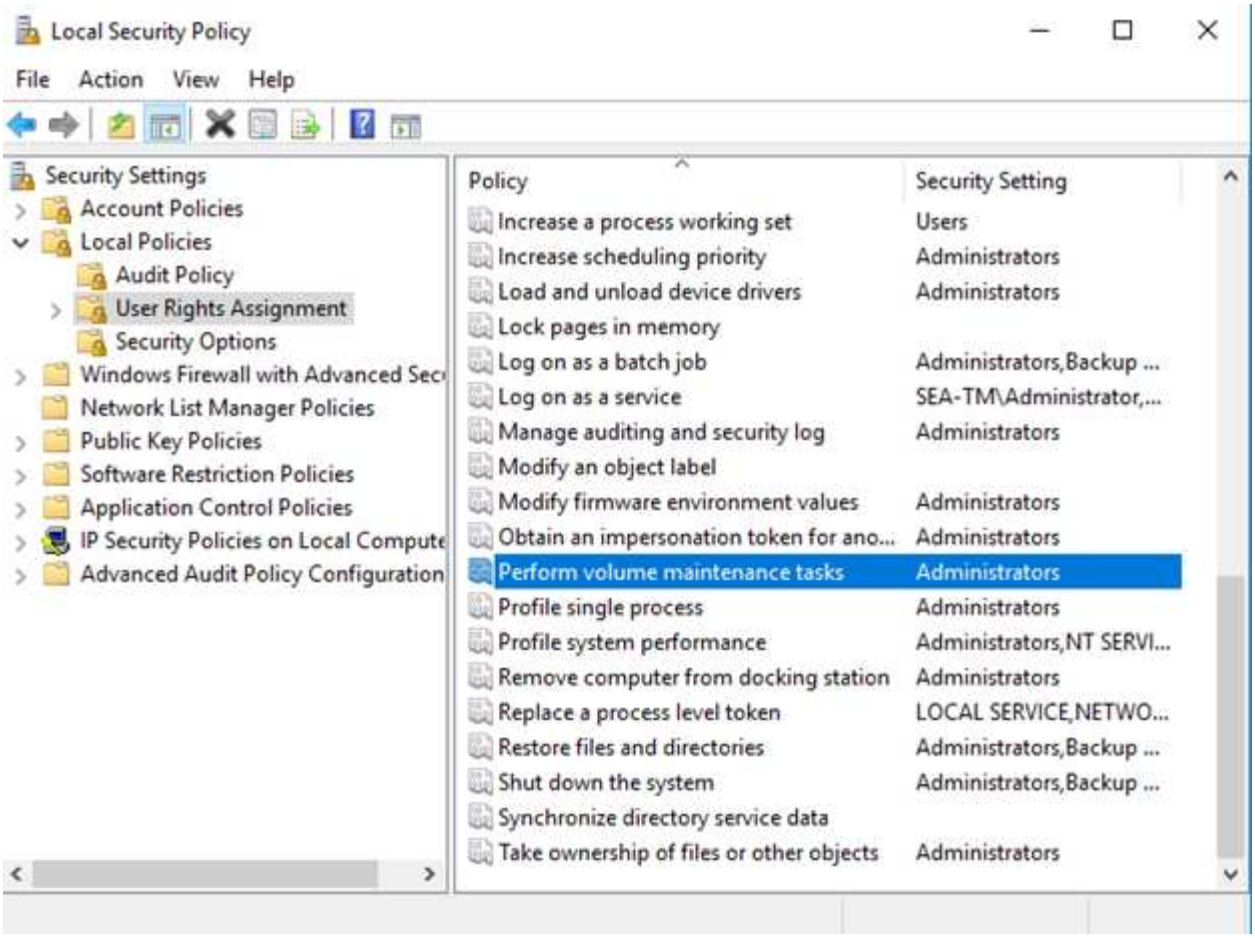
NetApp consiglia che tutti i file nel singolo filegroup abbiano le stesse dimensioni iniziali e parametri di crescita automatica, con la dimensione di crescita definita in megabyte piuttosto che in percentuali. Questo aiuta l'algoritmo di riempimento proporzionale a bilanciare uniformemente le attività di scrittura nei file di dati.

Ogni volta che SQL Server espande i file, riempie di zero lo spazio appena allocato. Questo processo blocca tutte le sessioni che devono scrivere nel file corrispondente o, in caso di crescita del log delle transazioni, genera record di log delle transazioni.

SQL Server azzera sempre il log delle transazioni e questo comportamento non può essere modificato. Tuttavia, è possibile controllare se i file di dati vengono azzerati attivando o disattivando l'inizializzazione istantanea dei file. L'attivazione dell'inizializzazione immediata dei file consente di velocizzare la crescita dei file di dati e di ridurre il tempo necessario per creare o ripristinare il database.

Un piccolo rischio per la sicurezza è associato all'inizializzazione immediata dei file. Quando questa opzione è attivata, le parti non allocate del file di dati possono contenere informazioni provenienti da file del sistema operativo eliminati in precedenza. Gli amministratori di database possono esaminare tali dati.

È possibile attivare l'inizializzazione immediata dei file aggiungendo l'autorizzazione SA_MANAGE_VOLUME_NAME, nota anche come "Esegui attività di manutenzione del volume" all'account di avvio di SQL Server. È possibile eseguire questa operazione nell'applicazione di gestione dei criteri di protezione locale (secpol.msc), come illustrato nella figura seguente. Aprire le proprietà per l'autorizzazione "Esegui attività di manutenzione del volume" e aggiungere l'account di avvio di SQL Server all'elenco degli utenti.



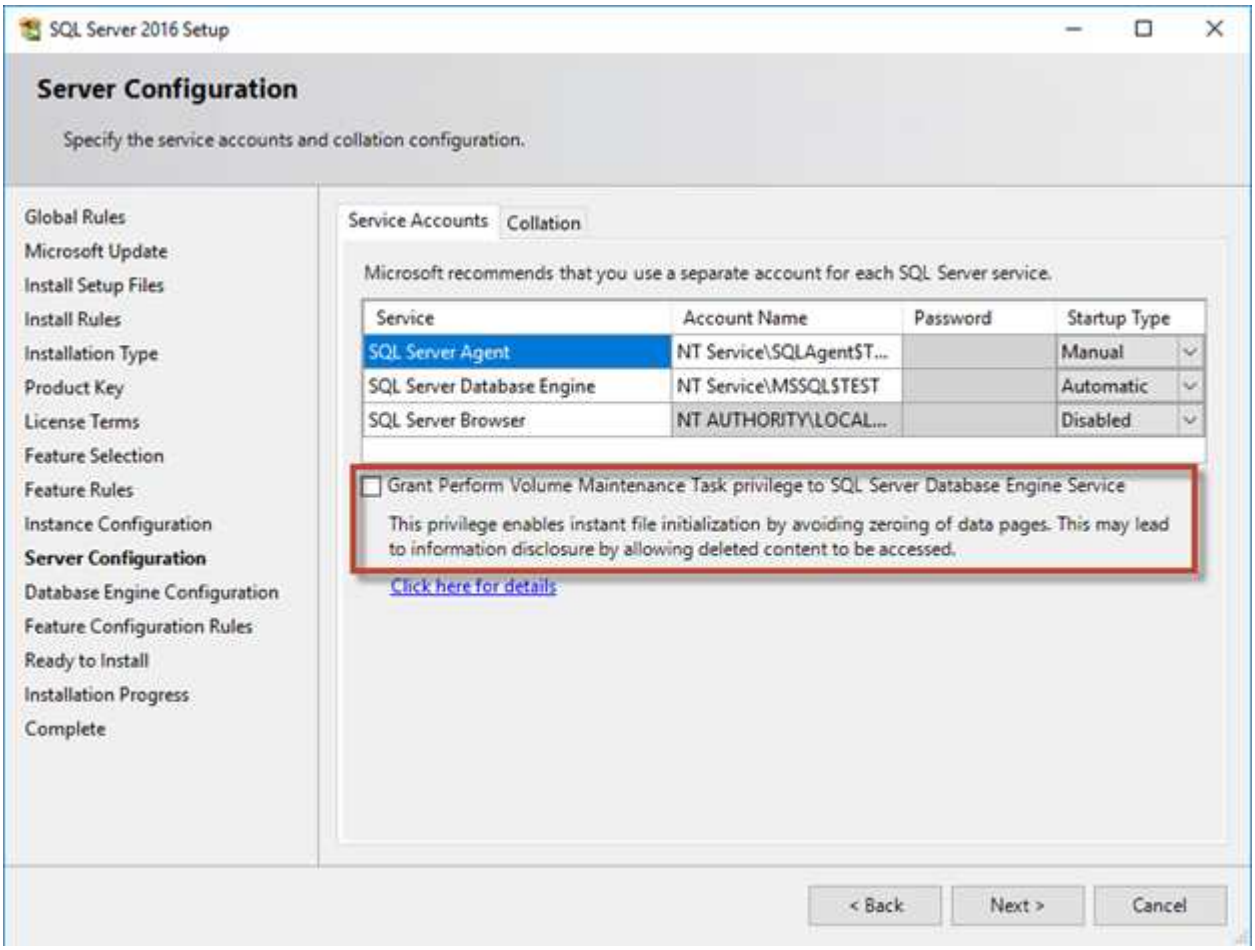
Per verificare se l'autorizzazione è attivata, è possibile utilizzare il codice riportato nell'esempio seguente. Questo codice imposta due flag di traccia che obbligano SQL Server a scrivere informazioni aggiuntive nel registro degli errori, a creare un database di piccole dimensioni e a leggere il contenuto del registro.

```
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO
```

Quando l'inizializzazione immediata del file non è attivata, il registro degli errori di SQL Server mostra che SQL Server sta azzerando il file di dati mdf oltre a azzerare il file di registro ldf, come illustrato nell'esempio seguente. Quando l'inizializzazione immediata del file è attivata, viene visualizzato solo l'azzeramento del file di registro.

	LogDate	ProcessInfo	Text
365	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 flush delta counts.
366	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 logging active xact info.
367	2017-02-09 08:10:07.750	spid53	Ckpt dbid 3 phase 1 ended (8)
368	2017-02-09 08:10:07.750	spid53	About to log Checkpoint end.
369	2017-02-09 08:10:07.880	spid53	Ckpt dbid 3 complete
370	2017-02-09 08:10:08.130	spid53	Starting up database 'DelMe'.
371	2017-02-09 08:10:08.150	spid53	FixupLogTail(progress) zeroing C:\Program Files\Micros
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQ
373	2017-02-09 08:10:08.170	spid53	Zeroing completed on C:\Program Files\Microsoft SQL
374	2017-02-09 08:10:08.710	spid53	Ckpt dbid 6 started
375	2017-02-09 08:10:08.710	spid53	About to log Checkpoint begin.

L'attività Esegui manutenzione volume è semplificata in SQL Server 2016 e viene fornita come opzione durante il processo di installazione. In questa figura viene visualizzata l'opzione per concedere al servizio del motore di database di SQL Server il privilegio di eseguire l'attività di manutenzione del volume.



Un'altra importante opzione del database che controlla le dimensioni dei file di database è l'autohrink. Quando questa opzione è attivata, SQL Server riduce regolarmente i file di database, ne riduce le dimensioni e rilascia spazio al sistema operativo. Questa operazione richiede molte risorse ed è raramente utile perché i file di database crescono di nuovo dopo un certo periodo di tempo quando nuovi dati entrano nel sistema. Il collegamento automatico non deve essere attivato nel database.

Directory di log

La directory di registro è specificata in SQL Server per memorizzare i dati di backup del registro delle transazioni a livello di host. Se si utilizza SnapCenter per eseguire il backup dei file di registro, ciascun host SQL Server utilizzato da SnapCenter deve disporre di una directory di registro host configurata per eseguire i backup dei registri.

Posizionare la directory di registro su un'unità di archiviazione dedicata. La quantità di dati nella directory del registro host dipende dalle dimensioni dei backup e dal numero di giorni in cui i backup vengono conservati. SnapCenter consente una sola directory di registro host per host SQL Server. È possibile configurare le directory del registro host in SnapCenter -→ host -→ Configura plug-in.

NetApp consiglia quanto segue per una directory del registro host:

- Assicurarsi che la directory del registro host non sia condivisa da altri tipi di dati che potrebbero danneggiare i dati dello snapshot di backup.
- Creare la directory del registro host su un'unità di archiviazione dedicata in cui SnapCenter copia i registri delle transazioni.
- Se si utilizza un'istanza cluster di failover sempre attiva, l'unità di storage utilizzata per la directory di log dell'host deve essere una risorsa disco del cluster nello stesso gruppo di cluster dell'istanza di SQL Server di cui si esegue il backup in SnapCenter.

Protezione dei dati

Le strategie di backup dei database devono basarsi su requisiti aziendali identificati, non su capacità teoriche. Combinando la tecnologia Snapshot di ONTAP e sfruttando le API di Microsoft SQL Server, è possibile eseguire rapidamente un backup coerente delle applicazioni indipendentemente dalle dimensioni dei database utente. Per requisiti di gestione dei dati più avanzati o scale-out, NetApp offre SnapCenter.

SnapCenter

SnapCenter è il software di data Protection di NetApp per le applicazioni aziendali. I database di SQL Server possono essere protetti in modo rapido e semplice con il plug-in SnapCenter per SQL Server e con operazioni del sistema operativo gestite dal plug-in SnapCenter per Microsoft Windows.

L'istanza di SQL Server può essere un'installazione autonoma, un'istanza cluster di failover o può essere un gruppo di disponibilità sempre attivo. Il risultato è che, grazie a un singolo pannello di controllo, i database possono essere protetti, clonati e ripristinati da una copia primaria o secondaria. SnapCenter può gestire database SQL Server sia on-premise, nel cloud che in configurazioni ibride. Le copie dei database possono essere create in pochi minuti sull'host originale o alternativo per lo sviluppo o per il reporting.

SQL Server richiede inoltre un coordinamento tra il sistema operativo e lo storage per garantire che i dati corretti siano presenti negli snapshot al momento della creazione. Nella maggior parte dei casi, l'unico metodo sicuro per eseguire questa operazione è SnapCenter o T-SQL. Gli snapshot creati senza questo coordinamento aggiuntivo potrebbero non essere recuperabili in modo affidabile.

Per ulteriori informazioni sul plug-in di SQL Server per SnapCenter, vedere "TR-4714: Guida alle Best practice per SQL Server con NetApp SnapCenter".

Protezione del database mediante snapshot T-SQL

In SQL Server 2022, Microsoft ha introdotto le istantanee T-SQL che offrono un percorso per la creazione di script e l'automazione delle operazioni di backup. Invece di eseguire copie di dimensioni normali, è possibile preparare il database per le snapshot. Una volta che il database è pronto per il backup, è possibile sfruttare le API REST di ONTAP per creare snapshot.

Di seguito è riportato un esempio di flusso di lavoro di backup:

1. Bloccare un database con il comando ALTER. In questo modo il database viene preparato per uno snapshot coerente sullo storage sottostante. Dopo il blocco è possibile scongelare il database e registrare lo snapshot con il comando di BACKUP.
2. Eseguire istantanee di più database sulle unità di archiviazione contemporaneamente con il nuovo GRUPPO DI BACKUP e i comandi del SERVER DI BACKUP.
3. Se il carico di lavoro del database viene eseguito su più unità di storage, creare gruppi di coerenza per semplificare l'attività di gestione. Il gruppo di coerenza è un insieme di unità di archiviazione gestite come unità singola.
4. Eseguire backup COMPLETI o backup COMPLETI COPY_ONLY. Anche questi backup sono registrati in msdb.
5. Eseguire il recovery point-in-time utilizzando i backup di log eseguiti con il normale approccio di streaming dopo il backup COMPLETO delle snapshot. Se lo si desidera, sono supportati anche i backup differenziali in streaming.

Per ulteriori informazioni, vedere "Documentazione Microsoft per conoscere le istantanee T-SQL".

NetApp recommended Using SnapCenter to create Snapshot copy. Anche il metodo T-SQL descritto sopra funziona, ma SnapCenter offre un'automazione completa sul processo di backup, ripristino e cloning. Esegue inoltre il rilevamento per garantire che vengano creati gli snapshot corretti.

Utilizzare SnapCenter se i database di SQL Server sono ospitati su disco virtuale eseguito in un ambiente virtualizzato VMware. Nel caso in cui i database siano ospitati su server fisici bare metal o su SCSI guest-Connect, eseguire il backup del database utilizzando il comando snapshot T-SQL.

Disaster recovery

Disaster recovery

I database e le infrastrutture applicative aziendali spesso richiedono la replica per proteggersi da disastri naturali o interruzioni impreviste del business, con tempi di inattività minimi.

La funzionalità di replica del gruppo di disponibilità always-on di SQL Server può essere un'opzione eccellente e NetApp offre opzioni per integrare la protezione dei dati con Always-on. In alcuni casi, tuttavia, è consigliabile prendere in considerazione la tecnologia di replica ONTAP utilizzando le seguenti opzioni.

SnapMirror

La tecnologia SnapMirror offre una soluzione aziendale rapida e flessibile per la replica dei dati su LAN e WAN. La tecnologia SnapMirror trasferisce solo i blocchi di dati modificati a destinazione dopo la creazione del

mirror iniziale, riducendo in modo significativo i requisiti di larghezza di banda della rete. Può essere configurato in modalità sincrona o asincrona. La replica sincrona SnapMirror in NetApp ASA è configurata con la sincronizzazione attiva SnapMirror.

Sincronizzazione attiva di SnapMirror

Per molti clienti, la business continuity richiede molto più del semplice possesso di una copia remota dei dati, ma richiede la capacità di sfruttarli rapidamente, cosa possibile in NetApp ONTAP utilizzando la sincronizzazione attiva di SnapMirror

Con la sincronizzazione attiva di SnapMirror, avrai essenzialmente due sistemi ONTAP diversi che mantengono copie indipendenti dei dati LUN, ma cooperano per presentare una singola istanza di tale LUN. Dal punto di vista dell'host, si tratta di una singola entità LUN. La sincronizzazione attiva di SnapMirror è supportata per LUN basata su iSCSI/FC.

La SnapMirror Active Sync può fornire una replica con RPO=0 ed è semplice da implementare tra due cluster indipendenti. Una volta sincronizzate le due copie di dati, i due cluster devono solo mirrorare le scritture. Quando si verifica una scrittura su un cluster, questa viene replicata nell'altro cluster. La scrittura viene riconosciuta all'host solo quando la scrittura è stata completata su entrambi i siti. A parte questo comportamento di suddivisione del protocollo, i due cluster sono altrimenti normali cluster ONTAP.

Un caso d'utilizzo chiave per SM-AS è la replica granulare. A volte non vuoi replicare tutti i dati come una singola unità oppure devi eseguire il failover selettivo su alcuni carichi di lavoro.

Un altro caso d'utilizzo chiave per SM-AS è per operazioni Active-Active, dove desideri che siano disponibili copie dei dati completamente utilizzabili su due cluster diversi situati in due posizioni diverse con caratteristiche di performance identiche e, se desiderato, non richiedere l'estensione della SAN tra i siti. Le applicazioni potranno essere già in esecuzione su entrambi i siti, a condizione che sia supportata un'applicazione, con una conseguente riduzione dell'RTO complessivo durante le operazioni di failover.

SnapMirror

Di seguito sono riportati alcuni consigli su SnapMirror per SQL Server:

- Utilizza la replica sincrona con la sincronizzazione attiva SnapMirror, in cui la richiesta di un recovery di dati rapido è maggiore e soluzioni asincrone per la flessibilità negli RPO.
- Se stai utilizzando SnapCenter per eseguire il backup dei database e replicare le snapshot su un cluster remoto, non pianificare gli aggiornamenti SnapMirror dai controller per motivi di coerenza. Attiva invece gli update di SnapMirror da SnapCenter per aggiornare SnapMirror al termine del backup completo o del log.
- Bilancia le unità di storage che contengono dati SQL Server nei diversi nodi del cluster per consentire a tutti i nodi del cluster di condividere l'attività di replica di SnapMirror. Questa distribuzione ottimizza l'utilizzo delle risorse dei nodi.

Per ulteriori informazioni su SnapMirror, vedere "TR-4015: Guida alle Best practice e alla configurazione di SnapMirror per ONTAP 9".

Sincronizzazione attiva di SnapMirror

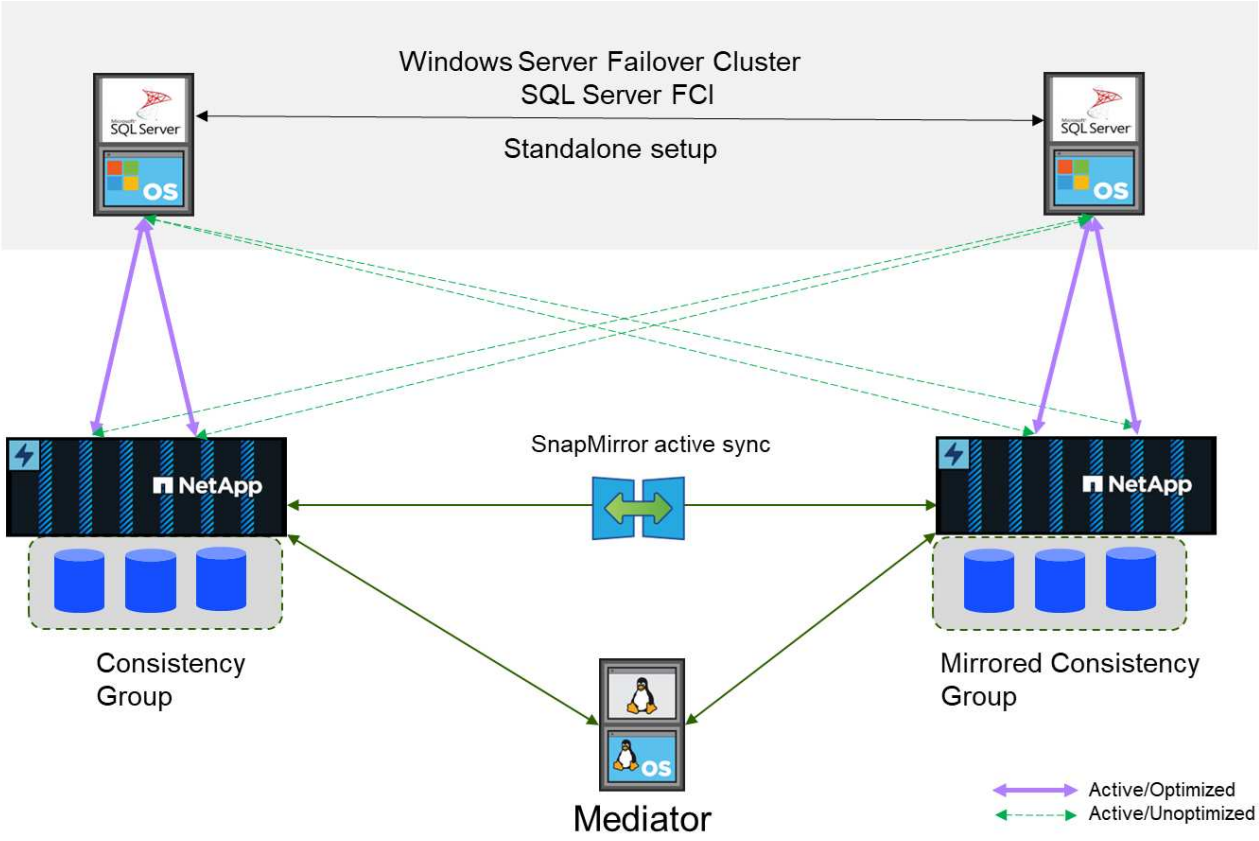
Panoramica

La sincronizzazione attiva di SnapMirror consente a singoli database e applicazioni SQL Server di continuare le operazioni durante interruzioni di rete e storage, con failover trasparente dello storage senza interventi manuali.

SnapMirror Active Sync supporta l'architettura Active/Active simmetrica che offre replica bidirezionale sincrona per la business continuity e il disaster recovery. Protegge l'accesso ai dati per carichi di lavoro critici SAN con accesso simultaneo in lettura e scrittura ai dati in diversi domini di errore, garantendo operazioni senza interruzioni e riducendo al minimo i downtime in caso di disastri o errori di sistema.

Gli host SQL Server accedono allo storage utilizzando LUN Fiber Channel (FC) o iSCSI. Replica tra ciascun cluster in cui risiede una copia dei dati replicati. Poiché questa funzionalità è la replica a livello di storage, le istanze di SQL Server in esecuzione su istanze di cluster di failover o host standalone possono eseguire operazioni di lettura/scrittura in entrambi i cluster. Per le fasi di pianificazione e configurazione, fare riferimento a "Documentazione di ONTAP sulla sincronizzazione attiva di SnapMirror" .

Architettura di SnapMirror Active con Active/Active simmetrico



Replica sincrona

Durante l'utilizzo normale, ciascuna copia è una replica sincrona con RPO=0/7, con un'unica eccezione. Se i dati non possono essere replicati, ONTAP rilascerà il requisito di replicare i dati e riprendere la distribuzione io su un sito mentre le LUN dell'altro sito vengono portate offline.

Hardware di archiviazione

A differenza di altre soluzioni di disaster recovery per lo storage, SnapMirror Active Sync offre una flessibilità asimmetrica della piattaforma. Non è necessario che l'hardware di ciascun sito sia identico. Questa funzionalità consente di dimensionare correttamente l'hardware utilizzato per supportare la sincronizzazione attiva di SnapMirror. Il sistema di storage remoto può essere identico al sito primario se deve supportare un carico di lavoro di produzione completo, ma se un disastro determina una riduzione dell'i/o, rispetto a un sistema più piccolo nel sito remoto potrebbe risultare più conveniente.

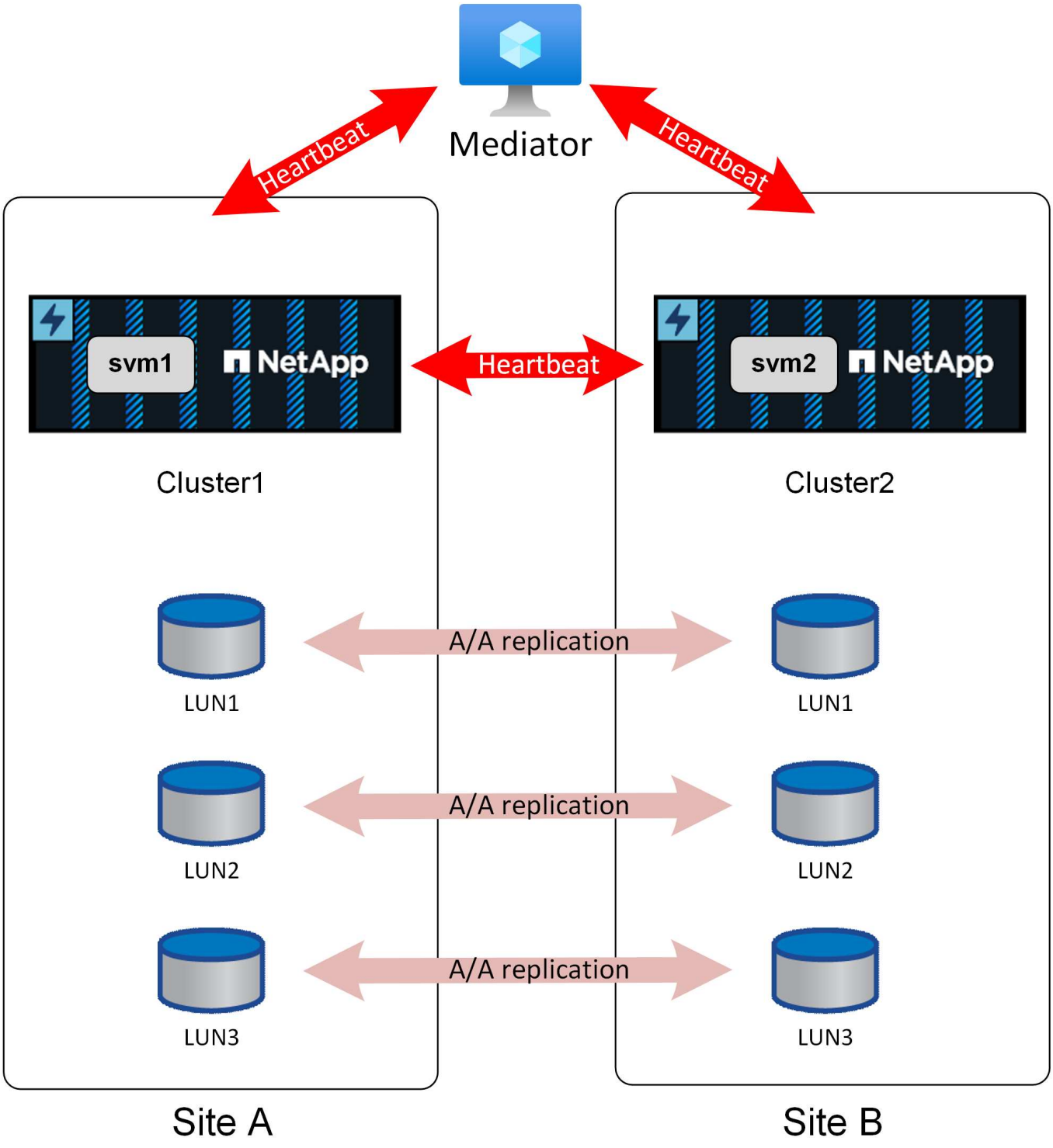
ONTAP mediatore

ONTAP Mediator è un'applicazione software che viene scaricata dal supporto NetApp e che viene in genere distribuita su una piccola macchina virtuale. Il ONTAP Mediator non è un tiebreaker. È un canale di comunicazione alternativo per i due cluster che partecipano alla replica sincrona attiva SnapMirror. Le operazioni automatizzate sono gestite da ONTAP in base alle risposte ricevute dal partner tramite connessioni dirette e tramite il mediatore.

Mediatore ONTAP

Il mediatore è necessario per automatizzare in modo sicuro il failover. Idealmente, sarebbe posizionato su un sito 3rd indipendente, ma può comunque funzionare per la maggior parte delle esigenze se colocated con uno dei cluster che partecipano alla replica.

Il mediatore non è realmente un tiebreaker, anche se quella è effettivamente la funzione che fornisce. Non esegue alcuna azione, ma fornisce un canale di comunicazione alternativo per la comunicazione tra cluster e cluster.



La sfida #1 con il failover automatizzato è il problema split-brain, e questo problema sorge se i due siti perdono la connettività tra loro. Che cosa dovrebbe accadere? Non si desidera che due siti diversi si designino come copie sopravvissute dei dati, ma in che modo un singolo sito può distinguere tra la perdita effettiva del sito opposto e l'impossibilità di comunicare con il sito opposto?

Qui entra il mediatore nell'immagine. Se si trova in un sito 3rd e ciascun sito dispone di una connessione di rete separata per tale sito, è disponibile un percorso aggiuntivo per ciascun sito per convalidare lo stato dell'altro. Esaminare nuovamente l'immagine sopra riportata e considerare i seguenti scenari.

- Cosa succede se il mediatore non riesce o è irraggiungibile da uno o entrambi i siti?
  - I due cluster possono ancora comunicare tra loro sullo stesso link utilizzato per i servizi di replica.
  - I dati sono ancora serviti con protezione RPO=0/7
- Cosa succede se il sito A non funziona?
  - Il sito B vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito B sostituirà i servizi dati, ma senza RPO = mirroring 0:1
- Cosa succede se il sito B non funziona?
  - Il sito A vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito A sostituirà i servizi dati, ma senza RPO = mirroring 0:1

Esiste un altro scenario da considerare: La perdita del collegamento di replica dei dati. In caso di perdita del link di replica tra i siti, RPO=0 Mirroring sarà ovviamente impossibile. Che cosa dovrebbe accadere allora?

Questo è controllato dallo stato del sito preferito. In una relazione SM-AS, uno dei siti è secondario all'altro. Questo non ha alcun effetto sulle normali operazioni e tutto l'accesso ai dati è simmetrico, ma se la replica viene interrotta, il legame dovrà essere interrotto per riprendere le operazioni. Ne risulta che il sito preferito continuerà le operazioni senza mirroring e il sito secondario interromperà l'elaborazione io fino al ripristino della comunicazione di replica.

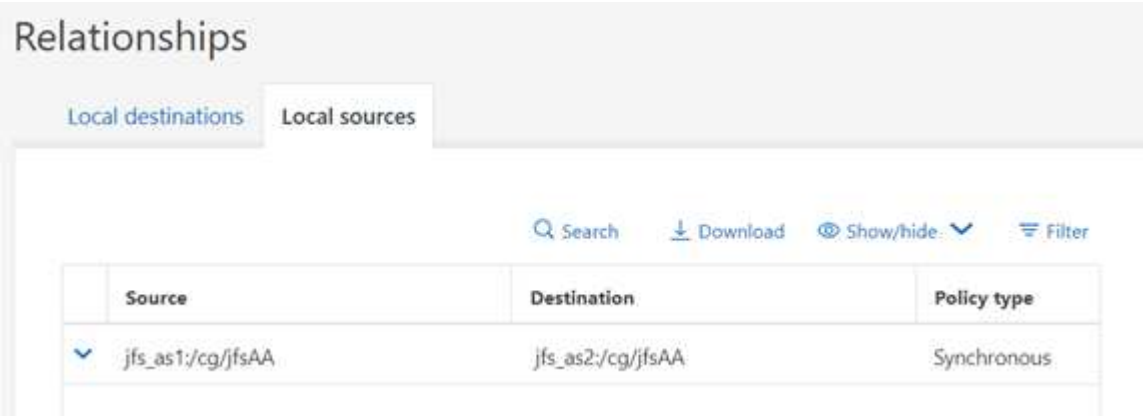
Sito preferito

Il comportamento di sincronizzazione attiva di SnapMirror è simmetrico, con una eccezione importante: Configurazione del sito preferita.

La sincronizzazione attiva di SnapMirror considererà un sito la "fonte" e l'altro la "destinazione". Ciò implica una relazione di replica unidirezionale, ma ciò non si applica al comportamento io. La replica è bidirezionale e simmetrica, mentre i tempi di risposta io sono identici su entrambi i lati del mirror.

La `source` designazione è controlla il sito preferito. In caso di perdita del link di replica, i percorsi delle LUN nella copia di origine continueranno a fornire i dati mentre i percorsi delle LUN nella copia di destinazione non saranno disponibili finché la replica non viene ristabilita e SnapMirror ritorna allo stato sincrono. I percorsi riprenderanno a fornire i dati.

La configurazione di origine/destinazione può essere visualizzata tramite SystemManager:



O all'interfaccia CLI:

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                      Destination Path: jfs_as2:/cg/jfsAA
                     Relationship Type: XDP
                Relationship Group Type: consistencygroup
                    SnapMirror Schedule: -
                 SnapMirror Policy Type: automated-failover-duplex
                      SnapMirror Policy: AutomatedFailOverDuplex
                            Tries Limit: -
                      Throttle (KB/sec): -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

La chiave è che source è l'SVM su cluster1. Come menzionato sopra, i termini "origine" e "destinazione" non descrivono il flusso di dati replicati. Entrambi i siti possono elaborare una scrittura e replicarla nel sito opposto. In effetti, entrambi i cluster sono origini e destinazioni. L'effetto della designazione di un cluster come origine controlla semplicemente quale cluster sopravvive come sistema di storage in lettura e scrittura in caso di perdita del link di replica.

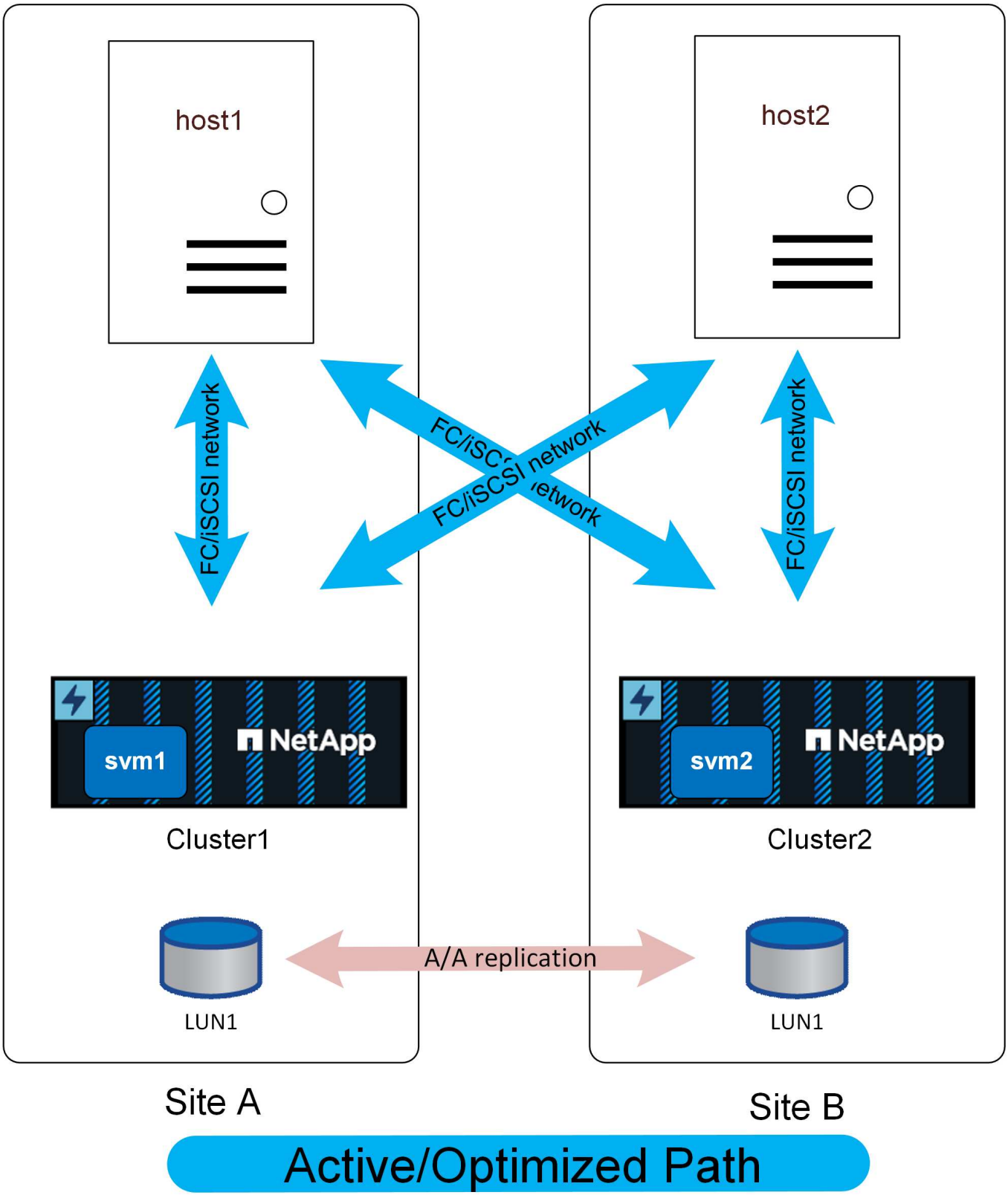
Topologia di rete

Accesso uniforme

Una rete ad accesso uniforme significa che gli host sono in grado di accedere ai percorsi su entrambi i siti (o ai domini di errore all'interno dello stesso sito).

Una caratteristica importante di SM-AS è la capacità di configurare i sistemi storage per sapere dove si trovano gli host. Quando si mappano i LUN a un determinato host, è possibile indicare se sono prossimali o meno a un determinato sistema di archiviazione.

I sistemi NetApp ASA offrono multipathing Active-Active su tutti i percorsi di un cluster. Questo vale anche per le configurazioni SM-AS.



Con un accesso uniforme, io attraverserebbe la WAN. Si tratta di un cluster completamente mesh collegato in rete e questo può essere o meno auspicabile per tutti i casi di utilizzo.

Se i due siti fossero a una distanza di 100 metri con connettività in fibra non dovrebbe esserci una latenza aggiuntiva rilevabile che attraversa la WAN, ma se i siti fossero a lunga distanza gli uni dagli altri, le

performance in lettura risulterebbero compromesse su entrambi i siti. ASA con rete di accesso non uniforme sarebbe un'opzione per ottenere i vantaggi in termini di costi e funzionalità di ASA senza incorrere in penalizzazioni per l'accesso alla latenza tra siti o utilizzare la funzionalità di prossimità dell'host per consentire l'accesso in lettura/scrittura locale tra siti per entrambi i siti.

ASA con SM-as in una configurazione a bassa latenza offre due benefici interessanti. In primo luogo, essenzialmente **raddoppia** le prestazioni per ogni singolo host perché io può essere gestito dal doppio dei controller utilizzando il doppio dei percorsi. In secondo luogo, in un ambiente a sito singolo offre una disponibilità estrema, perché è possibile perdere un intero sistema storage senza interrompere l'accesso degli host.

Impostazioni di prossimità

La prossimità si riferisce a una configurazione per cluster che indica che un determinato WWN host o ID iniziatore iSCSI appartiene a un host locale. Si tratta di un secondo passo opzionale per la configurazione dell'accesso LUN.

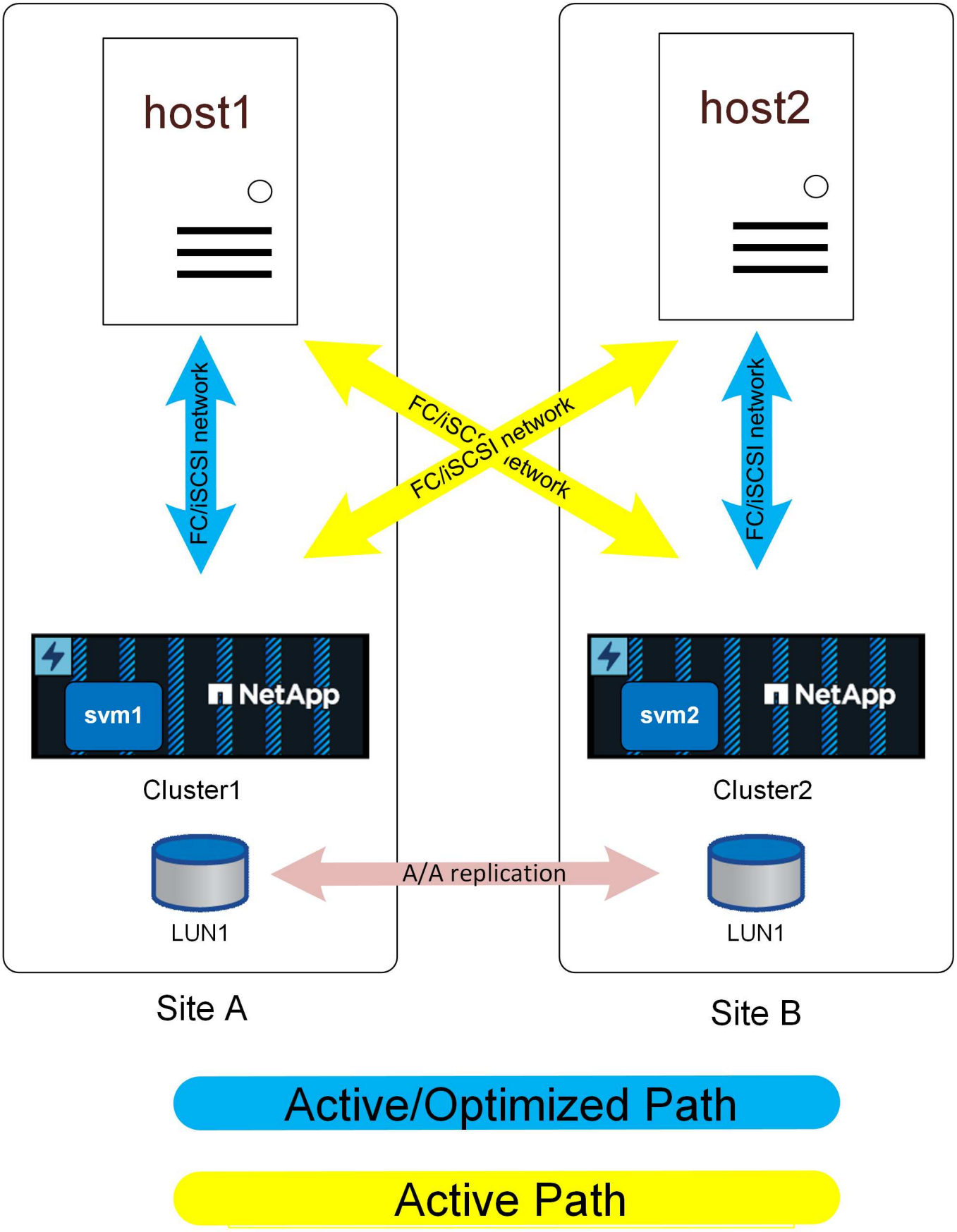
Il primo passo è la normale configurazione di igroup. Ogni LUN deve essere mappato a un igroup che contiene gli ID WWN/iSCSI degli host che devono accedere a quel LUN. Controlla quale host ha *accesso* a un LUN.

Il secondo passo, opzionale, consiste nel configurare la prossimità all'host. Questo non controlla l'accesso, controlla *priority*.

Ad esempio, un host del sito A potrebbe essere configurato in modo da accedere a una LUN protetta dalla sincronizzazione attiva di SnapMirror e, poiché la SAN è estesa tra i siti, i percorsi sono disponibili per tale LUN utilizzando lo storage sul sito A o lo storage sul sito B.

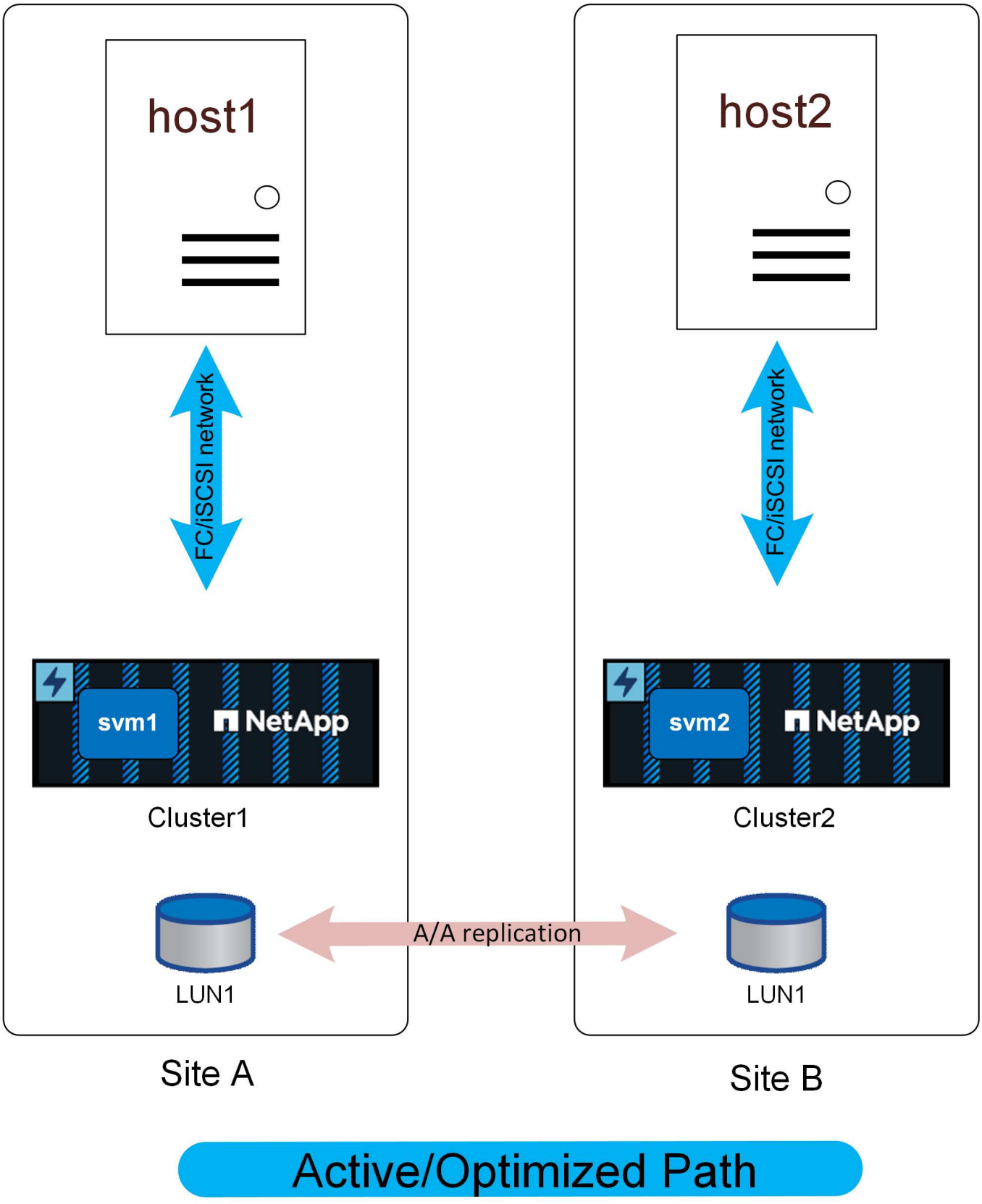
Senza impostazioni di prossimità, l'host utilizzerà entrambi i sistemi storage allo stesso modo perché entrambi i sistemi storage pubblicizzeranno i percorsi attivi/ottimizzati. Se la latenza SAN e/o la larghezza di banda tra i siti è limitata, questa operazione potrebbe non essere disattivabile e potrebbe essere necessario assicurarsi che durante il normale funzionamento ogni host utilizzi preferenzialmente i percorsi verso il sistema di storage locale. Viene configurato aggiungendo l'ID WWN/iSCSI dell'host al cluster locale come host prossimale. Questa operazione può essere eseguita dalla CLI o da SystemManager.

I percorsi vengono visualizzati come mostrato di seguito quando è stata configurata la prossimità host.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
NetApp
svm2
NetApp
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

Accesso non uniforme

Una rete di accesso non uniforme significa che ogni host ha solo accesso alle porte sul sistema di storage locale. LA SAN non è estesa a più siti (o presenta errori nei domini dello stesso sito).



Il vantaggio principale di questo approccio è la semplicità DELLE SAN, eliminando l'esigenza di stretching di una SAN via rete. Alcuni clienti non dispongono di connettività a latenza sufficientemente bassa tra i siti o non dispongono dell'infrastruttura per il tunnel del traffico FC SAN su una rete intersito.

Lo svantaggio legato all'accesso non uniforme è che alcuni scenari di errore, inclusa la perdita del collegamento di replica, provocheranno la perdita dell'accesso allo storage da parte di alcuni host. In caso di interruzione della connettività dello storage locale, le applicazioni eseguite come istanze singole, come ad esempio i database non in cluster, eseguiti in maniera intrinseca solo su un singolo host in uno qualsiasi dei supporti di montaggio, si guasterebbero. I dati sarebbero comunque protetti, ma il server di database non avrebbe più accesso. Dovrebbe essere riavviato su un sito remoto, preferibilmente tramite un processo automatizzato. Ad esempio, VMware ha è in grado di rilevare una situazione di tutti i percorsi verso l'esterno su un server e di riavviare una macchina virtuale su un altro server in cui sono disponibili i percorsi.

Al contrario, un'applicazione in cluster come Oracle RAC può offrire un servizio disponibile contemporaneamente in due siti diversi. Perdere un sito non significa perdere il servizio dell'applicazione nel suo complesso. Le istanze sono ancora disponibili e in esecuzione nel sito sopravvissuto.

In molti casi, l'overhead della latenza aggiuntivo di un'applicazione che accede allo storage attraverso un collegamento da sito a sito sarebbe inaccettabile. Ciò significa che la migliore disponibilità di una rete uniforme è minima, poiché la perdita di storage su un sito comporterebbe comunque la necessità di arrestare i servizi sul sito in cui si è verificato l'errore.

Esistono percorsi ridondanti attraverso il cluster locale non mostrati in questi diagrammi per semplicità. I sistemi di storage ONTAP sono ad alta disponibilità, pertanto un guasto a un controller non dovrebbe causare guasti nel sito. Ciò dovrebbe comportare solo una modifica dei percorsi locali utilizzati nel sito interessato.

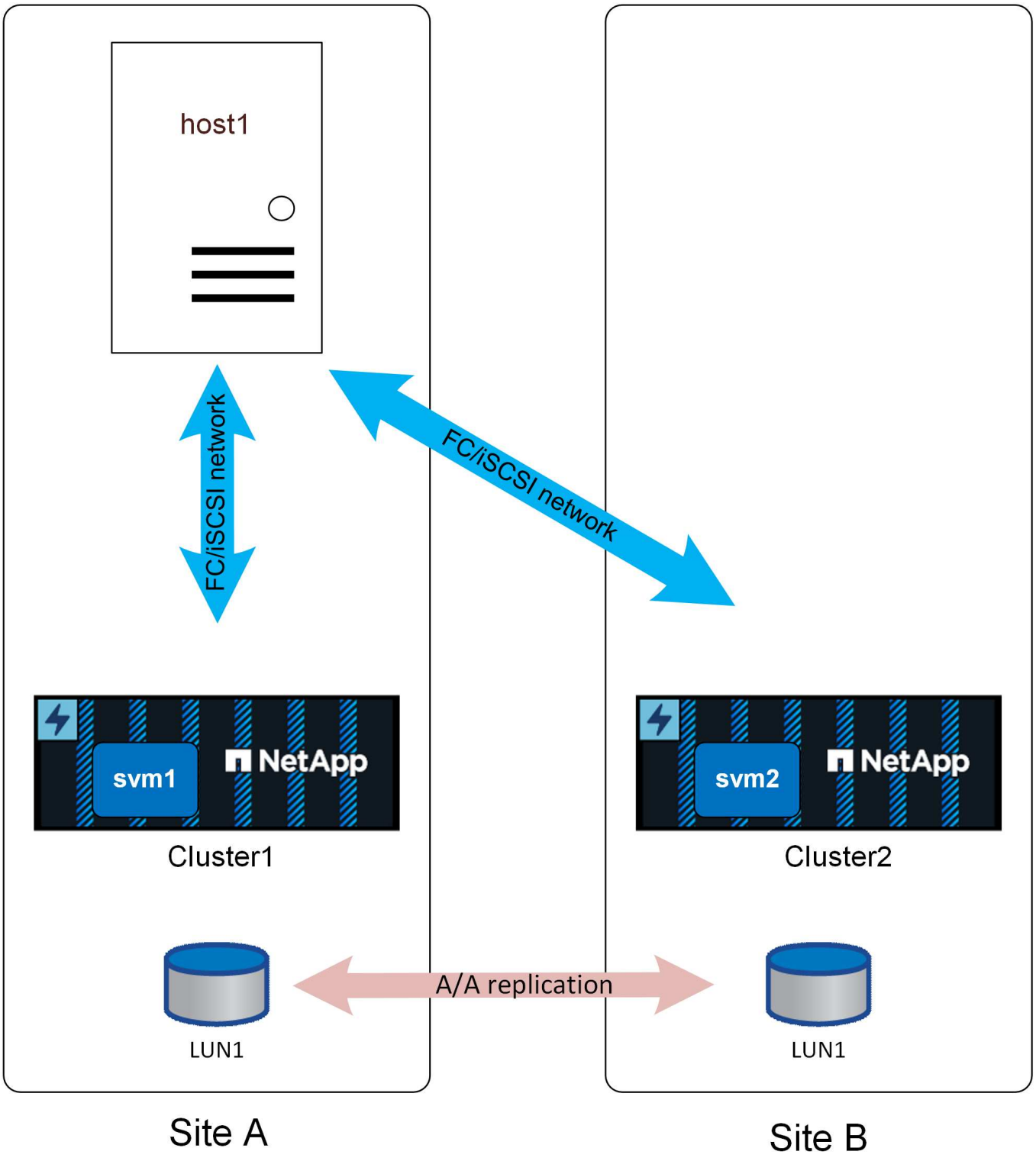
Panoramica

SQL Server può essere configurato per funzionare con la sincronizzazione attiva di SnapMirror in diversi modi. La risposta giusta dipende dalla connettività di rete disponibile, dai requisiti RPO e dai requisiti di disponibilità.

Istanza standalone di SQL Server

Le procedure consigliate per il layout dei file e la configurazione del server sono identiche a quelle consigliate nella "Server SQL su ONTAP" documentazione.

Con una configurazione standalone, SQL Server potrebbe essere eseguito solo in un sito. Presumibilmente "uniforme"si userebbe l'accesso.



Grazie a un accesso uniforme, un guasto dello storage in uno dei siti non interromperebbe le operazioni di database. Un errore completo del sito che includeva il server di database comporterebbe, ovviamente, un'interruzione del servizio.

Alcuni clienti potrebbero configurare un sistema operativo in esecuzione nel sito remoto con un'installazione di SQL Server preconfigurata, aggiornata con una versione di build equivalente a quella dell'istanza di produzione. Il failover richiede l'attivazione dell'istanza standalone di SQL Server nel sito alternativo, il rilevamento dei LUN e l'avvio del database. È possibile automatizzare il processo completo con il cmdlet di

Windows PowerShell, in quanto non sono richieste operazioni dal lato storage.

"Non uniforme" è possibile utilizzare anche l'accesso, ma il risultato sarebbe un'interruzione del database se il sistema storage in cui si trovava il server del database non avesse avuto esito positivo, perché il database non avrebbe percorsi disponibili per lo storage. In alcuni casi ciò può ancora essere accettabile. La sincronizzazione attiva di SnapMirror garantirebbe comunque una data Protection con RPO=0 e, in caso di guasto del sito, la copia restante sarebbe attiva e pronta a riprendere le operazioni seguendo la stessa procedura utilizzata con un accesso uniforme descritta sopra.

Un processo di failover semplice e automatizzato può essere configurato più facilmente con l'uso di un host virtualizzato. Ad esempio, se i file di dati di SQL Server vengono replicati in modo sincrono su storage secondario insieme a un VMDK di avvio, in caso di disastro è possibile attivare l'ambiente completo nel sito alternativo. Un amministratore può attivare manualmente l'host nel sito rimasto o automatizzare il processo tramite un servizio come VMware ha.

Istanza del cluster di failover di SQL Server

Le istanze di failover di SQL Server possono essere ospitate anche in un cluster di failover di Windows in esecuzione su un server fisico o virtuale come sistema operativo guest. Questa architettura multi-host offre un'istanza di SQL Server e la resilienza dello storage. Tale implementazione è utile in ambienti a domanda elevata che richiedono solidi processi di failover senza rinunciare a prestazioni avanzate. In una configurazione del cluster di failover, quando un host o uno storage primario viene colpito, SQL Services eseguirà il failover sull'host secondario e, allo stesso tempo, lo storage secondario sarà disponibile per servire io. Non è richiesto alcuno script di automazione o intervento dell'amministratore.

Scenari di errore

La pianificazione di un'architettura completa dell'applicazione SnapMirror Active Sync richiede la comprensione del modo in cui SM-AS risponderà in vari scenari di failover pianificati e non pianificati.

Per gli esempi seguenti, si supponga che il sito A sia configurato come sito preferito.

Interruzione della connettività di replica

Se la replica SM-AS viene interrotta, non è possibile completare la scrittura io perché sarebbe impossibile per un cluster replicare le modifiche al sito opposto.

Sito A (sito preferito)

Il risultato dell'errore del collegamento di replica sul sito preferito sarà una pausa di circa 15 secondi nell'elaborazione io in scrittura, poiché ONTAP ritenta le operazioni di scrittura replicate prima di determinare che il collegamento di replica è veramente irraggiungibile. Trascorsi i 15 secondi, il sistema del sito A riprende l'elaborazione io in lettura e scrittura. I percorsi SAN non vengono modificati e i LUN rimangono online.

Sito B

Poiché il sito B non è il sito preferito di sincronizzazione attiva SnapMirror, i relativi percorsi LUN non saranno più disponibili dopo circa 15 secondi.

Errore del sistema di storage

Il risultato di un errore del sistema di storage è quasi identico al risultato della perdita del collegamento di replica. Il sito sopravvissuto dovrebbe subire una pausa io di circa 15 secondi. Trascorso questo periodo di 15

secondi, io riprenderà sul sito come di consueto.

Perdita del mediatore

Il servizio di mediazione non controlla direttamente le operazioni di storage. Funziona come un percorso di controllo alternativo tra cluster. Esiste principalmente per automatizzare il failover senza il rischio di uno scenario split-brain. Durante l'utilizzo normale, ogni cluster replica le modifiche al partner e pertanto ogni cluster può verificare che il cluster partner sia online e fornisca i dati. Se il collegamento di replica non è riuscito, la replica viene interrotta.

Il motivo per cui è necessario un mediatore per il failover automatizzato sicuro è perché altrimenti sarebbe impossibile per un cluster di storage determinare se la perdita di comunicazione bidirezionale fosse il risultato di un'interruzione della rete o di un errore effettivo dello storage.

Il mediatore fornisce un percorso alternativo per ciascun cluster per verificare lo stato di salute del partner. Gli scenari sono i seguenti:

- Se un cluster può contattare direttamente il partner, i servizi di replica sono operativi. Non è richiesta alcuna azione.
- Se un sito preferito non può contattare direttamente il proprio partner o tramite il mediatore, presuppone che il partner sia effettivamente non disponibile oppure è stato isolato e ha portato i percorsi LUN offline. Il sito preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'io in lettura e in scrittura.
- Se un sito non preferito non può contattare direttamente il proprio partner, ma può contattarlo tramite il mediatore, prenderà i suoi percorsi offline e attenderà il ritorno della connessione di replica.
- Se un sito non preferito non può contattare direttamente il proprio partner o tramite un mediatore operativo, supporterà che il partner sia effettivamente non disponibile, oppure che sia stato isolato e che abbia portato i percorsi LUN offline. Il sito non preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'i/o in lettura e scrittura. Assumerà il ruolo dell'origine della replica e diventerà il nuovo sito preferito.

Se il mediatore non è completamente disponibile:

- In caso di guasto dei servizi di replica per qualsiasi motivo, incluso un guasto del sito o del sistema storage non preferito, il sito preferito rilascerà lo stato RPO=0 e riprenderà l'elaborazione i/o in lettura e scrittura. Il sito non preferito prenderà i suoi percorsi offline.
- Il guasto del sito preferito causerà un'interruzione poiché il sito non preferito non sarà in grado di verificare che il sito opposto sia effettivamente offline e quindi non sarebbe sicuro per il sito non preferito riprendere i servizi.

Ripristino dei servizi in corso

Dopo aver risolto un errore, come il ripristino della connettività da sito a sito o l'accensione di un sistema guasto, gli endpoint di sincronizzazione attivi di SnapMirror rilevano automaticamente la presenza di una relazione di replica difettosa e la riportano allo stato RPO=0. Una volta ristabilita la replica sincrona, i percorsi non riusciti torneranno in linea.

In molti casi, le applicazioni in cluster rilevano automaticamente la restituzione dei percorsi non riusciti e tali applicazioni tornano online. In altri casi, potrebbe essere necessaria una scansione SAN a livello di host oppure potrebbe essere necessario riportare le applicazioni online manualmente. Dipende dall'applicazione e dal modo in cui è configurata, e in generale tali attività possono essere facilmente automatizzate. ONTAP si sta auto-riparando e non deve richiedere alcun intervento da parte dell'utente per riprendere le operazioni di storage RPO = 0 KB.

Failover manuale

La modifica del sito preferito richiede un'operazione semplice. I/o si fermeranno per un secondo o due come autorità sugli switch del comportamento di replica tra i cluster, ma in caso contrario i/o non vengono influenzati.

Sicurezza dei dati

La protezione di un ambiente di database SQL Server è un'operazione multidimensionale che va oltre la gestione del database stesso. ONTAP offre diverse funzioni esclusive progettate per proteggere gli aspetti dello storage dell'infrastruttura di database.

Copie Snapshot

Le snapshot di storage sono repliche point-in-time dei dati di destinazione. L'implementazione di ONTAP include le funzionalità per impostare varie policy e memorizzare fino a 1024 snapshot per volume. Le Snapshot in ONTAP sono efficienti in termini di spazio. Lo spazio viene consumato solo quando viene modificato il set di dati originale. Sono anche di sola lettura. Uno snapshot può essere eliminato, ma non può essere modificato.

In alcuni casi, le snapshot possono essere pianificate direttamente su ONTAP. In altri casi, software come SnapCenter potrebbe essere necessario per orchestrare le operazioni dell'applicazione o del sistema operativo prima di creare snapshot. Qualunque sia l'approccio migliore per i tuoi workload, un'aggressiva strategia di snapshot può garantire sicurezza dei dati tramite un accesso frequente e facilmente accessibile ai backup di ogni elemento, dalle LUN di avvio ai database mission-critical.

> Un volume flessibile ONTAP, o più semplicemente, un volume non è sinonimo di LUN. I volumi sono container di gestione per dati come file o LUN. Ad esempio, un database può essere posizionato su un set di stripe da 8 LUN, con tutti i LUN contenuti in un singolo volume.

Per ulteriori informazioni sulle istantanee, consultare la "Documentazione ONTAP."

Snapshot a prova di manomissione

A partire da ONTAP 9.12.1, gli snapshot non sono solo di lettura, ma possono anche essere protetti da eliminazioni accidentali o intenzionali. La funzione è denominata istantanee antimanomissione. È possibile impostare e applicare un periodo di conservazione tramite policy snapshot. Gli snapshot risultanti non possono essere eliminati fino a quando non hanno raggiunto la data di scadenza. Non sono presenti sostituzioni amministrative o del centro di supporto.

In questo modo, un intruso, un malintenzionato o persino un attacco ransomware non sono in grado di compromettere i backup, anche nel caso in cui abbiano accesso al sistema ONTAP stesso. Se combinato con una pianificazione degli snapshot frequente, offre una data Protection estremamente potente con un RPO molto basso.

> Non è possibile eseguire il tiering delle snapshot a prova di manomissione utilizzando il fabric pool.

Per ulteriori informazioni sugli snapshot antimanomissione, consultare la "Documentazione ONTAP."

> Nella nuova piattaforma ASA, replicare le snapshot nel cluster remoto utilizzando il criterio *vault*, quindi bloccare la destinazione per rendere le snapshot a prova di manomissione.

Replica SnapMirror

Gli snapshot possono anche essere replicati su un sistema remoto. Sono inclusi gli snapshot antimanomissione, in cui il periodo di conservazione viene applicato e applicato sul sistema remoto. Come risultato otterrai gli stessi vantaggi di protezione dei dati delle snapshot locali, ma i dati verranno posizionati in un secondo storage array. In questo modo si garantisce che la distruzione dell'array originale non comprometta i backup.

Un secondo sistema apre anche nuove opzioni per la sicurezza amministrativa. Ad esempio, alcuni clienti NetApp segregano le credenziali di autenticazione per i sistemi di storage primario e secondario. Nessun utente amministrativo singolo ha accesso a entrambi i sistemi, il che significa che un amministratore malintenzionato non può eliminare tutte le copie dei dati.

Per ulteriori informazioni su SnapMirror, consultare la "Documentazione ONTAP."

Macchine virtuali di storage

Un sistema di storage ONTAP appena configurato è simile a un server VMware ESX appena configurato, perché nessuno di questi può supportare gli utenti fino alla creazione di una macchina virtuale. Con ONTAP viene creata una Storage Virtual Machine (SVM) che diventa l'unità di gestione dello storage più base. Ciascuna SVM dispone di risorse di storage, configurazioni di protocolli, indirizzi IP e WWN FCP. Questa è la base della multi-tenancy di ONTAP.

Ad esempio, è possibile configurare una SVM per i carichi di lavoro di produzione critici e una seconda SVM su un segmento di rete diverso per le attività di sviluppo. Quindi, è possibile limitare l'accesso alla SVM di produzione a determinati amministratori, garantendo al contempo agli sviluppatori un controllo più esteso sulle risorse storage nella SVM di sviluppo. Potrebbe anche essere necessario fornire una terza SVM ai tuoi team finanziari e delle risorse umane per memorizzare dati particolarmente critici solo per gli occhi.

Per ulteriori informazioni sulle SVM, consulta la "Documentazione ONTAP."

RBAC amministrativo

ONTAP offre un potente role-based access control (RBAC) per gli accessi amministrativi. Alcuni amministratori potrebbero aver bisogno di un accesso completo al cluster, altri invece potrebbero aver bisogno solo dell'accesso a determinate SVM. Il personale avanzato dell'helpdesk potrebbe aver bisogno di aumentare le dimensioni dei volumi. Il risultato è la possibilità di concedere agli utenti amministrativi l'accesso necessario per eseguire le proprie responsabilità lavorative, e niente di più. Inoltre, è possibile proteggere questi accessi utilizzando PKI di vari fornitori, limitare l'accesso solo alle chiavi ssh e applicare blocchi dei tentativi di accesso non riusciti.

Per ulteriori informazioni sul controllo dell'accesso amministrativo, consultare la "Documentazione ONTAP."

Autenticazione multifattore (MFA)

ONTAP e alcuni altri prodotti NetApp supportano ora l'autenticazione a più fattori (MFA) utilizzando una vasta gamma di metodi. Il risultato è un nome utente/password compromesso da solo non è un thread di sicurezza senza i dati del secondo fattore, come un FOB o un'applicazione per smartphone.

Per ulteriori informazioni sull'autenticazione multifattore (MFA), vedere "Documentazione ONTAP."

RBAC API

L'automazione richiede chiamate API, ma non tutti gli strumenti richiedono un accesso amministrativo completo. Per contribuire a proteggere i sistemi di automazione, RBAC è disponibile anche a livello di API. È possibile limitare gli account utente di automazione alle chiamate API richieste. Ad esempio, il software di monitoraggio non richiede l'accesso alle modifiche, ma solo l'accesso in lettura. I workflow che forniscono storage non hanno bisogno della capacità di eliminare lo storage.

Per ulteriori informazioni su RBAC API, consulta la "Documentazione ONTAP."

Verifica multi-admin (MAV)

L'autenticazione a più "fattori" può essere ulteriormente eseguita richiedendo l'approvazione di determinate attività da parte di due amministratori diversi, ciascuno con le proprie credenziali. Ciò include la modifica delle autorizzazioni di accesso, l'esecuzione dei comandi diagnostici e l'eliminazione dei dati.

Per ulteriori informazioni sulla verifica multi-admin (MAV), vedere "Documentazione ONTAP."

MySQL

Panoramica

MySQL e le sue varianti, tra cui MariaDB e Percona MySQL, è il database più diffuso al mondo.

Questa documentazione su ONTAP e il database MySQL sostituisce il database *TR-4722: MySQL pubblicato in precedenza sulle Best practice di ONTAP.*

ONTAP è una piattaforma ideale per database MySQL, perché ONTAP è letteralmente progettato per database. Sono state create numerose funzionalità come le ottimizzazioni della latenza io random per la qualità del servizio avanzata fino alle funzionalità FlexClone di base per rispondere specificamente alle esigenze dei carichi di lavoro dei database.

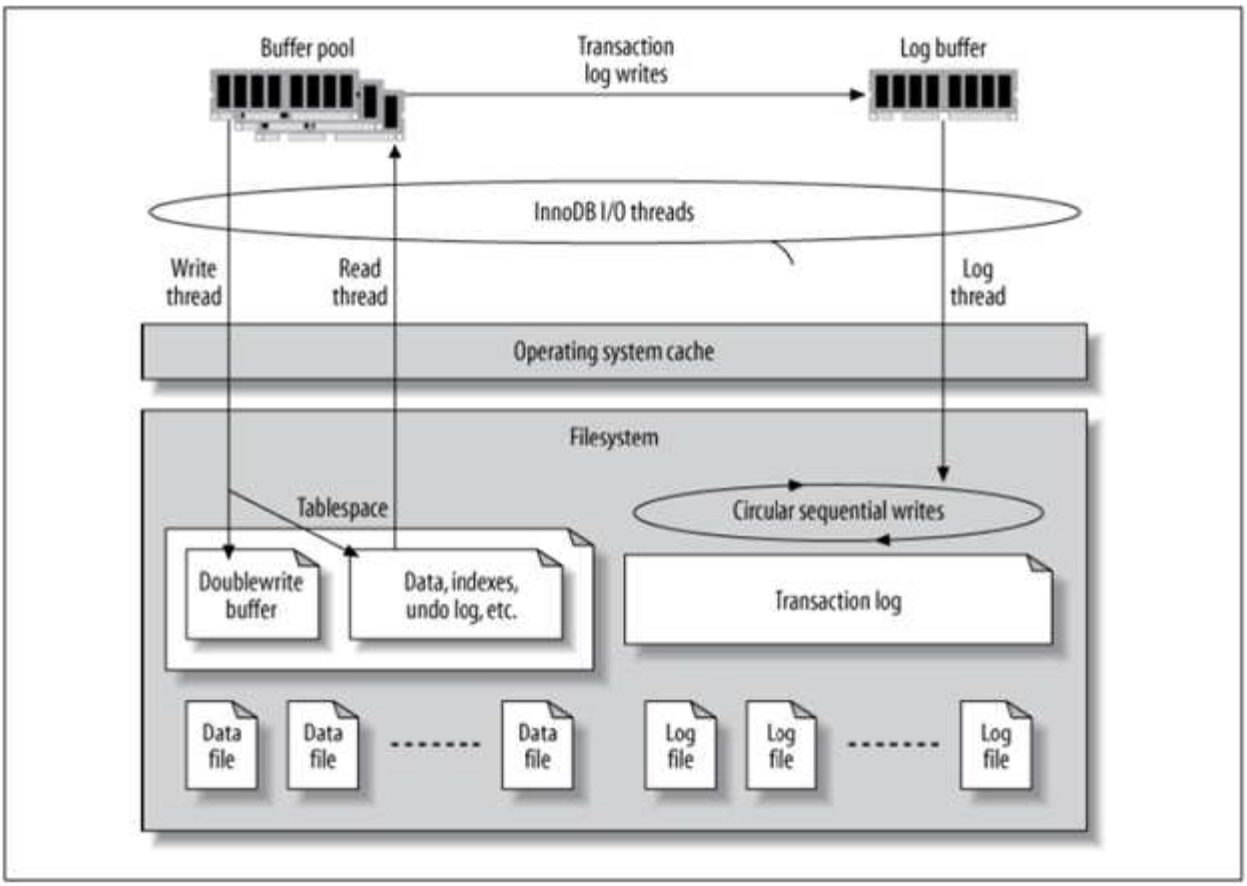
Funzioni aggiuntive come gli aggiornamenti senza interruzioni, (inclusa la sostituzione dello storage) garantiscono la disponibilità dei database critici. Puoi anche disporre di un disaster recovery istantaneo per ambienti di grandi dimensioni tramite MetroCluster o selezionare database tramite la sincronizzazione attiva di SnapMirror.

Soprattutto, ONTAP offre prestazioni senza pari con la possibilità di dimensionare la soluzione in base alle proprie esigenze specifiche. I nostri sistemi high-end possono offrire oltre 1M IOPS con latenze misurate in microsecondi, ma se ti servono solo 100K IOPS, puoi dimensionare correttamente la tua soluzione storage con un controller più piccolo che esegue ancora lo stesso sistema operativo per lo storage.

Configurazione del database

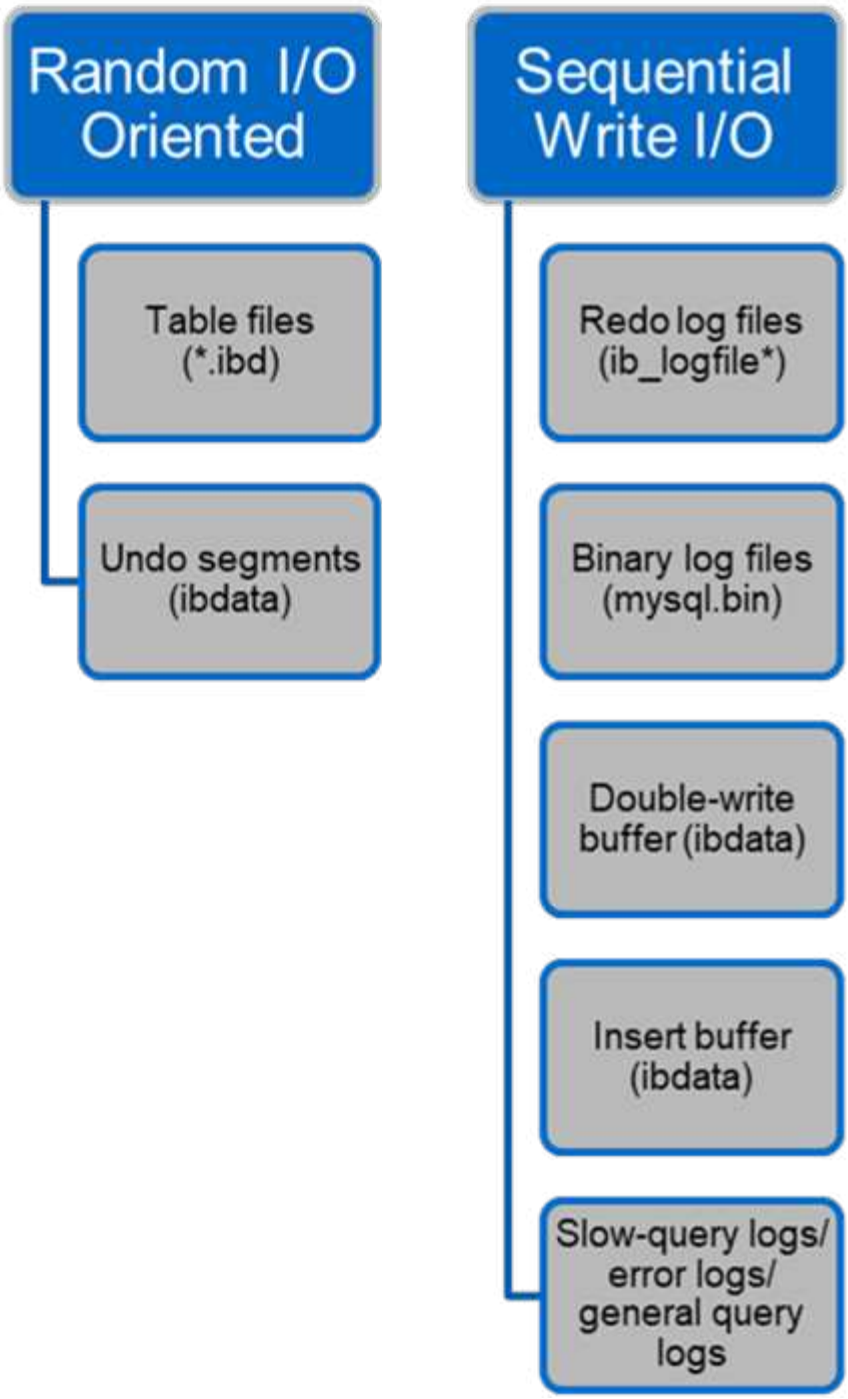
Struttura dei file

InnoDB funge da livello intermedio tra lo storage e il server MySQL, e memorizza i dati nelle unità.



I/o MySQL è suddiviso in due tipi:

- I/o di file casuali
- I/o di file sequenziale



I file di dati vengono letti e sovrascritti in modo casuale, con conseguente aumento degli IOPS. Pertanto, si consiglia di utilizzare l'unità SSD.

I file di log di ripristino e i file di log binari sono registri transazionali. Vengono scritti in sequenza, così potrai ottenere buone performance sul disco HDD con cache in scrittura. Al momento del ripristino si verifica una lettura sequenziale, che raramente causa problemi di prestazioni, poiché le dimensioni dei file di registro sono in genere inferiori ai file di dati e le letture sequenziali sono più veloci delle letture casuali (che si verificano sui file di dati).

Il buffer double-write è una caratteristica speciale di InnoDB. InnoDB prima scrive le pagine svuotate nel buffer di doppia scrittura e poi scrive le pagine nelle posizioni corrette sui file di dati. Questo processo impedisce il

danneggiamento della pagina. Senza il buffer di scrittura doppia, la pagina potrebbe danneggiarsi se si verifica un'interruzione dell'alimentazione durante il processo di scrittura su unità. La scrittura nel buffer double-write è sequenziale, pertanto è altamente ottimizzato per gli HDD. Al momento del ripristino vengono eseguite letture sequenziali.

Poiché la NVRAM ONTAP fornisce già la protezione in scrittura, non è necessario il doppio buffer in scrittura. MySQL ha un parametro, `skip_innodb_doublewrite`, per disattivare il buffer di doppia scrittura. Questa funzione può migliorare notevolmente le prestazioni.

Il buffer insert è anche una caratteristica speciale di InnoDB. Se i blocchi di indice secondari non univoci non sono in memoria, InnoDB inserisce le voci nel buffer di inserimento per evitare operazioni di i/o casuali. Periodicamente, il buffer di inserimento viene Unito agli alberi di indice secondari nel database. Il buffer di inserimento riduce il numero di operazioni di i/o unendo le richieste di i/o allo stesso blocco; le operazioni di i/o casuali possono essere sequenziali. Anche il buffer di inserimento è altamente ottimizzato per gli HDD. Durante le normali operazioni, vengono eseguite operazioni di scrittura e lettura sequenziali.

I segmenti di annullamento sono orientati all'i/o casuale. Per garantire la concorrenza multi-versione (MVCC), InnoDB deve registrare le vecchie immagini nei segmenti di annullamento. La lettura delle immagini precedenti dai segmenti di annullamento richiede letture casuali. Se si esegue una transazione lunga con letture ripetibili (come mysqldump, una singola transazione) o si esegue una query lunga, è possibile che si verifichino letture casuali. Pertanto, in questo caso è preferibile memorizzare i segmenti di annullamento negli SSD. Se si eseguono solo transazioni o query brevi, le letture casuali non costituiscono un problema.

NetApp consiglia il seguente layout di progettazione dello storage a causa delle caratteristiche i/o di InnoDB.

- Un unico volume per memorizzare i file di MySQL orientati ai/o casuali e sequenziali
- Un altro volume per memorizzare i file di MySQL orientati a i/o puramente sequenziali

Questo layout aiuta inoltre a progettare politiche e strategie di protezione dei dati.

Parametri di configurazione

NetApp consiglia alcuni importanti parametri di configurazione di MySQL per ottenere prestazioni ottimali.

Parametri	Valori
innodb_log_file_size	256M
innodb_flush_log_at_trx_commit	2
innodb_doublewrite	0
innodb_flush_method	fsync
innodb_buffer_pool_size	11G
innodb_io_capacity	8192
innodb_buffer_pool_instances	8
innodb_lru_scan_depth	8192
open_file_limit	65535

Per impostare i parametri descritti in questa sezione, è necessario modificarli nel file di configurazione MySQL (my.cnf). Le Best practice di NetApp sono il risultato di test eseguiti internamente.

innodb_log_file_size

La scelta della dimensione corretta per il file di log InnoDB è importante per le operazioni di scrittura e per avere un tempo di ripristino decente dopo un arresto anomalo del server.

Poiché molte transazioni sono registrate nel file, la dimensione del file di registro è importante per le operazioni di scrittura. Quando i record vengono modificati, la modifica non viene immediatamente riscritta nello spazio di tabella. La modifica viene invece registrata alla fine del file di registro e la pagina viene contrassegnata come sporca. InnoDB utilizza il proprio registro per convertire l'i/o casuale in i/o sequenziale

Quando il log è pieno, la pagina sporca viene scritta nello spazio di tabella in sequenza per liberare spazio nel file di log. Ad esempio, si supponga che un server si blocchi nel corso di una transazione e che le operazioni di scrittura vengano registrate solo nel file di registro. Prima che il server possa tornare attivo, deve passare attraverso una fase di recupero in cui vengono riprodotte le modifiche registrate nel file di registro. Maggiore è il numero di voci presenti nel file di registro, maggiore sarà il tempo necessario al server per il ripristino.

In questo esempio, la dimensione del file di registro influisce sia sul tempo di ripristino che sulle prestazioni di scrittura. Quando si sceglie il numero giusto per la dimensione del file di registro, bilanciare il tempo di ripristino rispetto alle prestazioni di scrittura. In genere, qualsiasi valore compreso tra 128M e 512M è un buon valore.

innodb_flush_log_at_trx_commit

In caso di modifica dei dati, la modifica non viene immediatamente scritta nell'archivio.

I dati vengono invece registrati in un buffer di registro, che è una porzione di memoria allocata da InnoDB alle modifiche del buffer registrate nel file di registro. InnoDB svuota il buffer nel file di registro quando viene eseguito il commit di una transazione, quando il buffer diventa pieno o una volta al secondo, a seconda dell'evento che si verifica per primo. La variabile di configurazione che controlla questo processo è innodb_Flush_log_at_trx_commit. Le opzioni valore includono:

- Quando si imposta `innodb_flush_log_trx_at_commit=0`, InnoDB scrive i dati modificati (nel pool di buffer InnoDB) nel file di log (ib_logfile) e scarica il file di log (write to storage) ogni secondo. Tuttavia, non fa nulla quando la transazione è impegnata. Se si verifica un'interruzione dell'alimentazione o un arresto anomalo del sistema, nessuno dei dati non scaricati è recuperabile perché non vengono scritti né nel file di registro né nelle unità.
- Quando si imposta `innodb_flush_log_trx_commit=1`, InnoDB scrive il buffer di log nel log delle transazioni e lo svuota nello storage durevole per ogni transazione. Ad esempio, per tutti i commit delle transazioni, InnoDB scrive nel registro e quindi nello storage. La lentezza dello storage influisce negativamente sulle performance, ad esempio riducendo il numero di transazioni InnoDB al secondo.
- Quando si imposta `innodb_flush_log_trx_commit=2`, InnoDB scrive il buffer di log nel file di log ad ogni commit; tuttavia, non scrive dati nell'archivio. InnoDB scarica i dati una volta al secondo. Anche in caso di interruzione dell'alimentazione o arresto anomalo del sistema, i dati dell'opzione 2 sono disponibili nel file di registro ed è recuperabile.

Se l'obiettivo principale è la prestazione, impostare il valore su 2. Poiché InnoDB scrive sui dischi una volta al secondo, non per ogni commit delle transazioni, le performance migliorano in modo significativo. Se si verifica un'interruzione dell'alimentazione o un arresto anomalo, i dati possono essere recuperati dal registro delle transazioni.

Se l'obiettivo principale è la sicurezza dei dati, impostare il valore su 1 in modo che per ogni commit di transazione, InnoDB si scarichi sulle unità. Tuttavia, le prestazioni potrebbero risentirne.

NetApp recommended impostare il valore innodb_Flush_log_trx_commit su 2 per ottenere prestazioni migliori.

innodb_doublewrite

Quando `innodb_doublewrite` È attivato (impostazione predefinita), InnoDB memorizza tutti i dati due volte: Prima nel buffer di doppia scrittura e poi nei file di dati effettivi.

È possibile disattivare questo parametro con `--skip-innodb_doublewrite` per i benchmark o quando siete più preoccupati per le prestazioni superiori che l'integrità dei dati o possibili guasti. InnoDB utilizza una tecnica di scaricamento file chiamata double-write. Prima di scrivere le pagine nei file di dati, InnoDB le scrive in un'area contigua denominata buffer double-write. Una volta completata la scrittura e lo scarico nel buffer di doppia scrittura, InnoDB scrive le pagine nelle posizioni corrette nel file di dati. Se il sistema operativo o un processo mysqld si blocca durante la scrittura di una pagina, InnoDB può in seguito trovare una buona copia della pagina dal buffer di doppia scrittura durante il recupero del crash.

NetApp recommended disabilitare il buffer double-write. La NVRAM ONTAP svolge la stessa funzione. Il doppio buffering danneggia inutilmente le prestazioni.

innodb_buffer_pool_size

Il pool di buffer InnoDB è la parte più importante di qualsiasi attività di ottimizzazione.

InnoDB si affida in gran parte al pool di buffer per la memorizzazione nella cache degli indici e il reming dei dati, all'indice hash adattivo, al buffer insert e a molte altre strutture di dati utilizzate internamente. Il pool di buffer memorizza inoltre le modifiche ai dati in modo che le operazioni di scrittura non debbano essere eseguite immediatamente nello storage, migliorando così le prestazioni. Il pool di buffer è parte integrante di InnoDB e le sue dimensioni devono essere regolate di conseguenza. Per impostare le dimensioni del pool di buffer, tenere conto dei seguenti fattori:

- Per una macchina dedicata solo InnoDB, impostare la dimensione del pool di buffer su 80% o più della RAM disponibile.
- Se non si tratta di un server dedicato MySQL, impostare la dimensione al 50% della RAM.

innodb_flush_method

Il parametro innodb_Flush_Method specifica come InnoDB apre e svuota i file di log e di dati.

Ottimizzazioni

Nell'ottimizzazione InnoDB, l'impostazione di questo parametro modifica le prestazioni del database, se applicabile.

Le seguenti opzioni consentono di svuotare i file tramite InnoDB:

- `fsync`. InnoDB utilizza `fsync()` chiamata di sistema per cancellare sia i file di dati che i file di registro.

Questa opzione è l'impostazione predefinita.

- `O_DSYNC`. InnoDB utilizza `O_DSYNC` possibilità di aprire e svuotare i file di log e fsync() per svuotare i file di dati. InnoDB non utilizza `O_DSYNC` Direttamente, perché ci sono stati problemi con esso su molte varietà di UNIX.
- `O_DIRECT`. InnoDB utilizza `O_DIRECT` (oppure `directio()` Su Solaris) per aprire i file di dati e gli usi `fsync()` per cancellare sia i file di dati che i file di registro. Questa opzione è disponibile su alcune versioni di GNU/Linux, FreeBSD e Solaris.
- `O_DIRECT_NO_FSYNC`. InnoDB utilizza `O_DIRECT` Durante lo spurgo dell'i/o, tuttavia, salta `fsync()` chiamata di sistema successiva. Questa opzione non è adatta per alcuni tipi di file system (ad esempio, XFS). Se non si è certi che il file system richieda un `fsync()` chiamata di sistema, ad esempio per conservare tutti i metadati dei file, utilizzare `O_DIRECT` invece.

Osservazione

Nei test di laboratorio di NetApp, il `fsync` L'opzione predefinita è stata utilizzata su NFS e SAN ed è stata un'improvvisazione per le prestazioni eccezionale rispetto a `O_DIRECT`. Mentre si utilizza il metodo di lavaggio come `O_DIRECT` Con ONTAP, abbiamo osservato che il client scrive molte scritture a byte singolo al margine del blocco 4096 in modo seriale. Queste operazioni di scrittura hanno aumentato la latenza sulla rete e degradato le performance.

innodb_io_capacity

Nel plug-in InnoDB è stato aggiunto un nuovo parametro chiamato innodb_io_Capacity da MySQL 5,7.

Controlla il numero massimo di IOPS eseguiti da InnoDB (che include la velocità di scaricamento delle pagine sporche e la dimensione batch del buffer di inserimento [ibuf]). Il parametro innodb_io_Capacity imposta un limite massimo per le IOPS da parte delle attività in background di InnoDB, come il lavaggio delle pagine dal pool di buffer e l'Unione dei dati dal buffer di modifica.

Impostare il parametro innodb_io_Capacity sul numero approssimativo di operazioni di i/o che il sistema può eseguire al secondo. Idealmente, mantenere l'impostazione più bassa possibile, ma non così bassa che le attività in background rallentano. Se l'impostazione è troppo alta, i dati vengono rimossi dal pool di buffer e il buffer di inserimento troppo rapidamente per la memorizzazione nella cache, per fornire un vantaggio significativo.

NetApp consiglia che, se si utilizza questa impostazione su NFS, analizzi il risultato del test di IOPS (SysBench/FiO) e imposti il parametro di conseguenza. Utilizzare il valore più piccolo possibile per lo spurgo e lo spurgo per continuare a meno che non vengano visualizzate pagine modificate o sporche di quanto si desidera nel pool di buffer InnoDB.

Non utilizzare valori estremi come 20.000 o più a meno che non si sia dimostrato che valori inferiori non sono sufficienti per il carico di lavoro.

Il parametro InnoDB_io_Capacity regola le velocità di lavaggio e i/o correlati

È possibile danneggiare seriamente le prestazioni impostando questo parametro o il parametro innodb_io_Capacity_max troppo alto e wastin

innodb_lru_scan_depth

Il `innodb_lru_scan_depth` Parametro influenza gli algoritmi e le euristiche dell'operazione di scaricamento per il pool di buffer InnoDB.

Questo parametro è principalmente di interesse per gli esperti di performance che ottimizzano i carichi di lavoro i/o-intensive. Per ogni istanza del pool di buffer, questo parametro specifica fino a che punto nell'elenco di pagine LRU (Last Recently Used) il thread di pulitura della pagina deve continuare la scansione, cercando le pagine sporche da eliminare. Questa operazione in background viene eseguita una volta al secondo.

È possibile regolare il valore verso l'alto o verso il basso per ridurre al minimo il numero di pagine libere. Non impostare un valore molto superiore al necessario, poiché le scansioni possono avere un costo significativo in termini di prestazioni. Inoltre, è consigliabile regolare questo parametro quando si modifica il numero di istanze del pool di buffer, perché `innodb_lru_scan_depth * innodb_buffer_pool_instances` definisce la quantità di lavoro eseguito dal filo del pulitore di pagina ogni secondo.

Un'impostazione più piccola di quella predefinita è adatta per la maggior parte dei carichi di lavoro. Considerare l'aumento del valore solo se si dispone di capacità i/o di riserva con un workload tipico. Per contro, se un carico di lavoro con un numero elevato di operazioni di scrittura satura la capacità i/o, diminuirne il valore, soprattutto se si dispone di un pool di buffer di grandi dimensioni.

open_file_limits

Il `open_file_limits` parametro determina il numero di file che il sistema operativo consente a mysqld di aprire.

Il valore di questo parametro in fase di esecuzione è il valore reale consentito dal sistema e potrebbe essere diverso dal valore specificato all'avvio del server. Il valore è 0 sui sistemi in cui MySQL non può modificare il numero di file aperti. L'efficace `open_files_limit` il valore si basa sul valore specificato all'avvio del sistema (se presente) e sui valori di `max_connections` e. `table_open_cache` utilizzando queste formule:

- 10 + `max_connections` + (`table_open_cache` x 2)
- `max_connections` x 5
- Limite del sistema operativo se positivo
- Se il limite del sistema operativo è infinito: `open_files_limit` il valore viene specificato all'avvio; 5.000 se nessuno

Il server tenta di ottenere il numero di descrittori di file utilizzando il massimo di questi quattro valori. Se non è possibile ottenere molti descrittori, il server tenta di ottenere il numero di descrittori consentito dal sistema.

Configurazione dell'host

Containerizzazione

La containerizzazione dei database MySQL sta diventando sempre più diffusa.

La gestione di container a basso livello viene quasi sempre eseguita con Docker. Le piattaforme di gestione dei container come OpenShift e Kubernetes semplificano ulteriormente la gestione di ambienti container di grandi dimensioni. I vantaggi della containerizzazione includono una riduzione dei costi, poiché non è necessario acquistare una licenza per un hypervisor. Inoltre, i container consentono l'esecuzione di più

database isolati l'uno dall'altro, condividendo lo stesso kernel e sistema operativo sottostanti. È possibile eseguire il provisioning dei container in microsecondi.

NetApp offre Astra Trident per fornire funzionalità di gestione avanzate dello storage. Ad esempio, Astra Trident consente a un container creato in Kubernetes di eseguire il provisioning automatico del proprio storage nel Tier appropriato, applicare policy di esportazione, impostare policy di snapshot e persino clonare un container in un altro. Per ulteriori informazioni, vedere "Documentazione di Astra Trident".

NFSv3 tavoli con slot

NFSv3 le prestazioni di Linux dipendono da un parametro chiamato `tcp_max_slot_table_entries`.

Le tabelle degli slot TCP sono l'equivalente di NFSv3 della profondità della coda degli HBA (host Bus Adapter). Queste tabelle controllano il numero di operazioni NFS che possono essere in sospeso in qualsiasi momento. Il valore predefinito è di solito 16, che è troppo basso per ottenere prestazioni ottimali. Il problema opposto si verifica sui kernel Linux più recenti, che possono aumentare automaticamente il limite della tabella degli slot TCP a un livello che satura il server NFS con le richieste.

Per prestazioni ottimali e per evitare problemi di prestazioni, regolare i parametri del kernel che controllano le tabelle degli slot TCP.

Eseguire `sysctl -a | grep tcp.*.slot_table` e osservare i seguenti parametri:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tutti i sistemi Linux dovrebbero includere `sunrpc.tcp_slot_table_entries`, ma solo alcuni includono `sunrpc.tcp_max_slot_table_entries`. Entrambi devono essere impostati su 128.

> La mancata impostazione di questi parametri può avere effetti significativi sulle prestazioni. In alcuni casi, le prestazioni sono limitate poiché il sistema operativo linux non fornisce i/o sufficienti In altri casi, le latenze i/o aumentano quando il sistema operativo linux tenta di emettere più i/o di quanto possa essere gestito.

Scheduler i/O.

Il kernel Linux permette un controllo di basso livello sul modo in cui l'i/o blocca i dispositivi è programmato.

Le impostazioni predefinite su varie distribuzioni di Linux variano notevolmente. MySQL consiglia di utilizzare `NOOP` oppure un `deadline` Scheduler i/o con i/o asincrono nativo (AIO) su Linux. In generale, i clienti NetApp e i test interni mostrano risultati migliori con NoOps.

Il motore di storage InnoDB di MySQL utilizza il sottosistema i/o asincrono (AIO nativo) su Linux per eseguire richieste di lettura e scrittura per le pagine dei file di dati. Questo comportamento è controllato da `innodb_use_native_aio` opzione di configurazione, attivata per impostazione predefinita. Con un sistema AIO nativo, il tipo di pianificatore i/o influisce maggiormente sulle prestazioni di i/O. Esegui benchmark per determinare quale scheduler i/o offrirà i risultati migliori per il tuo carico di lavoro e l'ambiente.

Per istruzioni sulla configurazione dello scheduler i/o, consultare la documentazione relativa a Linux e MySQL.

Descrittori di file

Per l'esecuzione, il server MySQL ha bisogno di descrittori di file, e i valori predefiniti non sono sufficienti.

Le utilizza per aprire nuove connessioni, archiviare tabelle nella cache, creare tabelle temporanee per risolvere query complesse e accedere a quelle persistenti. Se mysqld non è in grado di aprire nuovi file quando necessario, può smettere di funzionare correttamente. Un sintomo comune di questo problema è l'errore 24, "troppi file aperti". Il numero di descrittori di file che mysqld può aprire simultaneamente è definito dal `open_files_limit` opzione impostata nel file di configurazione (`/etc/my.cnf`). Ma `open_files_limit` dipende anche dai limiti del sistema operativo. Questa dipendenza rende l'impostazione della variabile più complicata.

MySQL non può impostare ITS `open_files_limit` opzione superiore a quanto specificato in `ulimit 'open files'`. Pertanto, è necessario impostare esplicitamente questi limiti a livello del sistema operativo per consentire a MySQL di aprire i file in base alle necessità. Ci sono due modi per controllare il limite dei file in Linux:

- Il `ulimit` command fornisce rapidamente una descrizione dettagliata dei parametri consentiti o bloccati. Le modifiche apportate eseguendo questo comando non sono permanenti e si cancellano dopo un riavvio del sistema.
- Modifiche al `/etc/security/limit.conf` i file sono permanenti e non sono interessati dal riavvio del sistema.

Assicurarsi di modificare sia i limiti hard che soft per l'utente mysql. I seguenti estratti provengono dalla configurazione:

```
mysql hard nofile 65535
mysql soft nofile 65353
```

In parallelo, aggiornare la stessa configurazione in `my.cnf` per utilizzare completamente i limiti dei file aperti.

Configurazione dello storage

NFS

La documentazione MySQL consiglia di utilizzare NFSv4 per le implementazioni NAS.

Dimensioni del trasferimento di NFS ONTAP

Per impostazione predefinita, ONTAP limiterà le dimensioni di i/o NFS a 64K. I/o casuali con un database MySQL utilizzano blocchi di dimensioni molto inferiori, che sono ben al di sotto del massimo di 64K KB. L'io a blocchi di grandi dimensioni è solitamente parallelizzato, quindi anche il massimo di 64K KB non costituisce un limite.

Ci sono alcuni carichi di lavoro in cui il massimo di 64K crea un limite. In particolare, le operazioni single-threaded, come le operazioni di backup della scansione completa del piano d'esame, verranno eseguite in modo più rapido ed efficiente se il database è in grado di eseguire un numero di io inferiore ma maggiore. La

dimensione ottimale di gestione io per ONTAP con carichi di lavoro del database è 256K. Le opzioni di montaggio NFS elencate per i sistemi operativi specifici elencati di seguito sono state aggiornate da 64K a 256K di conseguenza.

Le dimensioni massime di trasferimento per una SVM ONTAP possono essere modificate come segue:

```
Cluster01::> set advanced

Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.

Do you want to continue? {y|n}: y

Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
```

Non diminuire mai la dimensione di trasferimento massima consentita su ONTAP al di sotto del valore rsize/wsize dei filesystem NFS attualmente montati. In alcuni sistemi operativi, ciò può causare blocchi o addirittura danni ai dati. Ad esempio, se i client NFS sono attualmente impostati su un valore rsize/wsize di 65536, la dimensione massima di trasferimento ONTAP potrebbe essere regolata tra 65536 e 1048576 senza alcun effetto perché i client stessi sono limitati. La riduzione della dimensione massima di trasferimento inferiore a 65536 GB può danneggiare la disponibilità o i dati.

NetApp consiglia

Impostazione della seguente impostazione NFSv4 fstab (/etc/fstab):

```
nfs4 rw,
hard,nointr,bg,vers=4,proto=tcp,noatime,rsize=262144,wsize=262144
```

Un problema comune con NFSv3 è stato il blocco dei file di registro InnoDB dopo un'interruzione dell'alimentazione. Questo problema è stato risolto utilizzando i file di registro Time o Switching. Tuttavia, NFSv4 ha operazioni di blocco e tiene traccia dei file aperti e delle delegazioni.

SAN

È possibile collocare database di dimensioni inferiori su una coppia di LUN standard, a condizione che le richieste di i/o e capacità rientrino nei limiti di un singolo file system LUN. Ad esempio, un database che richiede circa 2K IOPS casuali può essere ospitato su un singolo file system su un singolo LUN. Analogamente, un database di sole 100GB GB di dimensioni dovrebbe adattarsi a un singolo LUN, senza creare problemi di gestione.

Database di dimensioni maggiori richiedono LUN multiple. Ad esempio, un database che richiede 100K IOPS avrà probabilmente bisogno di almeno otto LUN. Un singolo LUN sarebbe diventato un collo di bottiglia a causa del numero inadeguato di canali SCSI per le unità. Analogamente, sarebbe difficile gestire un database da 10TB TB su un singolo LUN da 10TB GB. I gestori di volumi logici sono progettati per unire le funzionalità di performance e capacità di più LUN per migliorare le prestazioni e la gestibilità.

In entrambi i casi, dovrebbe essere sufficiente una coppia di ONTAP Volumes. Con una configurazione semplice, la LUN dei file di dati viene posizionata in un volume dedicato, come farebbe la LUN di log. Con una configurazione di volume manager logica, tutte le LUN del gruppo di volumi dei file di dati si troverebbero in un volume dedicato e le LUN del gruppo di volumi di log si troverebbero in un secondo volume dedicato.

NetApp consiglia di utilizzare due file system per le distribuzioni MySQL su SAN:

- Il primo file system memorizza tutti i dati MySQL inclusi tablespace, dati e indice.
- Il secondo file system archivia tutti i log (log binari, log lenti e log delle transazioni).

Esistono diverse ragioni per separare i dati in questo modo, tra cui:

- I modelli di i/o dei file di dati e di registro sono diversi. La loro separazione permetterebbe più opzioni con i controlli QoS.
- L'uso ottimale della tecnologia Snapshot richiede la capacità di ripristinare in maniera indipendente i file di dati. L'associazione di file di dati con file di registro interferisce con il ripristino dei file di dati.
- La tecnologia NetApp SnapMirror può essere utilizzata per fornire una semplice funzionalità di disaster recovery con RPO ridotto per un database; tuttavia, richiede diverse pianificazioni della replica per i file di dati e log.

Utilizzare questo layout di base a due volumi per rendere la soluzione a prova di futuro, in modo che tutte le funzioni di ONTAP possano essere utilizzate se necessario.

NetApp consiglia la formattazione dell'unità con il file system ext4, grazie alle seguenti funzioni:

- Approccio esteso alle funzioni di gestione dei blocchi utilizzate nel file system di journaling (JFS) e nelle funzioni di allocazione differita del file system esteso (XFS).
- ext4 permette file system fino a 1 exbibyte (2^60 byte) e file fino a 16 tebibyte (16 * 2^40 byte). Al contrario, il file system ext3 supporta solo file system di dimensioni massime pari a 16TB MB e file di dimensioni massime pari a 2TB MB.
- Nei file system ext4, l'allocazione di più blocchi (mballoc) alloca più blocchi per un file in un'unica operazione, invece di assegnarli uno alla volta, come in ext3. Questa configurazione riduce l'overhead di chiamata dell'allocatore di blocchi diverse volte e ottimizza l'allocazione di memoria.
- Anche se XFS è il default per molte distribuzioni Linux, gestisce i metadati in modo diverso e non è adatto per alcune configurazioni MySQL.

NetApp consiglia di utilizzare le opzioni di dimensione del blocco 4K con l'utilità mkfs per allinearsi alle dimensioni del LUN del blocco esistenti.

```
mkfs.ext4 -b 4096
```

Le LUN NetApp memorizzano dati in blocchi fisici da 4KB KB, ottenendo otto blocchi logici da 512 byte.

Se non si impostano le stesse dimensioni del blocco, l'i/o non verrà allineato correttamente con i blocchi fisici e potrebbe scrivere in due unità diverse in un gruppo RAID, con conseguente latenza.

È importante allineare l'i/o per semplificare le operazioni di lettura/scrittura. Tuttavia, quando l'i/o inizia ad un blocco logico che non si trova all'inizio di un blocco fisico, l'i/o è disallineato. Le operazioni di i/o sono allineate solo quando iniziano presso un blocco logico, il primo blocco logico in un blocco fisico.

Database Oracle

Database Oracle su ONTAP

ONTAP è progettato per i database Oracle. Per decenni, ONTAP è stato ottimizzato per le esigenze uniche di i/o dei database relazionali e sono state create più funzionalità di ONTAP appositamente per soddisfare le esigenze dei database Oracle e persino su richiesta della stessa Oracle Inc.

> Questa documentazione sostituisce i report tecnici precedentemente pubblicati *TR-3633: Database Oracle su ONTAP; TR-4591: Protezione dei dati Oracle: Backup, recovery, replica; TR-4592: Oracle su MetroCluster; e TR-4534: Migrazione dei database Oracle su sistemi di storage NetApp*

Oltre ai numerosi modi possibili in cui ONTAP apporta valore all'ambiente di database, esiste anche una vasta gamma di requisiti utente, incluse le dimensioni del database, i requisiti di performance e le esigenze di protezione dei dati. Le distribuzioni note di storage NetApp includono tutto, da un ambiente virtualizzato di circa 6.000 database in esecuzione su VMware ESX a un data warehouse a singola istanza, di dimensioni attuali pari a 996TB TB e in crescita. Di conseguenza, sono disponibili alcune Best practice chiare per la configurazione di un database Oracle su storage NetApp.

I requisiti per l'utilizzo di un database Oracle su storage NetApp vengono risolti in due modi. In primo luogo, quando esiste una buona pratica chiara, essa verrà richiamata in modo specifico. A un livello generale, verranno spiegate molte considerazioni di progettazione che i progettisti delle soluzioni di storage Oracle devono affrontare in base ai loro specifici requisiti di business.

Configurazione di ONTAP

RAID

RAID si riferisce all'utilizzo della ridondanza per proteggere i dati dalla perdita di un'unità.

Occasionalmente sorgono domande riguardanti i livelli RAID nella configurazione dello storage NetApp utilizzato per i database Oracle e altre applicazioni aziendali. Molte Best practice Oracle precedenti relative alla configurazione degli array di storage contengono avvisi sull'utilizzo del mirroring RAID e/o sull'eliminazione di determinati tipi di RAID. Sebbene sollevino punti validi, questi sorgenti non si applicano a RAID 4 e alle tecnologie NetApp RAID DP e RAID-TEC utilizzate in ONTAP.

RAID 4, RAID 5, RAID 6, RAID DP e RAID-TEC utilizzano tutti la parità per garantire che il guasto al disco non determini una perdita di dati. Queste opzioni RAID offrono un utilizzo dello storage migliore rispetto al mirroring, ma la maggior parte delle implementazioni RAID presenta uno svantaggio che influisce sulle operazioni di scrittura. Il completamento di un'operazione di scrittura su altre implementazioni RAID potrebbe richiedere letture di più unità per rigenerare i dati di parità, un processo comunemente chiamato penalizzazione RAID.

ONTAP, tuttavia, non subisce questa penalizzazione del RAID. Ciò è dovuto all'integrazione di NetApp WAFL (Write Anywhere file Layout) con il livello RAID. Le operazioni di scrittura vengono unite nella RAM e preparate come uno stripe RAID completo, inclusa la generazione della parità. ONTAP non ha bisogno di eseguire una lettura per completare una scrittura, il che significa che ONTAP e WAFL evitare la penalizzazione RAID. Le performance per le operazioni critiche in termini di latenza, come il logging di redo, vengono mantenute e le scritture random dei file di dati non comportano penalizzazioni RAID dovute alla necessità di rigenerare la

parità.

Per quanto riguarda l'affidabilità statistica, anche RAID DP offre una protezione migliore rispetto al mirroring RAID. Il problema principale è la richiesta fatta sui dischi durante una ricostruzione del RAID. Con un set RAID con mirroring, il rischio di perdita di dati causata da un guasto al disco e durante la ricostruzione nel partner nel set RAID è molto maggiore del rischio di un guasto a tre dischi in un set RAID DP.

Gestione della capacità

La gestione di un database o di un'altra applicazione aziendale con storage aziendale prevedibile, gestibile e ad alte prestazioni richiede spazio libero sulle unità per la gestione di dati e metadati. La quantità di spazio libero richiesta dipende dal tipo di unità utilizzata e dai processi aziendali.

Lo spazio libero viene definito come lo spazio non utilizzato per i dati effettivi e include lo spazio non allocato dell'aggregato e lo spazio non utilizzato all'interno dei volumi costituenti. È importante prendere in considerazione anche il thin provisioning. Ad esempio, un volume potrebbe contenere un LUN da 1TB GB, di cui solo il 50% viene utilizzato dai dati reali. In un ambiente con thin provisioning, questo sembra consumare correttamente 500GB GB di spazio. Tuttavia, in un ambiente con provisioning completo, la capacità completa di 1TB TB sembra essere in uso. I 500GB GB di spazio non allocato sono nascosti. Questo spazio non è utilizzato dai dati effettivi e deve quindi essere incluso nel calcolo dello spazio libero totale.

Di seguito sono riportate le raccomandazioni NetApp per i sistemi storage utilizzati per le applicazioni aziendali:

Aggregati SSD, inclusi i sistemi AFF

NetApp consiglia almeno il 10% di spazio libero. Ciò comprende tutto lo spazio inutilizzato, compreso lo spazio libero all'interno dell'aggregato o di un volume ed eventuale spazio libero allocato a causa dell'utilizzo del provisioning completo, ma non utilizzato dai dati effettivi. Lo spazio logico non è importante, la domanda è quanto spazio fisico libero effettivo è disponibile per lo storage dei dati.

Il consiglio di liberare il 10% dello spazio è molto conservativo. Gli aggregati SSD possono supportare i carichi di lavoro a livelli di utilizzo ancora più elevati senza influire sulle performance. Tuttavia, con l'aumento dell'utilizzo dell'aggregato, aumenta anche il rischio di esaurimento dello spazio se l'utilizzo non viene monitorato con attenzione. Inoltre, mentre si utilizza un sistema al 99% della capacità potrebbe non verificarsi un peggioramento delle performance, tuttavia si verificherebbe un sforzo di gestione che impedirebbe il riempimento completo del sistema mentre si ordina hardware aggiuntivo e potrebbe essere necessario del tempo per l'acquisto e l'installazione di dischi aggiuntivi.

Aggregati HDD, compresi gli aggregati Flash Pool

NetApp consiglia almeno il 15% di spazio libero quando si utilizzano unità rotanti. Ciò comprende tutto lo spazio inutilizzato, compreso lo spazio libero all'interno dell'aggregato o di un volume ed eventuale spazio libero allocato a causa dell'utilizzo del provisioning completo, ma non utilizzato dai dati effettivi. Le prestazioni saranno influenzate dall'avvicinarsi del 10% dello spazio libero.

Macchine virtuali di storage

La gestione dello storage del database Oracle è centralizzata su una Storage Virtual

Machine (SVM)

Una SVM, nota come vserver nell'interfaccia a riga di comando di ONTAP, è un'unità funzionale di base dello storage ed è utile confrontare una SVM con un guest su un server VMware ESX.

Quando viene installato per la prima volta, ESX non dispone di funzionalità preconfigurate, come l'hosting di un sistema operativo guest o il supporto di un'applicazione per l'utente finale. Si tratta di un container vuoto fino a quando non viene definita una macchina virtuale (VM). ONTAP è simile. Quando viene installata per la prima volta, ONTAP non dispone di funzionalità di servizio dati fino a quando non viene creata una SVM. È il linguaggio della SVM che definisce i servizi dati.

Come per altri aspetti dell'architettura dello storage, le migliori opzioni per il design di SVM e interfaccia logica (LIF) dipendono in gran parte dai requisiti di scalabilità e dalle esigenze di business.

SVM

Non esistono Best practice ufficiali per il provisioning di SVM per ONTAP. Il giusto approccio dipende dai requisiti di gestione e sicurezza.

La maggior parte dei clienti utilizza una SVM primaria per la maggior parte delle loro esigenze quotidiane, quindi crea un piccolo numero di SVM per esigenze speciali. Ad esempio, è possibile creare:

- Una SVM per un database aziendale critico gestita da un team di specialisti
- Una SVM per un gruppo di sviluppo al quale è stato assegnato un controllo amministrativo completo in modo da poter gestire il proprio storage in maniera indipendente
- Una SVM per i dati di business sensibili, come le risorse umane o i dati di reporting finanziario, per cui il team di amministrazione deve essere limitato

In un ambiente multi-tenant, è possibile assegnare a ciascun tenant una SVM dedicata. Il limite per il numero di SVM e LIF per cluster, coppia ha e nodo dipende dal protocollo in uso, dal modello di nodo e dalla versione di ONTAP. Consultare "NetApp Hardware Universe" per questi limiti.

Gestione delle performance con QoS ONTAP

La gestione sicura ed efficiente di più database Oracle richiede un'efficace strategia di QoS. Il motivo è rappresentato dalle funzionalità di performance in costante aumento offerte da un sistema storage moderno.

Nello specifico, la maggiore adozione dello storage all-flash ha permesso il consolidamento dei carichi di lavoro. Gli storage array che si affidano a supporti rotanti tendevano a supportare solo un numero limitato di workload i/o-intensive a causa delle limitate funzionalità IOPS della tecnologia delle unità rotazionali meno recente. Uno o due database altamente attivi saturerebbero i dischi sottostanti molto prima che gli storage controller raggiungano i loro limiti. Questo è cambiato. La capacità di performance di un numero relativamente contenuto di dischi SSD è in grado di saturare anche gli storage controller più potenti. Ciò significa che è possibile sfruttare tutte le funzionalità dei controller senza la paura di un improvviso crollo delle performance con picchi di latenza dei supporti rotanti.

Come esempio di riferimento, un semplice sistema ha AFF A800 a due nodi è in grado di fornire fino a un milione di IOPS casuali prima che la latenza superi un millisecondo. Ci si aspetta che pochissimi carichi di lavoro singoli raggiungano tali livelli. L'utilizzo completo di questo array di sistema AFF A800 implicherà l'hosting di più carichi di lavoro, per questo motivo in modo sicuro, garantendo al contempo la prevedibilità dei requisiti di qualità del servizio.

Esistono due tipi di qualità del servizio (QoS) in ONTAP: IOPS e larghezza di banda. È possibile applicare controlli di qualità del servizio a SVM, volumi, LUN e file.

QoS (IOPS)

Un controllo della qualità del servizio IOPS si basa ovviamente sugli IOPS totali di una data risorsa, ma esistono alcuni aspetti della qualità del servizio IOPS che potrebbero non essere intuitivi. Alcuni clienti sono rimasti colpiti dall'apparente aumento della latenza al raggiungimento di una soglia IOPS. L'aumento della latenza è il risultato naturale della limitazione degli IOPS. Logicamente, funziona in modo simile a un sistema token. Ad esempio, se un dato volume contenente file di dati ha un limite di 10K IOPS, ogni i/o che arriva deve prima ricevere un token per continuare l'elaborazione. Fino a quando non sono stati consumati più di 10K gettoni in un dato secondo, non sono presenti ritardi. Se le operazioni io devono attendere per ricevere il token, questa attesa viene visualizzata come latenza aggiuntiva. Più un carico di lavoro supera il limite di qualità del servizio, più a lungo ogni i/o deve attendere in coda per l'elaborazione del proprio turno, che appare all'utente come una latenza più elevata.

Prestare attenzione nell'applicazione dei controlli QoS ai dati dei log di transazione/ripristino del database. Mentre le richieste di performance del logging di redo sono in genere molto, molto più basse dei data afiles, l'attività del log di redo è molto bursty. L'io avviene in brevi impulsi e un limite QoS che appare appropriato per i livelli di io di redo medi potrebbe essere troppo basso per i requisiti effettivi. Il risultato può essere una serie di limitazioni delle performance, mentre la qualità del servizio viene associata a ogni burst dei log di ripristino. In generale, il redo e la registrazione dell'archivio non devono essere limitati dalla QoS.

QoS della larghezza di banda

Non tutte le dimensioni i/o sono uguali. Ad esempio, un database potrebbe eseguire un elevato numero di piccoli blocchi di lettura con il raggiungimento della soglia IOPS, tuttavia, è possibile che i database eseguano anche un'operazione di scansione completa della tabella, che consisterebbe in un numero molto ridotto di letture di blocchi di grandi dimensioni, consumando una grande quantità di larghezza di banda ma con un numero relativamente basso di IOPS.

Allo stesso modo, un ambiente VMware potrebbe gestire un numero molto elevato di IOPS casuali durante l'avvio, ma eseguirebbe un numero minore di io, ma più grande, durante un backup esterno.

Una gestione efficace delle performance a volte richiede limiti di qualità del servizio (QoS) IOPS o larghezza di banda, o anche entrambi.

Qualità del servizio minima/garantita

Molti clienti cercano una soluzione che includa QoS garantita, che sia più difficile da raggiungere di quanto possa sembrare e che sia potenzialmente abbastanza dispendiosa. Ad esempio, collocare 10 database con una garanzia di 10K IOPS richiede il dimensionamento di un sistema per uno scenario in cui tutti i 10 database vengono eseguiti contemporaneamente a 10K IOPS, per un totale di 100K.

L'utilizzo ottimale per i controlli minimi della qualità del servizio è la protezione dei carichi di lavoro critici. Ad esempio, prendi in considerazione un controller ONTAP con un numero massimo di IOPS possibile di 500K e un mix di workload di produzione e sviluppo. È consigliabile applicare policy QoS massime ai carichi di lavoro di sviluppo per impedire a qualsiasi database di monopolizzare il controller. Quindi, ai carichi di lavoro di produzione si applicano policy minime di qualità del servizio per assicurarsi che dispongano sempre degli IOPS richiesti, quando necessario.

QoS adattiva

La qualità del servizio adattiva fa riferimento alla funzionalità ONTAP, in cui il limite della qualità del servizio si basa sulla capacità dell'oggetto storage. Viene utilizzata raramente con i database perché di solito non esiste alcun collegamento tra le dimensioni di un database e i relativi requisiti prestazionali. I database di grandi dimensioni possono essere quasi inerti, mentre quelli di dimensioni inferiori possono utilizzare un numero elevato di IOPS.

La qualità del servizio adattiva può rivelarsi molto utile con i datastore di virtualizzazione, perché i requisiti di IOPS di tali set di dati tendono a correlare le dimensioni totali del database. Un datastore più recente, che contiene 1TB TB di file VMDK, avrà probabilmente bisogno di circa la metà delle performance rispetto a un datastore da 2TB TB. La qualità del servizio adattiva ti consente di aumentare automaticamente i limiti della qualità del servizio, man mano che il datastore viene popolato con i dati.

Efficienza

Le funzionalità di efficienza dello spazio di ONTAP sono ottimizzate per i database Oracle. In quasi tutti i casi, l'approccio migliore è quello di lasciare le impostazioni predefinite con tutte le funzioni di efficienza attivate.

Le funzionalità di efficienza in termini di spazio, come compressione, compaction e deduplica, sono progettate per aumentare la quantità di dati logici applicabili a una determinata quantità di storage fisico. Il risultato è una riduzione dei costi e dell'overhead di gestione.

Ad un livello elevato, la compressione è un processo matematico in cui gli schemi nei dati vengono rilevati e codificati in modo da ridurre i requisiti di spazio. La deduplica, invece, rileva i blocchi di dati effettivi e ripetuti e rimuove le copie estranee. La tecnologia di compaction consente a più blocchi logici di dati di condividere lo stesso blocco fisico sui supporti.

> Per una spiegazione dell'interazione tra efficienza dello storage e prenotazione frazionata, vedere le sezioni seguenti sul thin provisioning.

Compressione

Prima della disponibilità dei sistemi storage all-flash, la compressione basata su array aveva un valore limitato, perché la maggior parte dei carichi di lavoro con i/o-intensive richiedeva un numero molto elevato di spindle per fornire performance accettabili. I sistemi storage contenevano invariabilmente una capacità superiore rispetto a quella richiesta come effetto collaterale dell'elevato numero di dischi. La situazione è cambiata con l'ascesa dello storage a stato solido. Non è più necessario effettuare un provisioning in eccesso significativo dei dischi solo per ottenere buone prestazioni. Lo spazio su disco di un sistema di storage può essere adattato alle effettive esigenze di capacità.

L'aumento della capacità degli IOPS dei dischi a stato solido (SSD) offre quasi sempre risparmi sui costi rispetto ai dischi rotanti, ma la compressione può ottenere ulteriori risparmi aumentando la capacità effettiva dei supporti a stato solido.

Esistono diversi modi per comprimere i dati. Molti database includono proprie funzionalità di compressione, sebbene raramente queste vengano osservate negli ambienti dei clienti. Il motivo è solitamente la penalizzazione delle prestazioni per una **modifica** dei dati compressi, mentre con alcune applicazioni vi sono elevati costi di licenza per la compressione a livello di database. Infine, ci sono le conseguenze globali delle performance sulle operazioni di database. Ha poco senso pagare un costo elevato di licenza per CPU per una CPU che esegue la compressione e la decompressione dei dati piuttosto che un vero lavoro di database. Un'opzione migliore è trasferire il lavoro di compressione sul sistema storage.

Compressione adattiva

La compressione adattiva è stata testata accuratamente con carichi di lavoro Enterprise senza effetti osservati sulle performance, anche in un ambiente all-flash in cui la latenza viene misurata in microsecondi. Alcuni clienti hanno anche segnalato un aumento delle performance con l'utilizzo della compressione, perché i dati rimangono compressi nella cache, aumentando di fatto la quantità di cache disponibile in un controller.

ONTAP gestisce i blocchi fisici in 4KB unità. La compressione adattiva utilizza dimensioni predefinite dei blocchi di compressione di 8KB KB, il che significa che i dati sono compressi in unità da 8KB KB. Corrisponde alle dimensioni dei blocchi di 8KB KB utilizzate più spesso dai database relazionali. Gli algoritmi di compressione diventano più efficienti con la compressione di un numero maggiore di dati come una singola unità. Una dimensione dei blocchi di compressione da 32KB KB sarebbe più efficiente in termini di spazio rispetto a un'unità dei blocchi di compressione da 8KB KB. Ciò significa che la compressione adattiva che utilizza le dimensioni predefinite dei blocchi di 8KB KB produce tassi di efficienza leggermente inferiori, ma esiste anche un vantaggio significativo nell'utilizzo di dimensioni inferiori dei blocchi di compressione. I carichi di lavoro dei database includono un'elevata attività di sovrascrittura. La sovrascrittura di un 8KB di un blocco di dati 32KB compresso richiede la lettura dell'intero 32KB di dati logici, la decompressione, l'aggiornamento della regione 8KB richiesta, la ricompressione e quindi la riscrittura dell'intero 32KB sui dischi. Si tratta di un'operazione molto costosa per un sistema storage ed è il motivo per cui alcuni storage array concorrenti basati su dimensioni dei blocchi di compressione più grandi implicano anche una significativa penalizzazione delle performance con i carichi di lavoro dei database.

Le dimensioni dei blocchi utilizzate dalla compressione adattiva possono essere aumentate fino a 32KB KB. Questo può migliorare l'efficienza di archiviazione e dovrebbe essere considerato per i file inattivi come i log delle transazioni e i file di backup quando una quantità sostanziale di tali dati è memorizzata nell'array. In alcune situazioni, i database attivi che utilizzano dimensioni blocco 16KB KB o 32KB KB possono anche trarre vantaggio dall'aumento delle dimensioni blocco della compressione adattiva per adeguarsi. Consulta un NetApp o un rappresentante del partner per ottenere indicazioni relative all'adeguatezza del tuo carico di lavoro.

Le dimensioni dei blocchi di compressione superiori a 8KB KB non devono essere utilizzate insieme alla deduplica nelle destinazioni di backup in streaming. Il motivo è che piccole modifiche ai dati di backup influiscono sulla finestra di compressione 32KB. Se la finestra si sposta, i dati compressi risultanti differiscono per l'intero file. La deduplica si verifica dopo la compressione, il che significa che il motore di deduplica vede ogni backup compresso in modo diverso. Se è richiesta la deduplica dei backup in streaming, è consigliabile utilizzare solo la compressione adattiva per blocchi da 8KB KB. La compressione adattiva è preferibile, perché funziona a blocchi di dimensioni inferiori e non interrompe l'efficienza di deduplica. Per motivi simili, la compressione lato host interferisce anche con l'efficienza della deduplica.

Allineamento delle compressioni

La compressione adattiva in un ambiente di database richiede alcune considerazioni sull'allineamento dei blocchi di compressione. Ciò rappresenta solo una preoccupazione per i dati che sono soggetti a sovrascritture casuali di blocchi molto specifici. Questo approccio è simile in teoria all'allineamento complessivo del file system, dove l'inizio di un file system deve essere allineato al limite di un dispositivo 4K e la dimensione di blocco di un file system deve essere un multiplo di 4K.

Ad esempio, una scrittura 8KB in un file viene compressa solo se si allinea con un limite 8KB all'interno del file system stesso. Questo punto significa che deve rientrare nel primo 8KB del file, nel secondo 8KB del file e così via. Il modo più semplice per garantire un corretto allineamento è utilizzare il tipo di LUN corretto, ogni partizione creata dovrebbe avere un offset dall'inizio del dispositivo che è un multiplo di 8K, e utilizzare una dimensione del blocco del file system che è un multiplo della dimensione del blocco del database.

Dati come backup o log delle transazioni sono operazioni scritte in sequenza che coprono più blocchi, tutti compressi. Pertanto, non è necessario considerare l'allineamento. L'unico modello di i/o che desta preoccupazione sono le sovrascritture casuali dei file.

Compaction dei dati

La data compaction è una tecnologia che migliora l'efficienza di compressione. Come indicato in precedenza, la sola compressione adattiva può garantire risparmi 2:1:1 al meglio, perché è limitata alla memorizzazione di un i/o da 8KB KB in un blocco WAFL da 4KB KB. I metodi di compressione con dimensioni dei blocchi maggiori garantiscono una maggiore efficienza. Tuttavia, non sono adatte per i dati che sono soggetti a piccole sovrascritture dei blocchi. La decompressione di 32KB unità di dati, l'aggiornamento di una porzione 8KB, la ricompressione e la riscrittura sui dischi crea overhead.

La data compaction opera consentendo di memorizzare più blocchi logici all'interno dei blocchi fisici. Ad esempio, un database con dati altamente comprimibili come testo o blocchi parzialmente completi può comprimere da 8KB a 1KB. Senza la compaction, quei 1KB PB di dati continuerebbero ad occupare un intero blocco da 4KB KB. Inline data compaction per memorizzare 1KB TB di dati compressi in sole 1KB:1 di spazio fisico insieme ad altri dati compressi. Non si tratta di una tecnologia di compressione, ma semplicemente di un metodo più efficiente per allocare spazio sulle unità e quindi non dovrebbe creare alcun effetto rilevabile sulle prestazioni.

Il grado di risparmio ottenuto varia. I dati già compressi o crittografati non possono in genere essere ulteriormente compressi, e pertanto tali set di dati non traggono vantaggio dalla compattazione. Al contrario, i file di dati appena inizializzati contenenti poco più dei metadati dei blocchi e la compressione di zeri fino a 80:1.

Efficienza di conservazione sensibile alla temperatura

L'efficienza di stoccaggio sensibile alla temperatura (TSSE) è disponibile in ONTAP 9.8 e versioni successive. Si affida alle mappe termiche di accesso ai blocchi per identificare i blocchi a cui si accede raramente e comprimerli con una maggiore efficienza.

Deduplica

La deduplica consiste nella rimozione di dimensioni dei blocchi duplicate da un set di dati. Ad esempio, se lo stesso blocco 4KB esistesse in 10 file diversi, la deduplica reindirizzerebbe quel blocco 4KB in tutti i file 10 allo stesso blocco fisico da 4KB KB. Il risultato sarebbe un miglioramento di 10:1 volte in efficienza per quei dati.

Dati come i LUN di avvio guest di VMware si deduplicano in genere in modo estremamente efficace poiché sono costituiti da più copie degli stessi file del sistema operativo. Sono state osservate un'efficienza pari o superiore a 100:1.

Alcuni dati non contengono dati duplicati. Ad esempio, un blocco Oracle contiene un'intestazione univoca a livello globale per il database e un trailer quasi univoco. Di conseguenza, la deduplica di un database Oracle raramente offre un risparmio superiore al 1%. La deduplica con i database MS SQL è leggermente migliore, ma i metadati univoci a livello di blocco rimangono un limite.

In pochi casi, sono stati osservati risparmi di spazio fino al 15% nei database con blocchi di dimensioni grandi e 16KB. Il 4KB iniziale di ciascun blocco contiene la testata unica a livello globale, mentre il 4KB finale contiene il rimorchio quasi unico. I blocchi interni sono candidati per la deduplica, sebbene in pratica ciò sia quasi interamente attribuito alla deduplica di dati azzerati.

Molti array della concorrenza rivendicano la capacità di deduplicare i database sulla base del presupposto che un database venga copiato più volte. Anche in questo caso è possibile utilizzare la deduplica NetApp, ma ONTAP offre un'opzione migliore: La tecnologia FlexClone di NetApp. Il risultato finale è lo stesso; vengono

create più copie di un database che condividono la maggior parte dei blocchi fisici sottostanti. L'utilizzo di FlexClone è molto più efficiente della necessità di dedicare tempo alla copia e alla deduplica dei file di database. In effetti, non viene effettuata alcuna duplicazione piuttosto che deduplica, poiché al primo posto non viene mai creato un duplicato.

Efficienza e thin provisioning

Le funzionalità di efficienza sono forme di thin provisioning. Ad esempio, una LUN da 100GB GB che occupa un volume da 100GB GB potrebbe comprimere fino a 50GB GB. Non ci sono risparmi effettivi ancora realizzati perché il volume è ancora 100GB. Le dimensioni del volume devono essere innanzitutto ridotte in modo che lo spazio salvato possa essere utilizzato in un'altra posizione del sistema. Se successivamente le modifiche apportate al LUN da 100GB GB rendono i dati meno comprimibili, il LUN aumenta le dimensioni e il volume potrebbe riempirsi.

Il thin provisioning è vivamente consigliato in quanto consente di semplificare la gestione, offrendo al contempo un sostanziale miglioramento della capacità utilizzabile con conseguenti risparmi sui costi. Il motivo è semplice: Gli ambienti di database includono spesso molto spazio vuoto, un elevato numero di volumi e LUN e dati comprimibili. Il thick provisioning crea la riserva di spazio sullo storage per volumi e LUN, nel caso in cui un giorno raggiungano il 100% di riempimento e contengano dati non comprimibili al 100%. È improbabile che ciò accada mai. Il thin provisioning consente di recuperare lo spazio e di utilizzarlo altrove e consente la gestione della capacità basata sul sistema storage stesso piuttosto che su molti volumi e LUN più piccoli.

Alcuni clienti preferiscono utilizzare il thick provisioning, per carichi di lavoro specifici o generalmente basato su pratiche operative e di approvvigionamento consolidate.

> Se un volume viene sottoposto a thick provisioning, è necessario fare attenzione a disattivare completamente tutte le funzionalità di efficienza per quel volume, inclusa la decompressione e la rimozione della deduplica tramite il `sis undo` comando. Il volume non dovrebbe comparire nell' `volume efficiency show`output. In tal caso, il volume è ancora parzialmente configurato per le funzioni di efficienza. Di conseguenza, la sovrascrittura garantisce un funzionamento diverso, aumentando le possibilità che le sovrascritture causino l'esaurimento inaspettato dello spazio del volume, con conseguenti errori di i/o del database.

Best practice di efficienza

NetApp consiglia quanto segue:

Valori predefiniti AFF

I volumi creati su ONTAP in esecuzione su un sistema AFF all-flash vengono sottoposti a thin provisioning con tutte le funzionalità di efficienza inline abilitate. Sebbene in genere i database non beneficino della deduplica e possano includere dati non comprimibili, le impostazioni predefinite sono comunque appropriate per quasi tutti i carichi di lavoro. ONTAP è progettato per elaborare in modo efficiente tutti i tipi di dati e gli schemi i/o, indipendentemente dal fatto che comportino risparmi. Le impostazioni predefinite devono essere modificate solo se le ragioni sono pienamente comprese e se vi è un vantaggio a deviare.

Raccomandazioni generali

- Se i volumi e/o le LUN non sono dotati di thin provisioning, è necessario disabilitare tutte le impostazioni di efficienza perché queste funzionalità non offrono risparmi e la combinazione del thick provisioning con l'efficienza dello spazio può causare comportamenti imprevisti, inclusi errori di spazio esaurito.
- Se i dati non sono soggetti a sovrascritture, ad esempio con i backup o i log delle transazioni dei database, puoi ottenere una maggiore efficienza abilitando TSSE con un periodo di raffreddamento ridotto.

- Alcuni file potrebbero contenere una quantità significativa di dati non comprimibili, ad esempio quando la compressione è già abilitata a livello di applicazione dei file sono crittografati. Se uno di questi scenari è vero, considerare la possibilità di disattivare la compressione per consentire un funzionamento più efficiente su altri volumi che contengono dati comprimibili.
- Non utilizzare sia la compressione 32KB che la deduplica con i backup del database. Vedere la sezione Compressione adattiva per ulteriori informazioni.

Thin provisioning

Il thin provisioning per un database Oracle richiede un'attenta pianificazione, perché ne consegue che è possibile configurare più spazio su un sistema di storage rispetto a quello necessariamente fisicamente disponibile. Vale la pena di fare tutto questo perché, se eseguito correttamente, il risultato è un notevole risparmio sui costi e un miglioramento della gestibilità.

Il thin provisioning è disponibile in molte forme e rappresenta parte integrante di molte funzionalità offerte da ONTAP a un ambiente applicativo aziendale. Il thin provisioning è inoltre strettamente correlato alle tecnologie di efficienza per lo stesso motivo: Le funzionalità di efficienza consentono di memorizzare dati più logici rispetto a quanto tecnicamente esistente nel sistema storage.

Quasi tutti gli utilizzi delle snapshot implicano il thin provisioning. Ad esempio, un tipico database da 10TB TB su storage NetApp include circa 30 giorni di snapshot. Questa disposizione risulta in circa 10TB di dati visibili nel file system attivo e 300TB dedicati agli snapshot. In genere, il 310TB GB di storage totale risiede su un totale di circa 12TB - 15TB GB di spazio. Il database attivo consuma 10TB e i restanti 300TB di dati richiedono solo da 2TB a 5TB di spazio, in quanto vengono memorizzate solo le modifiche apportate ai dati originali.

Anche il cloning è un esempio di thin provisioning. Un importante cliente NetApp ha creato 40 cloni di un database da 80TB TB per l'utilizzo da parte dello sviluppo. Se tutti i 40 sviluppatori che utilizzano questi cloni sovrascrivono ogni blocco in ogni file dati, sarebbero necessari oltre 3,2PB TB di storage. In pratica, il turnover è basso e il requisito di spazio collettivo è più vicino a 40TB, perché solo le modifiche sono memorizzate sui drive.

Gestione dello spazio

È necessario prestare particolare attenzione al thin provisioning di un ambiente applicativo, perché la velocità di modifica dei dati può aumentare inaspettatamente. Ad esempio, il consumo di spazio dovuto agli snapshot può crescere rapidamente se le tabelle di database vengono riindicizzate o se viene applicata una patch su larga scala ai guest VMware. Un backup posizionato in modo errato può scrivere una grande quantità di dati in un tempo molto breve. Infine, può essere difficile recuperare alcune applicazioni se un file system esaurisce inaspettatamente lo spazio libero.

Fortunatamente, questi rischi possono essere risolti con un'attenta configurazione di `volume-autogrow` e. `snapshot-autodelete` criteri: Come indicato dai nomi, queste opzioni consentono a un utente di creare policy in grado di liberare automaticamente lo spazio occupato dalle snapshot o di far crescere un volume per ospitare dati aggiuntivi. Sono disponibili molte opzioni e le esigenze variano in base al cliente.

Vedere "documentazione per la gestione logica dello storage" per una discussione completa di queste funzioni.

Prenotazioni frazionarie

Riserva frazionaria si riferisce al comportamento di un LUN in un volume rispetto all'efficienza dello spazio. Quando l'opzione `fractional-reserve` è impostato al 100%, tutti i dati nel volume possono subire un turnover del 100% con qualsiasi modello di dati senza esaurire lo spazio sul volume.

Ad esempio, si consideri un database su una singola LUN da 250GB GB in un volume da 1TB GB. La creazione di uno snapshot comporterebbe immediatamente la riserva di ulteriori 250GB GB di spazio nel volume per garantire che il volume non esaurisca lo spazio per alcun motivo. L'utilizzo di riserve frazionarie comporta in genere uno spreco di risorse poiché è estremamente improbabile che ogni byte nel volume del database debba essere sovrascritto. Non c'è motivo di riservare spazio per un evento che non si verifica mai. Tuttavia, se un cliente non è in grado di monitorare il consumo di spazio in un sistema di storage e deve essere certo che lo spazio non si esaurisce mai, sarebbero necessarie prenotazioni frazionarie del 100% per utilizzare gli snapshot.

Compressione e deduplica

Compressione e deduplica sono entrambe forme di thin provisioning. Ad esempio, un impatto dei dati di 50TB:1 potrebbe comprimere fino a 30TB:1, ottenendo un risparmio di 20TB:1. Affinché la compressione possa produrre vantaggi, alcuni di questi 20TB TB devono essere utilizzati per altri dati, altrimenti il sistema storage deve essere acquistato con meno di 50TB TB. In questo modo è possibile memorizzare una quantità di dati superiore rispetto a quella tecnicamente disponibile sul sistema storage. Dal punto di vista dei dati, i dati sono 50TB, anche se occupano solo 30TB sulle unità.

Esiste sempre la possibilità che la compressibilità di un set di dati cambi, con conseguente aumento del consumo di spazio reale. Questo aumento dei consumi implica che la compressione deve essere gestita come con altre forme di thin provisioning in termini di monitoraggio e utilizzo `volume-autogrow` e. `snapshot-autodelete`.

Compressione e deduplica sono descritte in dettaglio nella sezione xref:./oracle/efficiency.html

Compressioni e prenotazioni frazionarie

La compressione è una forma di thin provisioning. Le prenotazioni frazionarie influiscono sull'utilizzo della compressione, con una nota importante; lo spazio viene riservato prima della creazione dell'istantanea. Normalmente, la riserva frazionaria è importante solo se esiste uno snapshot. Se non è presente uno snapshot, la riserva frazionaria non è importante. Questo non è il caso della compressione. Se viene creata una LUN su un volume con compressione, ONTAP preserva lo spazio per ospitare uno snapshot. Questo comportamento può creare confusione durante la configurazione, ma è previsto.

Ad esempio, consideriamo un volume da 10GB GB con una LUN da 5GB GB compressa a 2,5GB GB senza snapshot. Considerare questi due scenari:

- Riserva frazionaria = 100 risultati in 7,5GB utilizzo
- Riserva frazionaria = 0 risultati in 2,5GB utilizzo

Il primo scenario include 2,5GB di consumo di spazio per i dati attuali e 5GB di spazio per rappresentare il 100% di fatturato della fonte in previsione dell'utilizzo di snapshot. Il secondo scenario non riserva spazio aggiuntivo.

Sebbene questa situazione possa sembrare confusa, è improbabile che si verifichi nella pratica. La compressione implica thin provisioning e il thin provisioning in un ambiente LUN richiede prenotazioni frazionarie. È sempre possibile sovrascrivere i dati compressi con qualcosa di non comprimibile, il che significa che un volume deve essere sottoposto a thin provisioning per la compressione per consentire qualsiasi risparmio.

NetApp consiglia le seguenti configurazioni riservate:

- Impostare `fractional-reserve` a 0 quando è in atto il monitoraggio della capacità di base con `volume-autogrow` e. `snapshot-autodelete`.
- Impostare `fractional-reserve` a 100 se non vi è alcuna capacità di monitoraggio o se è impossibile scaricare lo spazio in qualsiasi circostanza.

Spazio libero e allocazione di spazio LVM

L'efficienza del thin provisioning delle LUN attive in un ambiente di file system può andare persa nel tempo quando i dati vengono eliminati. A meno che i dati eliminati non vengano sovrascritti con zero (vedere anche "ASMRU" Oppure lo spazio viene liberato con il recupero dello spazio TRIM/UNMAP, i dati "cancellati" occupano sempre più spazi vuoti non allocati nel file system. Inoltre, in molti ambienti di database il thin provisioning delle LUN attive è limitato, in quanto i file di dati vengono inizializzati alle dimensioni massime al momento della creazione.

Un'attenta pianificazione della configurazione LVM può migliorare l'efficienza e ridurre al minimo la necessità di provisioning dello storage e di ridimensionamento delle LUN. Quando si utilizza un LVM come Veritas VxVM o Oracle ASM, le LUN sottostanti vengono suddivise in estensioni che vengono utilizzate solo quando necessario. Ad esempio, se un set di dati inizia a 2TB TB ma potrebbe crescere fino a 10TB TB con il passare del tempo, è possibile inserire il set di dati in 10TB LUN con thin provisioning organizzati in un gruppo di dischi LVM. Occupa solo 2TB GB di spazio al momento della creazione e richiederebbe spazio aggiuntivo solo se le estensioni sono allocate per ospitare la crescita dei dati. Questo processo è sicuro finché lo spazio è monitorato.

Failover/switchover di ONTAP

Le operazioni di takeover e switchover dello storage devono garantire l'integrità delle operazioni dei database Oracle. Inoltre, gli argomenti utilizzati dalle operazioni di takeover e switchover possono influire sull'integrità dei dati se utilizzati in modo errato.

- In condizioni normali, le scritture in arrivo su un dato controller vengono mirrorate in modo sincrono per il partner. In un ambiente NetApp MetroCluster, le scritture vengono anche mirrorate su un controller remoto. Fino a quando non viene memorizzata in un supporto non volatile in tutte le posizioni, la scrittura non viene riconosciuta all'applicazione host.
- Il supporto che memorizza i dati di scrittura è chiamato memoria non volatile o NVMEM. Viene anche talvolta indicata come memoria non volatile ad accesso casuale (NVRAM, nonvolatile Random Access Memory), e può essere considerata come una cache di scrittura anche se funziona come un journal. In condizioni normali, i dati provenienti da NVMEM non vengono letti; vengono utilizzati solo per proteggere i dati in caso di guasti software o hardware. Quando i dati vengono scritti sulle unità disco rigido, i dati vengono trasferiti dalla RAM nel sistema, non da NVMEM.
- Durante un'operazione di takeover, un nodo di una coppia ha (high Availability) assume il controllo delle operazioni dal partner. Lo switchover è praticamente identico, ma si applica alle configurazioni MetroCluster in cui un nodo remoto assume le funzioni di un nodo locale.

Durante le normali operazioni di manutenzione, un'operazione di takeover o switchover dello storage deve essere trasparente, ad eccezione di una potenziale breve pausa nelle operazioni in base al cambiamento dei percorsi di rete. Il networking può rivelarsi complesso, tuttavia, ed è facile commettere errori, pertanto NetApp consiglia vivamente di eseguire accuratamente le operazioni di takeover e switchover prima di mettere in produzione un sistema storage. In questo modo, è possibile verificare che tutti i percorsi di rete siano configurati correttamente. In un ambiente SAN, controllare attentamente l'output del comando `sanlun lun`

`show -p` per assicurarsi che tutti i percorsi primario e secondario previsti siano disponibili.

Occorre prestare attenzione quando si rilascia un'acquisizione forzata o uno switchover. Imporre una modifica alla configurazione dello storage con queste opzioni significa che lo stato del controller proprietario delle unità non viene preso in considerazione e il nodo alternativo assume forzatamente il controllo delle unità. Una forzatura non corretta di un takeover può causare la perdita o il danneggiamento dei dati. Questo perché un takeover o uno switchover forzato può scartare il contenuto di NVMEM. Una volta completato il takeover o lo switchover, la perdita dei dati potrebbe riportare i dati memorizzati nelle unità a uno stato leggermente più vecchio dal punto di vista del database.

Raramente dovrebbe essere necessario un takeover forzato con una normale coppia ha. In quasi tutti gli scenari di errore, un nodo si arresta e informa il partner in modo che si verifichi un failover automatico. In alcuni casi, ad esempio in caso di guasto permanente che causa la perdita dell'interconnessione tra i nodi e la perdita di un controller, è necessario eseguire un takeover forzato. In una situazione del genere, il mirroring tra i nodi viene perso prima del guasto del controller, il che significa che il controller rimasto non avrebbe più una copia delle scritture in corso. Il takeover deve quindi essere forzato, il che significa che potenzialmente i dati vengono persi.

La stessa logica si applica a uno switchover MetroCluster. In condizioni normali, lo switchover è quasi trasparente. Tuttavia, un disastro può causare una perdita di connettività tra il sito rimasto e il sito disastroso. Dal punto di vista del sito sopravvissuto, il problema potrebbe essere nient'altro che un'interruzione della connettività tra i siti, e il sito originale potrebbe ancora elaborare i dati. Se un nodo non è in grado di verificare lo stato del controller primario, è possibile solo uno switchover forzato.

NetApp raccomanda adottare le seguenti precauzioni:

- Prestare molta attenzione a non forzare accidentalmente un'acquisizione o uno switchover. In genere, non è necessario forzare e forzare la modifica può causare la perdita di dati.
- Se è necessario un takeover o uno switchover forzato, assicurarsi che le applicazioni vengano arrestate, che tutti i file system vengano dismontati e che i gruppi di volumi LVM (Logical Volume Manager) siano diversi. I gruppi di dischi ASM devono essere smontati.
- In caso di switchover MetroCluster forzato, scollegare il nodo guasto da tutte le risorse di storage rimaste. Per ulteriori informazioni, consultare la Guida alla gestione e al ripristino di emergenza di MetroCluster per la versione pertinente di ONTAP.

MetroCluster e aggregati multipli

MetroCluster è una tecnologia di replica sincrona che passa alla modalità asincrona in caso di interruzione della connettività. Questa è la richiesta più comune da parte dei clienti, perché la replica sincrona garantita significa che l'interruzione della connettività del sito porta a uno stallo completo dell'i/o del database, mettendo il database fuori servizio.

Con MetroCluster, gli aggregati vengono sincronizzati rapidamente dopo il ripristino della connettività. A differenza di altre tecnologie di storage, MetroCluster non dovrebbe mai richiedere un reindirizzamento completo in seguito a un guasto del sito. È necessario spedire solo le modifiche delta.

Nei set di dati estesi agli aggregati c'è solo un piccolo rischio che occorrano passaggi aggiuntivi di recovery dei dati in uno scenario di emergenza regolare. In particolare, se (a) la connettività tra siti viene interrotta, (b) la connettività viene ripristinata, (c) gli aggregati raggiungono uno stato in cui alcuni sono sincronizzati e alcuni non lo sono, quindi (d) il sito primario viene perso, il risultato è un sito sopravvissuto in cui gli aggregati non sono sincronizzati tra loro. In tal caso, parti del set di dati vengono sincronizzate tra loro e non è possibile ripristinare applicazioni, database o datastore senza un recovery. Se un set di dati si estende agli aggregati, NetApp consiglia vivamente di sfruttare i backup basati su snapshot con uno dei molti strumenti disponibili, per

verificare la possibilità di recupero rapido in questo scenario insolito.

Configurazione del database

Dimensioni dei blocchi

ONTAP utilizza internamente dimensioni dei blocchi variabili, il che significa che è possibile configurare i database Oracle con le dimensioni desiderate per i blocchi. Tuttavia, le dimensioni dei blocchi del file system possono influire sulle prestazioni e in alcuni casi una maggiore dimensione dei blocchi di ripristino può migliorare le prestazioni.

Dimensioni dei blocchi di dati

Alcuni sistemi operativi offrono una scelta di dimensioni dei blocchi del file system. Per i file system che supportano i file di dati Oracle, quando si utilizza la compressione le dimensioni del blocco devono essere pari a 8KB KB. Quando la compressione non è necessaria, è possibile utilizzare dimensioni del blocco pari a 8KB K o 4KB K.

Se un file dati viene inserito in un file system con un blocco di 512 byte, è possibile che i file non siano allineati correttamente. Il LUN e il file system potrebbero essere allineati correttamente in base ai consigli di NetApp, ma l'i/o del file non sarebbe allineato correttamente. Un tale disallineamento causerebbe gravi problemi di prestazioni.

I file system che supportano i log di ripristino devono utilizzare una dimensione blocco pari a un multiplo della dimensione del blocco di ripristino. In genere, questo richiede che sia il file system del redo log sia il redo log stesso utilizzino una dimensione del blocco di 512 byte.

Ripristina le dimensioni dei blocchi

A velocità di ripristino molto elevate, è possibile che le dimensioni dei blocchi di 4KB KB funzionino meglio, perché alte velocità di ripristino consentono di eseguire l'i/o in un numero inferiore di operazioni più efficienti. Se le velocità di ripristino sono superiori a 50Mbps KB, valutare la possibilità di testare dimensioni blocco di 4KB KB.

Sono stati identificati alcuni problemi dei clienti con i database che utilizzano i log di redo con dimensioni dei blocchi di 512 byte in un file system con dimensioni dei blocchi di 4KB KB e molte transazioni di dimensioni molto ridotte. L'overhead coinvolto nell'applicazione di modifiche multiple a 512 byte a un singolo blocco di file system da 4KB ha portato a problemi di performance che sono stati risolti modificando il file system in modo da utilizzare dimensioni dei blocchi di 512 byte.

NetApp consiglia di non modificare le dimensioni del blocco di redo se non dietro indicazione di un'organizzazione di assistenza clienti o servizi professionali o se la modifica si basa sulla documentazione ufficiale del prodotto.

db_file_multiblock_read_count

Il `db_file_multiblock_read_count` Parametro controlla il numero massimo di blocchi di database Oracle che Oracle legge come singola operazione durante l'i/o sequenziale

Questo parametro non influisce tuttavia sul numero di blocchi letti da Oracle durante qualsiasi e in tutte le

operazioni di lettura, né sull'i/o casuale Ciò influisce solo sulle dimensioni del blocco degli i/o sequenziali.

Oracle consiglia all'utente di lasciare il parametro non impostato. In questo modo, il software del database può impostare automaticamente il valore ottimale. Questo generalmente significa che questo parametro è impostato su un valore che fornisce una dimensione i/o di 1MB. Ad esempio, una lettura 1MB di 8KB blocchi richiederebbe la lettura di 128 blocchi e il valore predefinito per questo parametro sarebbe 128.

La maggior parte dei problemi di prestazioni del database osservati da NetApp presso le sedi dei clienti implica un'impostazione errata per questo parametro. Ci sono motivi validi per modificare questo valore con le versioni 8 e 9 di Oracle. Di conseguenza, il parametro potrebbe essere inconsapevolmente presente in `init.ora` Perché il database è stato aggiornato in posizione a Oracle 10 e versioni successive. Un'impostazione legacy di 8 o 16, rispetto a un valore predefinito di 128, danneggia significativamente le prestazioni i/o sequenziali.

NetApp recommended impostando `db_file_multiblock_read_count` il parametro non deve essere presente in `init.ora` file. NetApp non ha mai riscontrato una situazione in cui la modifica di questo parametro ha migliorato le prestazioni, ma in molti casi ha causato evidenti danni al throughput i/o sequenziale.

filesystemio_options

Parametro di inizializzazione Oracle `filesystemio_options` Controlla l'utilizzo dell'i/o asincrono e diretto

Contrariamente a quanto si crede, l'i/o asincrono e diretto non si escludono a vicenda. NetApp ha osservato che questo parametro è spesso configurato in modo non corretto negli ambienti dei clienti e che questa errata configurazione è direttamente responsabile di molti problemi di prestazioni.

L'i/o asincrono significa che le operazioni i/o di Oracle possono essere parallelizzate. Prima della disponibilità di i/o asincrono su vari sistemi operativi, gli utenti hanno configurato numerosi processi dbwriter e modificato la configurazione del processo del server. Con l'i/o asincrono, il sistema operativo stesso esegue i/o per conto del software di database in modo altamente efficiente e parallelo. La procedura non pone i dati a rischio e le operazioni critiche, come il logging di redo di Oracle, vengono comunque eseguite in maniera sincrona.

L'i/o diretto ignora la cache buffer del sistema operativo. L'i/o su un sistema UNIX scorre normalmente attraverso la cache del buffer del sistema operativo. Ciò è utile per le applicazioni che non mantengono una cache interna, ma Oracle dispone di una propria cache buffer all'interno di SGA. In quasi tutti i casi, è meglio abilitare l'i/o diretto e allocare la RAM del server all'SGA piuttosto che affidarsi alla cache del buffer del sistema operativo. Oracle SGA utilizza la memoria in modo più efficiente. Inoltre, quando l'i/o fluisce attraverso il buffer del sistema operativo, è soggetto a un'ulteriore elaborazione, che aumenta le latenze. L'aumento delle latenze è particolarmente percepibile con un elevato i/o in scrittura quando una bassa latenza è un requisito critico.

Le opzioni per `filesystemio_options` sono:

- **Async.** Oracle invia le richieste di i/o al sistema operativo per l'elaborazione. Questo processo consente a Oracle di eseguire altri lavori anziché attendere il completamento dell'i/o e quindi aumentare la parallelizzazione i/O.
- **Directio.** Oracle esegue l'i/o direttamente sui file fisici piuttosto che instradare l'i/o attraverso la cache del sistema operativo host.
- **None.** Oracle utilizza i/o sincroni e bufferizzati In questa configurazione, la scelta tra processi server condivisi e dedicati e il numero di dbwriter è più importante.
- **Setall.** Oracle utilizza i/o sia asincrono che diretto In quasi tutti i casi, l'uso di `setall` è ottimale.

Il `filesystemio_options` Parametro non ha effetto negli ambienti DNFS e ASM. L'utilizzo di DNFS o ASM comporta automaticamente l'utilizzo dell'i/o asincrono e diretto

In passato alcuni clienti hanno riscontrato problemi di i/o asincrono, in particolare con le precedenti versioni di Red Hat Enterprise Linux 4 (RHEL4). Alcuni consigli aggiornati su Internet suggeriscono comunque di evitare l'i/o asincrono a causa di informazioni non aggiornate. L'i/o asincrono è stabile su tutti i sistemi operativi correnti. Non c'è motivo di disabilitarlo, in assenza di un bug noto con il sistema operativo.

Se un database utilizza l'i/o con buffer, un passaggio all'i/o diretto potrebbe anche richiedere una modifica delle dimensioni SGA. La disattivazione dell'i/o con buffer elimina i vantaggi prestazionali che la cache del sistema operativo host fornisce al database. L'aggiunta di RAM alla SGA risolve questo problema. Il risultato netto dovrebbe essere un miglioramento delle performance di i/O.

Sebbene sia quasi sempre meglio utilizzare la RAM per Oracle SGA piuttosto che per il caching del buffer del sistema operativo, potrebbe essere impossibile determinare il valore migliore. Ad esempio, potrebbe essere preferibile utilizzare i/o con buffer di dimensioni SGA molto ridotte su un server di database con molte istanze Oracle attive in modo intermittente. Questa disposizione consente l'utilizzo flessibile della RAM disponibile rimanente sul sistema operativo da parte di tutte le istanze di database in esecuzione. Si tratta di una situazione molto insolita, ma è stata osservata presso alcune sedi dei clienti.

NetApp recommended `filesystemio_options` a. `setall`, Ma essere consapevoli che in alcune circostanze la perdita della cache del buffer host potrebbe richiedere un aumento nella SGA di Oracle.

Timeout RAC

Oracle RAC è un prodotto clusterware con diversi tipi di processi heartbeat interni che monitorano lo stato del cluster.

Le informazioni contenute in "errore di montaggio" La sezione contiene informazioni critiche per gli ambienti Oracle RAC che utilizzano lo storage di rete e, in molti casi, è necessario modificare le impostazioni predefinite di Oracle RAC per garantire che il cluster RAC sopravviva alle modifiche del percorso di rete e alle operazioni di failover/switchover dello storage.

disktimeout

Il parametro RAC relativo allo storage primario è `disktimeout`. Questo parametro controlla la soglia entro la quale l'i/o del file di voting deve essere completato. Se il `disktimeout` Il parametro viene superato, quindi il nodo RAC viene eliminato dal cluster. Il valore predefinito per questo parametro è 200. Questo valore dovrebbe essere sufficiente per le procedure standard di takeover e giveback dello storage.

NetApp consiglia di eseguire il test approfondito delle configurazioni RAC prima di metterle in produzione, perché molti fattori influiscono su un takeover o un giveback. Oltre al tempo richiesto per il completamento del failover dello storage, è necessario ulteriore tempo affinché le modifiche LACP (link Aggregation Control Protocol) vengano propagate. Inoltre, il software multipathing SAN deve rilevare un timeout i/o e riprovare su un percorso alternativo. Se un database è estremamente attivo, è necessario accodare e rieseguire una grande quantità di i/o prima di elaborare l'i/o del disco di voting.

Nel caso in cui non sia possibile eseguire un takeover o un giveback effettivo dello storage, l'effetto può essere simulato con test di pull dei cavi sul server di database.

NetApp consiglia quanto segue:

- Lasciando il `disktimeout` parametro al valore predefinito di 200.
- Verificare sempre accuratamente la configurazione di un RAC.

errore di montaggio

Il `misscount` In genere, il parametro influisce solo sull'heartbeat di rete tra i nodi RAC. Il valore predefinito è 30 secondi. Se i binari della griglia si trovano su un array di storage o l'unità di avvio del sistema operativo non è locale, questo parametro potrebbe diventare importante. Ciò comprende host con unità di boot ubicate su una SAN FC, sistemi operativi basati su NFS e unità di boot ubicate in datastore di virtualizzazione, come un file VMDK.

Se l'accesso a un'unità di boot viene interrotto da un takeover o un giveback dello storage, è possibile che la posizione binaria della griglia o l'intero sistema operativo si blocchi temporaneamente. Il tempo necessario affinché ONTAP completi l'operazione di storage e affinché il sistema operativo modifichi i percorsi e riprenda l'i/o potrebbe superare l' `misscount` soglia. Di conseguenza, un nodo viene eliminato immediatamente dopo il ripristino della connettività al LUN di avvio o ai binari della griglia. Nella maggior parte dei casi, l'eviction e il successivo riavvio si verificano senza messaggi di registrazione per indicare il motivo del riavvio. Non tutte le configurazioni sono interessate dal problema, pertanto è possibile testare host basati su boot SAN, NFS o datastore in un ambiente RAC in modo che RAC rimanga stabile in caso di interruzione della comunicazione con il disco di avvio.

Nel caso di unità di avvio non locali o di un hosting di file system non locale `grid` binari, il `misscount` deve essere modificato per corrispondere `disktimeout`. Se questo parametro viene modificato, eseguire ulteriori test per identificare anche eventuali effetti sul comportamento RAC, come il tempo di failover dei nodi.

NetApp consiglia quanto segue:

- Lasciare la `misscount` parametro al valore predefinito di 30 a meno che non si verifichi una delle seguenti condizioni:
  - `grid` I file binari sono collocati in un disco collegato in rete, inclusi dischi NFS, iSCSI, FC e basati su datastore.
  - Il sistema operativo viene avviato SAN.
- In questi casi, valutare l'effetto delle interruzioni di rete che influiscono sull'accesso al sistema operativo o. `GRID_HOME` file system. In alcuni casi, tali interruzioni causano lo stallo dei daemon Oracle RAC, che può portare a un `misscount`timeout e sfratto basati su -. Il valore predefinito del timeout è 27 secondi, ovvero il valore di `misscount` meno `reboottime`. In questi casi, aumentare `misscount` a 200 per la corrispondenza `disktimeout`.

Configurazione dell'host

AIX

Argomenti di configurazione per database Oracle su IBM AIX con ONTAP.

I/o simultanei

Per ottenere prestazioni ottimali su IBM AIX è necessario utilizzare l'i/o simultaneo Senza i/o simultaneo, è probabile che le limitazioni delle prestazioni siano dovute al fatto che AIX esegue i/o atomico serializzato, che comporta un overhead significativo.

In origine, NetApp ha consigliato di utilizzare `cio` Opzione di montaggio per forzare l'uso di i/o simultanei sul file system, ma questo processo ha avuto degli inconvenienti e non è più necessario. Dall'introduzione di AIX 5,2 e Oracle 10gR1, Oracle su AIX può aprire singoli file per i/o simultanei, anziché forzare i/o simultanei sull'intero file system.

Il metodo migliore per abilitare l'i/o simultaneo è impostare `init.ora` parametro `filesystemio_options` a. `setall`. In questo modo, Oracle può aprire file specifici da utilizzare con i/o simultanei

Utilizzo di `cio` Come opzione di montaggio, l'utilizzo di i/o simultanei può avere conseguenze negative. Ad esempio, forzando i/o simultanei si disabilita la lettura dei file system, che può danneggiare le prestazioni dell'i/o al di fuori del software del database Oracle, come la copia dei file e l'esecuzione di backup su nastro. Inoltre, prodotti come Oracle GoldenGate e SAP BR*Tools non sono compatibili con l'uso di `cio` Montare l'opzione con alcune versioni di Oracle.

NetApp consiglia quanto segue:

- Non utilizzare `cio` opzione di montaggio a livello di file system. Abilitare invece l'i/o simultaneo tramite l'utilizzo di `filesystemio_options=setall`.
- Utilizzare solo l' `cio` l'opzione di montaggio dovrebbe essere impostata se non è possibile `filesystemio_options=setall`.

Opzioni di montaggio NFS AIX

Nella tabella seguente sono elencate le opzioni di montaggio NFS AIX per i database Oracle a istanza singola.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, intr`

Nella tabella seguente sono elencate le opzioni di montaggio NFS AIX per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac`

Tipo di file	Opzioni di montaggio
CRS/Voting	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac
Dedicato ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144
Condiviso ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare la cache del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati.

Anche se si utilizza il `cio` montare l'opzione e. `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac`. `noac` è obbligatorio per condiviso `ORACLE_HOME` Implementazioni per facilitare la coerenza di file quali file di password Oracle e. `spfile` file di parametri. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è necessario.

Opzioni di montaggio di AIX jfs/JFS2

Nella tabella seguente sono elencate le opzioni di montaggio di AIX jfs/JFS2.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	Valori predefiniti
File di controllo File di dati Registri di ripristino	Valori predefiniti
ORACLE_HOME	Valori predefiniti

Prima di utilizzare AIX `hdisk` i dispositivi in qualsiasi ambiente, inclusi i database, controllano il parametro `queue_depth`. Questo parametro non è la profondità della coda HBA, bensì la profondità della coda SCSI dell'individuo `hdisk device. Depending on how the LUNs are configured, the value for` `queue_depth` potrebbe essere troppo basso per garantire buone prestazioni. I test hanno dimostrato che il valore ottimale è 64.

HP-UX

Argomenti di configurazione per database Oracle su HP-UX con ONTAP.

Opzioni di montaggio NFS HP-UX

Nella tabella seguente sono elencate le opzioni di montaggio NFS HP-UX per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144,forc edirectio, nointr,suid`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`

Nella tabella seguente sono elencate le opzioni di montaggio NFS HP-UX per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, noac,suid`
File di controllo File di dati Registri di ripristino	`rw, bg,hard, [vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,forcedirectio,suid`
CRS/votazione	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac, forcedirectio,suid`
Dedicato `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`
Condiviso `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,suid`

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` e. `forcedirectio` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac` e. `forcedirectio`.

Il motivo `noac` è obbligatorio per condiviso `ORACLE_HOME` Le distribuzioni consentono di semplificare la coerenza di file quali file di password Oracle e file spfile. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è richiesto.

Opzioni di montaggio VxFS HP-UX

Utilizzare le seguenti opzioni di montaggio per i file system che ospitano file binari Oracle:

```
delaylog,nodatainlog
```

Utilizzare le seguenti opzioni di montaggio per i file system contenenti file di dati, log di ripristino, log di archivio e file di controllo in cui la versione di HP-UX non supporta i/o simultanei:

```
nodatainlog,mincache=direct,convosync=direct
```

Quando l'i/o simultaneo è supportato (VxFS 5.0.1 e versioni successive o con ServiceGuard Storage Management Suite), utilizzare queste opzioni di montaggio per i file system contenenti file di dati, log di ripristino, log di archivio e file di controllo:

```
delaylog,cio
```

Il parametro `db_file_multiblock_read_count` È particolarmente critico negli ambienti VxFS. Oracle consiglia di non impostare questo parametro in Oracle 10g R1 e versioni successive, a meno che non sia diversamente specificato. L'impostazione predefinita con dimensioni blocco Oracle 8KB è 128 KB. Se il valore di questo parametro è forzato a 16 o inferiore, rimuovere l' `convosync=direct` Montare l'opzione perché può danneggiare le prestazioni i/o sequenziali. Questa operazione danneggia altri aspetti delle prestazioni e deve essere eseguita solo se il valore di `db_file_multiblock_read_count` deve essere modificato dal valore predefinito.

Linux

Argomenti di configurazione specifici del sistema operativo Linux.

Tabelle degli slot TCP per Linux NFSv3

Le tabelle degli slot TCP sono l'equivalente di NFSv3 della profondità della coda degli HBA (host Bus Adapter). Queste tabelle controllano il numero di operazioni NFS che possono essere in sospeso in qualsiasi momento. Il valore predefinito è di solito 16, che è troppo basso per ottenere prestazioni ottimali. Il problema opposto si verifica sui kernel Linux più recenti, che possono aumentare automaticamente il limite della tabella degli slot TCP a un livello che satura il server NFS con le richieste.

Per prestazioni ottimali e per evitare problemi di prestazioni, regolare i parametri del kernel che controllano le tabelle degli slot TCP.

Eseguire `sysctl -a | grep tcp.*.slot_table` e osservare i seguenti parametri:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tutti i sistemi Linux dovrebbero includere `sunrpc.tcp_slot_table_entries`, ma solo alcuni includono `sunrpc.tcp_max_slot_table_entries`. Entrambi devono essere impostati su 128.

La mancata impostazione di questi parametri può avere effetti significativi sulle prestazioni. In alcuni casi, le prestazioni sono limitate poiché il sistema operativo linux non fornisce i/o sufficienti In altri casi, le latenze i/o aumentano quando il sistema operativo linux tenta di emettere più i/o di quanto possa essere gestito.

Opzioni di montaggio NFS Linux

Nella tabella seguente sono elencate le opzioni di montaggio NFS Linux per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr`
ORACLE_HOME	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr`

Nella tabella seguente sono elencate le opzioni di montaggio NFS Linux per RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, actimeo=0`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,actimeo=0`
CRS/votazione	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,actimeo=0`
Dedicato `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
Condiviso `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,actimeo=0`

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `actimeo=0` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `actimeo=0`.

Il motivo `actimeo=0` è obbligatorio per condiviso `ORACLE_HOME` Le distribuzioni consentono di semplificare la coerenza di file quali file di password e file spfile di Oracle. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è necessario.

In genere, i file non di database devono essere montati con le stesse opzioni utilizzate per i file di dati a singola istanza, sebbene applicazioni specifiche possano avere requisiti diversi. Evitare le opzioni di montaggio `noac` e. `actimeo=0` se possibile perché queste opzioni disabilitano la lettura e il buffering a livello di file system. Ciò può causare gravi problemi di prestazioni per processi quali l'estrazione, la traduzione e il caricamento.

ACCESSO e GETATTR

Alcuni clienti hanno notato che un livello estremamente elevato di altri IOPS, come ACCESSO e GETATTR, può dominare i propri workload. In casi estremi, operazioni come letture e scritture possono arrivare fino al 10% del totale. Si tratta di un comportamento normale con qualsiasi database che include l'uso di `actimeo=0` e/o. `noac` Su Linux perché queste opzioni fanno sì che il sistema operativo Linux ricarichi costantemente i metadati dei file dal sistema di archiviazione. Operazioni quali ACCESS e GETATTR sono operazioni a basso impatto gestite dalla cache ONTAP in un ambiente di database. Non dovrebbero essere considerati IOPS autentici, come le letture e le scritture, che creano una vera domanda sui sistemi storage. Tuttavia, questi altri IOPS creano un certo carico, specialmente negli ambienti RAC. Per risolvere questo problema, abilitare DNFS, che ignora la cache buffer del sistema operativo ed evita queste operazioni non necessarie relative ai metadati.

Linux Direct NFS

Un'opzione di montaggio aggiuntiva, denominata `nosharecache`, È necessario quando (a) DNFS è abilitato e (b) un volume sorgente è montato più di una volta su un singolo server (c) con un mount NFS nidificato. Questa configurazione si osserva principalmente in ambienti che supportano applicazioni SAP. Ad esempio, un singolo volume di un sistema NetApp può avere una directory situata in `/vol/oracle/base` e un secondo a. `/vol/oracle/home`. Se `/vol/oracle/base` è montato su `/oracle` e. `/vol/oracle/home` è montato su `/oracle/home`, Il risultato sono montaggi NFS nidificati che hanno origine sulla stessa fonte.

Il sistema operativo è in grado di rilevare che `/oracle` e `/oracle/home` si trovano sullo stesso volume, che è lo stesso file system di origine. Il sistema operativo utilizza quindi lo stesso handle di dispositivo per l'accesso ai dati. In questo modo si migliora l'uso della cache del sistema operativo e di alcune altre operazioni, ma interferisce con DNFS. Se DNFS deve accedere a un file, come ad esempio `spfile` , on , `/oracle/home` potrebbe erroneamente tentare di utilizzare il percorso errato dei dati. Il risultato è un'operazione i/o non riuscita. In queste configurazioni, Aggiungi l' `nosharecache`opzione di montaggio a qualsiasi file system NFS che condivide un volume di origine con un altro file system NFS su quell'host. In questo modo, il sistema operativo Linux assegna un handle di dispositivo indipendente al file system.

Linux Direct NFS e Oracle RAC

L'uso di DNFS offre speciali vantaggi in termini di prestazioni per Oracle RAC sul sistema operativo Linux, poiché Linux non dispone di un metodo per forzare l'i/o diretto, necessario con RAC per la coerenza tra i nodi. Come soluzione, Linux richiede l'uso di `actimeo=0` Opzione di montaggio, che fa sì che i dati dei file scadano immediatamente dalla cache del sistema operativo. Questa opzione a sua volta obbliga il client NFS Linux a rileggere costantemente i dati degli attributi, danneggiando la latenza e aumentando il carico sullo storage controller.

Abilitando DNFS si ignora il client NFS dell'host ed evita questo danno. Diversi clienti hanno segnalato significativi miglioramenti delle performance sui cluster RAC e una significativa riduzione del carico ONTAP (soprattutto in relazione ad altri IOPS) quando si attiva DNFS.

Linux Direct NFS e file oranfstab

Quando si utilizza DNFS su Linux con l'opzione multipathing, è necessario utilizzare più sottoreti. Su altri sistemi operativi, è possibile stabilire più canali DNFS utilizzando `LOCAL` e. `DONTROUTE` Opzioni per configurare più canali DNFS su una singola subnet. Tuttavia, questo non funziona correttamente su Linux e

possono verificarsi problemi di prestazioni imprevisti. Con Linux, ogni NIC utilizzata per il traffico DNFS deve trovarsi su una subnet diversa.

Utilità di pianificazione i/O.

Il kernel Linux permette un controllo di basso livello sul modo in cui l'i/o blocca i dispositivi è programmato. Le impostazioni predefinite su varie distribuzioni di Linux variano notevolmente. I test dimostrano che la scadenza di solito offre i migliori risultati, ma a volte NOOP è stato leggermente migliore. La differenza di prestazioni è minima, ma è necessario verificare entrambe le opzioni se è necessario estrarre le massime prestazioni possibili da una configurazione di database. CFQ è l'impostazione predefinita in molte configurazioni e ha dimostrato di avere problemi significativi di prestazioni con i carichi di lavoro del database.

Per istruzioni sulla configurazione dello scheduler i/o, consultare la documentazione del fornitore di Linux pertinente.

Multipathing

Alcuni clienti hanno riscontrato arresti anomali durante l'interruzione della rete perché il daemon multipath non era in esecuzione sul proprio sistema. Nelle versioni recenti di Linux, il processo di installazione del sistema operativo e del demone multipathing potrebbero lasciare questi sistemi operativi vulnerabili a questo problema. I pacchetti sono installati correttamente, ma non sono configurati per l'avvio automatico dopo un riavvio.

Ad esempio, il valore predefinito per il daemon multipath su RHEL5,5 potrebbe essere il seguente:

```
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:off   3:off   4:off   5:off   6:off
```

Questo può essere corretto con i seguenti comandi:

```
[root@host1 iscsi]# chkconfig multipathd on
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:on    3:on    4:on    5:on    6:off
```

Mirroring ASM

Il mirroring ASM potrebbe richiedere modifiche alle impostazioni di multipath Linux per consentire ad ASM di riconoscere un problema e passare a un gruppo di errori alternativo. La maggior parte delle configurazioni ASM su ONTAP utilizza la ridondanza esterna, il che significa che la protezione dei dati è fornita dall'array esterno e ASM non esegue il mirroring dei dati. Alcuni siti utilizzano ASM con ridondanza normale per fornire il mirroring bidirezionale, in genere su siti diversi.

Le impostazioni di Linux visualizzate nella "Documentazione delle utilità host NetApp" Includi parametri multipath che determinano indefinite code di i/O. Ciò significa che un i/o su un dispositivo LUN senza percorsi attivi attende finché l'i/o non viene completato. Questo è solitamente consigliabile perché gli host Linux attendono il tempo necessario per il completamento delle modifiche al percorso SAN, per il riavvio degli switch FC o per il completamento di un failover da parte di un sistema di storage.

Questo comportamento di accodamento illimitato causa un problema con il mirroring ASM perché ASM deve ricevere un errore di i/o per consentire al reparto IT di riprovare l'i/o su un LUN alternativo.

Impostare i seguenti parametri in Linux `multipath.conf` File per i LUN ASM utilizzati con il mirroring ASM:

```
polling_interval 5
no_path_retry 24
```

Queste impostazioni creano un timeout di 120 secondi per i dispositivi ASM. Il timeout viene calcolato come `polling_interval` * `no_path_retry` in pochi secondi. In alcuni casi potrebbe essere necessario regolare il valore esatto, ma per la maggior parte degli utilizzi dovrebbe essere sufficiente un timeout di 120 secondi. In particolare, 120 secondi devono consentire il takeover o il giveback del controller senza produrre un errore di i/o che porterebbe il gruppo guasto a diventare offline.

Un più basso `no_path_retry` Il valore può ridurre il tempo richiesto per ASM per passare a un gruppo di errori alternativo, ma aumenta anche il rischio di un failover indesiderato durante attività di manutenzione come il takeover di un controller. Il rischio può essere mitigato tramite un attento monitoraggio dello stato di mirroring ASM. Se si verifica un failover indesiderato, è possibile risincronizzare rapidamente i mirror se la risincronizzazione viene eseguita in modo relativamente rapido. Per ulteriori informazioni, consultare la documentazione Oracle su ASM Fast Mirror Resync per la versione del software Oracle in uso.

Linux xfs, ext3, e ext4 opzioni di mount

NetApp recommended usando le opzioni di mount predefinite.

ASMILib/AFD (driver filtro ASM)

Argomenti di configurazione specifici per il sistema operativo Linux utilizzando AFD e ASMlib

Dimensioni dei blocchi ASMlib

ASMlib è una libreria di gestione ASM opzionale e le utilità associate. Il suo valore principale è la capacità di contrassegnare un LUN o un file basato su NFS come una risorsa ASM con un'etichetta leggibile da un utente.

Le versioni recenti di ASMlib rilevano un parametro LUN chiamato Logical Blocks per Physical Block Exponent (LBPPBE). Questo valore non è stato segnalato dal target SCSI ONTAP fino a poco tempo fa. Ora restituisce un valore che indica che è preferibile una dimensione blocco 4KB. Questa non è una definizione della dimensione del blocco, ma è un suggerimento per qualsiasi applicazione che utilizza LBPPBE che i/o di una certa dimensione potrebbero essere gestiti in modo più efficiente. ASMlib, tuttavia, interpreta LBPPBE come dimensione del blocco e contrassegna in modo permanente l'intestazione ASM quando viene creato il dispositivo ASM.

Questo processo può causare problemi di aggiornamento e migrazione in vari modi, tutti basati sull'impossibilità di combinare dispositivi ASMlib con dimensioni dei blocchi diverse nello stesso gruppo di dischi ASM.

Ad esempio, gli array meno recenti generalmente riportavano un valore LBPPBE pari a 0 o non riportavano affatto questo valore. ASMlib lo interpreta come una dimensione di blocco di 512 byte. Gli array più recenti dovrebbero essere interpretati come aventi una dimensione del blocco di 4KB KB. Non è possibile combinare dispositivi a 512 byte e 4KB nello stesso gruppo di dischi ASM. In questo modo, si impedirebbe a un utente di aumentare le dimensioni del gruppo di dischi ASM utilizzando LUN di due array o sfruttando ASM come strumento di migrazione. In altri casi, RMAN potrebbe non consentire la copia dei file tra un gruppo di dischi ASM con dimensioni del blocco di 512 byte e un gruppo di dischi ASM con dimensioni del blocco di 4KB KB.

La soluzione preferita è quella di tamponare ASMlib. L'ID del bug di Oracle è 13999609 e la patch è presente

in oracleasm-support-2,1.8-1 e versioni successive. Questo patch consente all'utente di impostare il parametro `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` a. `true` in `/etc/sysconfig/oracleasm` file di configurazione. In questo modo, ASMlib non utilizza il parametro LBPPBE, il che significa che i LUN del nuovo array sono ora riconosciuti come dispositivi a blocchi da 512 byte.

> L'opzione non modifica le dimensioni del blocco sui LUN precedentemente contrassegnati da ASMlib. Ad esempio, se un gruppo di dischi ASM con blocchi da 512 byte deve essere migrato in un nuovo sistema di storage che riporta un blocco da 4KB KB, è possibile scegliere questa opzione `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` Deve essere impostato prima che i nuovi LUN siano contrassegnati con ASMlib. Se i dispositivi sono già stati contrassegnati da oracleasm, è necessario riformattarli prima di essere contrassegnati con una nuova dimensione del blocco. Innanzitutto, deconfigurare il dispositivo con `oracleasm deletedisk`, E quindi cancellare i primi 1GB del dispositivo con `dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`. Infine, se il dispositivo era stato precedentemente partizionato, utilizzare `kpartx` Per rimuovere le partizioni obsolete o semplicemente riavviare il sistema operativo.

Se ASMlib non può essere aggiornato, ASMlib può essere rimosso dalla configurazione. Questa modifica comporta un'interruzione e richiede la rimozione dello stampaggio dei dischi ASM e la verifica che `asm_diskstring` parametro impostato correttamente. Questa modifica, tuttavia, non richiede la migrazione dei dati.

Dimensioni blocco comando filtro ASM (AFD)

AFD è una libreria di gestione ASM opzionale che sta diventando il sostituto di ASMlib. Dal punto di vista dello storage, è molto simile ad ASMlib, ma include funzionalità aggiuntive come la capacità di bloccare i/o non Oracle per ridurre le possibilità di errori di utenti o applicazioni che potrebbero danneggiare i dati.

Dimensioni dei blocchi dei dispositivi

Come ASMlib, anche AFD legge il parametro LUN Logical Blocks per Physical Block Exponent (LBPPBE) e per impostazione predefinita utilizza la dimensione fisica del blocco, non la dimensione logica del blocco.

Ciò potrebbe creare un problema se l'AFD viene aggiunto a una configurazione esistente in cui i dispositivi ASM sono già formattati come dispositivi a blocchi da 512 byte. Il driver AFD riconosce il LUN come un dispositivo 4K e la mancata corrispondenza tra l'etichetta ASM e il dispositivo fisico impedirebbe l'accesso. Allo stesso modo, le migrazioni sarebbero influenzate dal fatto che non è possibile combinare dispositivi a 512 byte e 4KB nello stesso gruppo di dischi ASM. In questo modo, si impedirebbe a un utente di aumentare le dimensioni del gruppo di dischi ASM utilizzando LUN di due array o sfruttando ASM come strumento di migrazione. In altri casi, RMAN potrebbe non consentire la copia dei file tra un gruppo di dischi ASM con dimensioni del blocco di 512 byte e un gruppo di dischi ASM con dimensioni del blocco di 4KB KB.

La soluzione è semplice: AFD include un parametro per controllare se utilizza le dimensioni del blocco logico o fisico. Si tratta di un parametro globale che interessa tutti i dispositivi del sistema. Per forzare AFD a utilizzare le dimensioni del blocco logico, impostare `options oracleafd oracleafd_use_logical_block_size=1` in `/etc/modprobe.d/oracleafd.conf` file.

Dimensioni di trasferimento multipath

Le recenti modifiche al kernel linux impongono restrizioni delle dimensioni di i/o inviate ai dispositivi multipath e AFD non rispetta queste restrizioni. Gli i/o vengono quindi rifiutati, il che causa la disconnessione del percorso LUN. Il risultato è un'impossibilità di installare Oracle Grid, configurare ASM o creare un database.

La soluzione consiste nel specificare manualmente la lunghezza massima di trasferimento nel file

multipath.conf per i LUN ONTAP:

```
devices {
            device {
                vendor "NETAPP"
                product "LUN.*"
                max_sectors_kb 4096
        }
    }
```

Anche se attualmente non esistono problemi, questo parametro deve essere impostato se si utilizza AFD per garantire che un futuro aggiornamento linux non causi inaspettatamente problemi.

Microsoft Windows

Argomenti di configurazione per database Oracle su Microsoft Windows con ONTAP.

NFS

Oracle supporta l'utilizzo di Microsoft Windows con il client NFS diretto. Questa funzionalità offre un percorso per i vantaggi di gestione di NFS, tra cui la possibilità di visualizzare i file tra più ambienti, ridimensionare dinamicamente i volumi e sfruttare un protocollo IP meno costoso. Consultare la documentazione ufficiale di Oracle per informazioni sull'installazione e la configurazione di un database in Microsoft Windows utilizzando DNFS. Non esistono Best practice speciali.

SAN

Per un'efficienza di compressione ottimale, assicurarsi che il file system NTFS utilizzi un'unità di allocazione di 8K GB o superiore. L'utilizzo di un'unità di allocazione 4K, generalmente predefinita, influisce negativamente sull'efficienza della compressione.

Solaris

Argomenti di configurazione specifici di Solaris.

Opzioni di montaggio NFS Solaris

Nella tabella seguente sono elencate le opzioni di montaggio di Solaris NFS per una singola istanza.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	`rw,bg,hard,[vers=3,vers=4.1], roto=tcp, timeo=600,rsize=262144,wsize=262144`
File di controllo File di dati Registri di ripristino	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,llock,suid`

Tipo di file	Opzioni di montaggio
ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,suid

L'utilizzo di `llock` è stato dimostrato di migliorare drasticamente le performance negli ambienti dei clienti rimuovendo la latenza associata all'acquisizione e al rilascio di blocchi sul sistema storage. Utilizzare questa opzione con attenzione negli ambienti in cui sono configurati numerosi server per montare gli stessi file system e Oracle è configurato per montare questi database. Sebbene si tratti di una configurazione molto insolita, viene utilizzata da un numero limitato di clienti. Se un'istanza viene avviata accidentalmente una seconda volta, i dati potrebbero danneggiarsi perché Oracle non è in grado di rilevare i file di blocco sul server esterno. I blocchi NFS non offrono altrimenti protezione; come nella versione 3 di NFS, sono solo di natura consultiva.

Perché il `llock` e. `forcedirectio` i parametri si escludono a vicenda, è importante che `filesystemio_options=setall` è presente in `init.ora` file in modo che `directio` viene utilizzato. Senza questo parametro, viene utilizzato il caching del buffer del sistema operativo host e le prestazioni possono essere compromesse.

Nella tabella seguente sono elencate le opzioni di montaggio Solaris NFS RAC.

Tipo di file	Opzioni di montaggio
Pagina iniziale ADR	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,noac
File di controllo File di dati Registri di ripristino	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,forcedirectio
CRS/votazione	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,forcedirectio
Dedicato ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,suid
Condiviso ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,suid

L'aggiunta fa la differenza principale tra le opzioni di montaggio RAC e a istanza singola `noac` e. `forcedirectio` alle opzioni di montaggio. Questa aggiunta ha l'effetto di disabilitare il caching del sistema operativo host, consentendo a tutte le istanze nel cluster RAC di avere una visione coerente dello stato dei dati. Anche se si utilizza il `init.ora` parametro `filesystemio_options=setall` ha lo stesso effetto di disabilitare la cache dell'host, è comunque necessario utilizzare `noac` e. `forcedirectio`.

Il motivo `actimeo=0` è obbligatorio per condiviso `ORACLE_HOME` Le distribuzioni consentono di semplificare la coerenza di file quali file di password Oracle e file spfile. Se ogni istanza di un cluster RAC dispone di un'istanza dedicata `ORACLE_HOME`, questo parametro non è richiesto.

Opzioni di montaggio UFS di Solaris

NetApp consiglia vivamente di utilizzare l'opzione di montaggio della registrazione in modo che l'integrità dei dati venga preservata in caso di arresto anomalo dell'host Solaris o di interruzione della connettività FC. L'opzione di montaggio della registrazione preserva anche l'usabilità dei backup Snapshot.

Solaris ZFS

Solaris ZFS deve essere installato e configurato con attenzione per garantire prestazioni ottimali.

mvector

Solaris 11 ha introdotto una modifica nel modo in cui elabora operazioni i/o di grandi dimensioni, che può causare gravi problemi di prestazioni sugli array di storage SAN. Il problema è documentato nel rapporto 630173 del bug di monitoraggio di NetApp, "regressione delle prestazioni di Solaris 11 ZFS".

Questo non è un bug di ONTAP. Si tratta di un difetto di Solaris rilevato in Solaris Defects 7199305 e 7082975.

È possibile consultare il supporto Oracle per scoprire se la versione di Solaris 11 in uso è interessata o per verificare la soluzione alternativa, passando `zfs_mvector_max_size` a un valore inferiore.

È possibile farlo eseguendo il seguente comando come root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t131072" |mdb -kw
```

Se da questa modifica emergono problemi imprevisti, è possibile annullarli facilmente eseguendo il seguente comando come root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t1048576" |mdb -kw
```

Kernel

Prestazioni ZFS affidabili richiedono un kernel Solaris con patch contro i problemi di allineamento LUN. La correzione è stata introdotta con la patch 147440-19 in Solaris 10 e con SRU 10,5 per Solaris 11. Utilizzare solo Solaris 10 e versioni successive con ZFS.

Configurazione del LUN

Per configurare un LUN, attenersi alla seguente procedura:

1. Creare un LUN di tipo `solaris`.
2. Installare l'host Utility Kit (HUK) appropriato specificato da "Tool di matrice di interoperabilità NetApp (IMT)".
3. Seguire esattamente le istruzioni nell'HUK come descritto. I passaggi di base sono descritti di seguito, ma fare riferimento a. "documentazione più recente" per la procedura corretta.
   a. Eseguire `host_config` utilità per aggiornare `sd.conf/sdd.conf` file. Questo consente alle unità SCSI di rilevare correttamente i LUN ONTAP.
   b. Seguire le istruzioni fornite da `host_config` Utility per abilitare l'input/output multipath (MPIO).

c. Reboot (Riavvia). Questa fase è necessaria per consentire il riconoscimento di eventuali modifiche nel sistema.

4. Partizionare i LUN e verificare che siano allineati correttamente. Vedere "Appendice B: Verifica dell'allineamento WAFL" per istruzioni su come eseguire direttamente il test e confermare l'allineamento.

zpool

Uno zpool deve essere creato solo dopo i passaggi nella "Configurazione LUN" vengono eseguite. Se la procedura non viene eseguita correttamente, le prestazioni potrebbero peggiorare notevolmente a causa dell'allineamento i/O. Per ottenere prestazioni ottimali con ONTAP è necessario allineare l'i/o a un confine di 4K su un'unità. I file system creati su uno zpool utilizzano una dimensione di blocco effettiva controllata tramite un parametro chiamato `ashift`, che può essere visualizzato eseguendo il comando `zdb -C`.

Il valore di `ashift` il valore predefinito è 9, ovvero 2^9 o 512 byte. Per prestazioni ottimali, la `ashift` Il valore deve essere 12 (2^12=4K). Questo valore viene impostato al momento della creazione di zpool e non può essere modificato, il che significa che i dati in zpool con `ashift` oltre a 12 deve essere eseguita la migrazione copiando i dati in uno zpool appena creato.

Dopo aver creato uno zpool, verificare il valore di `ashift` prima di procedere. Se il valore non è 12, i LUN non sono stati rilevati correttamente. Distruggere lo zpool, verificare che tutti i passaggi indicati nella relativa documentazione delle utilità host siano stati eseguiti correttamente e ricreare lo zpool.

Zpool e LDOM Solaris

Gli LDOM di Solaris creano un requisito aggiuntivo per assicurarsi che l'allineamento i/o sia corretto. Sebbene un LUN possa essere rilevato correttamente come un dispositivo 4K, un dispositivo vdsk virtuale su un LDOM non eredita la configurazione dal dominio i/O. Vdsk basato su tale LUN torna per impostazione predefinita a un blocco da 512 byte.

È necessario un file di configurazione aggiuntivo. In primo luogo, i singoli LDOM devono essere aggiornati per Oracle bug 15824910 per abilitare le opzioni di configurazione aggiuntive. Questa patch è stata trasferita in tutte le versioni attualmente utilizzate di Solaris. Una volta installato il software LDOM, è pronto per la configurazione dei nuovi LUN correttamente allineati come segue:

1. Identificare il LUN o i LUN da utilizzare nel nuovo zpool. In questo esempio, si tratta del dispositivo c2d1.

```
[root@LDOM1 ~]# echo | format
Searching for disks...done
AVAILABLE DISK SELECTIONS:
  0. c2d0 <Unknown-Unknown-0001-100.00GB>
     /virtual-devices@100/channel-devices@200/disk@0
  1. c2d1 <SUN-ZFS Storage 7330-1.0 cyl 1623 alt 2 hd 254 sec 254>
     /virtual-devices@100/channel-devices@200/disk@1
```

2. Recuperare l'istanza vdc dei dispositivi da utilizzare per un pool ZFS:

```
[root@LDOM1 ~]#  cat /etc/path_to_inst
#
# Caution! This file contains critical kernel state
#
"/fcoe" 0 "fcoe"
"/iscsi" 0 "iscsi"
"/pseudo" 0 "pseudo"
"/scsi_vhci" 0 "scsi_vhci"
"/options" 0 "options"
"/virtual-devices@100" 0 "vnex"
"/virtual-devices@100/channel-devices@200" 0 "cnex"
"/virtual-devices@100/channel-devices@200/disk@0" 0 "vdc"
"/virtual-devices@100/channel-devices@200/pciv-communication@0" 0 "vpci"
"/virtual-devices@100/channel-devices@200/network@0" 0 "vnet"
"/virtual-devices@100/channel-devices@200/network@1" 1 "vnet"
"/virtual-devices@100/channel-devices@200/network@2" 2 "vnet"
"/virtual-devices@100/channel-devices@200/network@3" 3 "vnet"
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc" << We want
this one
```

3. Modifica `/platform/sun4v/kernel/drv/vdc.conf`:

```
block-size-list="1:4096";
```

Ciò significa che all'istanza di dispositivo 1 viene assegnata una dimensione di blocco di 4096.

Come ulteriore esempio, si supponga che le istanze vdsk da 1 a 6 debbano essere configurate per una dimensione di blocco di 4K e. `/etc/path_to_inst` recita:

```
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc"
"/virtual-devices@100/channel-devices@200/disk@2" 2 "vdc"
"/virtual-devices@100/channel-devices@200/disk@3" 3 "vdc"
"/virtual-devices@100/channel-devices@200/disk@4" 4 "vdc"
"/virtual-devices@100/channel-devices@200/disk@5" 5 "vdc"
"/virtual-devices@100/channel-devices@200/disk@6" 6 "vdc"
```

4. La finale `vdc.conf` il file deve contenere quanto segue:

```
block-size-list="1:8192","2:8192","3:8192","4:8192","5:8192","6:8192";
```

Attenzione
L’LDOM deve essere riavviato dopo la configurazione di vdc.conf e la creazione di vdsk. Questa fase non può essere evitata. La modifica delle dimensioni del blocco ha effetto solo dopo un riavvio. Procedere con la configurazione di zpool e accertarsi che l’ashift sia impostato correttamente su 12 come descritto in precedenza.

ZFS Intent Log (ZIL)

In genere, non esiste alcun motivo per individuare ZFS Intent Log (ZIL) su un dispositivo diverso. Il registro può condividere lo spazio con il pool principale. L’uso principale di una ZIL separata è quando si utilizzano unità fisiche che non dispongono delle funzionalità di cache di scrittura nei moderni array di storage.

logbias

Impostare `logbias` Parametro sui file system ZFS che ospitano dati Oracle.

```
zfs set logbias=throughput <filesystem>
```

L’utilizzo di questo parametro riduce i livelli di scrittura complessivi. Per impostazione predefinita, i dati scritti vengono salvati prima nella ZIL e quindi nel pool di storage principale. Questo approccio è appropriato per una configurazione che utilizza una configurazione a disco normale, che include un dispositivo ZIL basato su SSD e supporti rotanti per il pool di storage principale. Questo perché consente l’esecuzione di un commit in una singola transazione i/o sul supporto con latenza più bassa disponibile.

Quando si utilizza un moderno storage array che include funzionalità di caching autonome, questo approccio generalmente non è necessario. In rare circostanze, potrebbe essere opportuno assegnare una scrittura con una singola transazione al registro, ad esempio un carico di lavoro costituito da scritture casuali altamente concentrate e sensibili alla latenza. Vi sono conseguenze sotto forma di amplificazione in scrittura poiché i dati registrati vengono infine scritti nel pool di archiviazione principale, con il risultato di raddoppiare l’attività di scrittura.

I/o diretto

Molte applicazioni, inclusi i prodotti Oracle, possono bypassare la cache del buffer host attivando l’i/o diretto Questa strategia non funziona come previsto con i file system ZFS. Anche se la cache del buffer host viene ignorata, ZFS continua a memorizzare i dati nella cache. Questa azione può produrre risultati fuorvianti quando si utilizzano strumenti come fio o sio per eseguire test delle prestazioni perché è difficile prevedere se l’i/o raggiunge il sistema di storage o se viene memorizzato nella cache locale del sistema operativo. Questa azione rende inoltre molto difficile l’utilizzo di tali test sintetici per confrontare le prestazioni di ZFS con altri file system. In pratica, le performance del file system differiscono da poco a nulla per i carichi di lavoro degli utenti reali.

Diversi zpool

Backup basati su snapshot, ripristini, cloni e archiviazione dei dati basati su ZFS devono essere eseguiti al livello di zpool e in genere richiedono più zpool. Uno zpool è analogo a un gruppo di dischi LVM e deve essere configurato utilizzando le stesse regole. Ad esempio, è probabilmente meglio disporre un database con i file di dati residenti su `zpool1` e i log di archivio, i file di controllo e i log di ripristino che risiedono su `zpool2`. Questo approccio consente un backup a caldo standard in cui il database viene posto in modalità hot backup, seguito da uno snapshot di `zpool1`. Il database viene quindi rimosso dalla modalità di backup a caldo, l’archivio di log viene forzato e viene creata una snapshot di `zpool2` viene creato. Un’operazione di ripristino

richiede lo smontaggio dei file system zfs e l'offlining completo di zpool, in seguito a un'operazione di ripristino di SnapRestore. Lo zpool può quindi essere portato nuovamente online e il database recuperato.

filesystemio_options

Parametro Oracle `filesystemio_options` Funziona in modo diverso con ZFS. Se `setall` oppure `directio` Viene utilizzato, le operazioni di scrittura sono sincrone e ignorano la cache del buffer del sistema operativo, ma le letture sono bufferizzate da ZFS. Questa azione causa difficoltà nell'analisi delle performance perché talvolta l'i/o viene intercettato e gestito dalla cache ZFS, rendendo la latenza dello storage e l'i/o totale inferiori a quanto pare.

Configurazione di rete

Interfacce logiche

I database Oracle devono accedere allo storage. Le interfacce logiche (LIF) sono le tubazioni di rete che collegano una Storage Virtual Machine (SVM) alla rete e a loro volta al database. La corretta progettazione della LIF è necessaria per garantire una larghezza di banda sufficiente per ogni carico di lavoro del database e il failover non comporta una perdita dei servizi storage.

Questa sezione offre una panoramica dei principali principi di progettazione della LIF. Per una documentazione più completa, vedere "Documentazione di gestione della rete ONTAP". Come per altri aspetti dell'architettura dei database, le migliori opzioni per la progettazione di una Storage Virtual Machine (SVM, nota come vserver all'interfaccia della CLI) e di un'interfaccia logica (LIF) dipendono in gran parte dai requisiti di scalabilità e dalle esigenze di business.

Durante la creazione di una strategia LIF, prendi in considerazione i seguenti argomenti principali:

- **Performance.** la larghezza di banda della rete è sufficiente?
- **Resilienza.** ci sono singoli punti di guasto nel progetto?
- **Gestibilità.** la rete può essere scalata senza interruzioni?

Gli argomenti trattati sono relativi alla soluzione end-to-end, dall'host fino agli switch fino al sistema storage.

Tipi di LIF

Esistono diversi tipi di LIF. "Documentazione ONTAP sui tipi di LIF" Fornisci informazioni più complete su questo argomento, ma da un punto di vista funzionale le LIF possono essere divise in gruppi:

- **LIF di gestione cluster e nodi.** LIF utilizzati per gestire il cluster storage.
- **LIF di gestione SVM.** interfacce che consentono l'accesso a una SVM tramite l'API REST o ONTAPI (nota anche come ZAPI) per funzioni come la creazione di snapshot o il ridimensionamento del volume. Prodotti come SnapManager for Oracle (SMO) devono avere accesso a una LIF di gestione SVM.
- **Interfacce LIF dati** per FC, iSCSI, NVMe/FC, NVMe/TCP, NFS, o dati SMB/CIFS.

Una LIF dati utilizzata per il traffico NFS può anche essere utilizzata per la gestione cambiando la policy del firewall da `data` a. `mgmt` O un'altra policy che consente HTTP, HTTPS o SSH. Questa modifica può semplificare la configurazione di rete evitando la configurazione di ciascun host per l'accesso sia alla LIF dati NFS che a una LIF di gestione separata. Non è possibile configurare un'interfaccia sia per iSCSI che per il traffico di gestione, nonostante entrambi utilizzino un protocollo IP. Negli ambienti iSCSI è necessaria una LIF di gestione separata.

Progettazione della SAN LIF

Il design di LIF in un ambiente SAN è relativamente semplice per un motivo: Il multipathing. Tutte le moderne implementazioni SAN consentono a un client di accedere ai dati su più percorsi di rete indipendenti e di selezionare i percorsi migliori per l'accesso. Di conseguenza, le performance rispetto alla progettazione LIF sono più semplici da gestire, perché i client SAN bilanciano automaticamente il carico dell'i/o nei migliori percorsi disponibili.

Se un percorso non è disponibile, il client seleziona automaticamente un percorso diverso. Grazie alla sua semplicità di progettazione, le LIF SAN sono generalmente più gestibili. Ciò non significa che un ambiente SAN sia sempre più facile da gestire, poiché vi sono molti altri aspetti dello storage SAN che sono molto più complicati di NFS. Significa semplicemente che la progettazione della SAN LIF è più semplice.

Performance

La considerazione più importante riguardo le performance di una LIF in un ambiente SAN è la larghezza di banda. Ad esempio, un cluster ONTAP AFF a due nodi con due porte FC da 16GB GB per nodo offre fino a 32GB Gbps di larghezza di banda da/per ciascun nodo.

Resilienza

Le LIF SAN non eseguono il failover su un sistema storage AFF. In caso di guasto di una LIF SAN a causa del failover del controller, il software multipath del client rileva la perdita di un percorso e reindirizza l'i/o a una diversa LIF. Con i sistemi storage ASA, il failover delle LIF dopo un breve ritardo, ma ciò non interrompe l'io perché ci sono percorsi già attivi sull'altro controller. Il processo di failover viene eseguito per ripristinare l'accesso dell'host su tutte le porte definite.

Gestibilità

La migrazione LIF è un task molto più comune in un ambiente NFS, perché la migrazione LIF è spesso associata alla riallocazione dei volumi nel cluster. Non è necessario migrare una LIF in un ambiente SAN quando i volumi vengono ricollocati nella coppia ha. Questo perché, una volta completato lo spostamento del volume, ONTAP invia una notifica alla SAN in merito a una modifica dei percorsi e i client SAN vengono automaticamente risottimizzati. La migrazione LIF con SAN è associata principalmente a importanti modifiche hardware fisiche. Ad esempio, per eseguire un upgrade senza interruzioni dei controller, viene eseguita la migrazione di una SAN LIF nel nuovo hardware. Se una porta FC è guasta, una LIF può essere migrata a una porta inutilizzata.

Raccomandazioni di progettazione

NetApp formula i seguenti consigli:

- Non creare più percorsi di quelli richiesti. Un numero eccessivo di percorsi complica la gestione complessiva e può causare problemi con il failover del percorso su alcuni host. Inoltre, alcuni host hanno limitazioni inattese del percorso per configurazioni come l'avvio SAN.
- Un numero molto ridotto di configurazioni deve richiedere più di quattro percorsi a un LUN. Il valore di avere più di due nodi che pubblicizzano i percorsi delle LUN è limitato perché l'aggregato che ospita un

LUN è inaccessibile in caso di guasto del nodo proprietario del LUN e del partner ha. In una situazione del genere, la creazione di percorsi su nodi diversi dalla coppia ha primaria non è d'aiuto.

- Sebbene il numero di percorsi LUN visibili possa essere gestito selezionando le porte incluse nelle zone FC, in genere è più semplice includere tutti i potenziali punti di destinazione nella zona FC e controllare la visibilità delle LUN a livello ONTAP.
- In ONTAP 8,3 e versioni successive, la funzione SLM (Selective LUN mapping) è quella predefinita. Con SLM, ogni nuova LUN viene automaticamente pubblicizzata dal nodo proprietario dell'aggregato sottostante e del partner ha del nodo. Questa disposizione evita la necessità di creare set di porte o configurare la suddivisione in zone per limitare l'accessibilità delle porte. Ogni LUN è disponibile sul numero minimo di nodi necessari per performance e resilienza ottimali.
  *Nel caso in cui sia necessario migrare un LUN all'esterno dei due controller, è possibile aggiungere i nodi aggiuntivi con `lun mapping add-reporting-nodes` In modo che le LUN vengano pubblicizzate sui nuovi nodi. In questo modo si creano ulteriori percorsi SAN alle LUN per la migrazione delle LUN. Tuttavia, l'host deve eseguire un'operazione di rilevamento per utilizzare i nuovi percorsi.
- Non preoccupatevi eccessivamente del traffico indiretto. Si consiglia di evitare il traffico indiretto in un ambiente i/o-intensive per il quale è critico ogni microsecondo di latenza, ma l'effetto visibile delle performance è trascurabile per i workload tipici.

Progettazione della LIF NFS

A differenza dei protocolli SAN, NFS ha una capacità limitata di definire percorsi multipli ai dati. Le estensioni Parallel NFS (pNFS) a NFSv4 risolvono questo limite, ma poiché le velocità ethernet hanno raggiunto 100GB Mbps e oltre, raramente è utile aggiungere altri percorsi.

Performance e resilienza

Sebbene la misurazione delle performance SAN LIF si debba principalmente calcolare la larghezza di banda totale da tutti i percorsi primari, la determinazione delle performance NFS LIF richiede un'analisi più approfondita dell'esatta configurazione di rete. Ad esempio, è possibile configurare due porte 10Gb come porte fisiche grezze oppure come gruppo di interfacce LACP (link Aggregation Control Protocol). Se sono configurati come gruppo di interfacce, sono disponibili più criteri di bilanciamento del carico che funzionano in modo diverso a seconda che il traffico sia commutato o instradato. Infine, Oracle Direct NFS (DNFS) offre configurazioni di bilanciamento del carico attualmente inesistenti in nessun client NFS del sistema operativo.

A differenza dei protocolli SAN, i file system NFS richiedono resilienza al livello del protocollo. Ad esempio, un LUN è sempre configurato con il multipathing attivato, ovvero sono disponibili più canali ridondanti per il sistema storage, ciascuno dei quali utilizza il protocollo FC. Un file system NFS, invece, dipende dalla disponibilità di un unico canale TCP/IP che può essere protetto solo a livello fisico. Questa disposizione è il motivo per cui esistono opzioni quali il failover della porta e l'aggregazione della porta LACP.

In un ambiente NFS, performance e resilienza sono fornite a livello del protocollo di rete. Di conseguenza, entrambi gli argomenti sono intrecciati e devono essere discussi insieme.

Associare le LIF ai gruppi di porte

Per associare una LIF a un gruppo di porte, associare l'indirizzo IP della LIF a un gruppo di porte fisiche. Il metodo principale per aggregare insieme le porte fisiche è LACP. La capacità di fault tolerance di LACP è abbastanza semplice; ogni porta di un gruppo LACP viene monitorata e rimossa dal gruppo di porte in caso di malfunzionamento. Esistono, tuttavia, molte idee sbagliate sul funzionamento di LACP in relazione alle prestazioni:

- LACP non richiede che la configurazione sullo switch corrisponda all'endpoint. Ad esempio, ONTAP può essere configurato con il bilanciamento del carico basato su IP, mentre uno switch può utilizzare il

bilanciamento del carico basato su MAC.

- Ogni endpoint che utilizza una connessione LACP può scegliere indipendentemente la porta di trasmissione del pacchetto, ma non può scegliere la porta utilizzata per la ricezione. Ciò significa che il traffico da ONTAP a una destinazione specifica è legato a una porta specifica e il traffico di ritorno potrebbe arrivare su un'interfaccia diversa. Ciò non causa tuttavia problemi.
- LACP non distribuisce uniformemente il traffico in ogni momento. In un ambiente di grandi dimensioni con molti client NFS, il risultato è generalmente l'utilizzo di tutte le porte in un'aggregazione LACP. Tuttavia, qualsiasi file system NFS nell'ambiente è limitato alla larghezza di banda di una sola porta, non all'intera aggregazione.
- Sebbene i criteri LACP di robin-robin siano disponibili su ONTAP, questi criteri non indirizzano la connessione da uno switch a un host. Ad esempio, una configurazione con un trunk LACP a quattro porte su un host e un trunk LACP a quattro porte su ONTAP è ancora in grado di leggere un file system utilizzando una sola porta. Sebbene ONTAP sia in grado di trasmettere dati attraverso tutte e quattro le porte, non sono attualmente disponibili tecnologie di switch che inviano dallo switch all'host attraverso tutte e quattro le porte. Ne viene utilizzato uno solo.

L'approccio più comune in ambienti di grandi dimensioni costituiti da molti host di database è quello di creare un aggregato LACP di un numero appropriato di interfacce 10Gb (o più veloce) utilizzando il bilanciamento del carico IP. Questo approccio consente a ONTAP di garantire l'uso uniforme di tutte le porte, purché esistano un numero sufficiente di client. Il bilanciamento del carico si interrompe quando nella configurazione sono presenti meno client, poiché il trunking LACP non ridistribuisce dinamicamente il carico.

Quando viene stabilita una connessione, il traffico in una determinata direzione viene posizionato su una sola porta. Ad esempio, un database che esegue una scansione completa della tabella su un file system NFS collegato tramite un trunk LACP a quattro porte legge i dati tramite una sola scheda di interfaccia di rete (NIC). Se in un tale ambiente sono presenti solo tre server di database, è possibile che tutti e tre stiano leggendo dalla stessa porta, mentre le altre tre porte sono inattive.

Lega le LIF alle porte fisiche

L'associazione di una LIF a una porta fisica dà come risultato un controllo più granulare della configurazione di rete, in quanto un dato indirizzo IP su un sistema ONTAP è associato a una sola porta di rete alla volta. La resilienza viene quindi ottenuta tramite la configurazione di gruppi di failover e policy di failover.

Criteri di failover e gruppi di failover

Il comportamento delle LIF durante un'interruzione di rete è controllato da policy di failover e gruppi di failover. Le opzioni di configurazione sono state modificate con le diverse versioni di ONTAP. Consultare "Documentazione sulla gestione della rete di ONTAP per gruppi e policy di failover" Per informazioni specifiche sulla versione di ONTAP distribuita.

ONTAP 8,3 (e versioni successive) consente la gestione del failover LIF in base ai domini di broadcast. Pertanto, un amministratore può definire tutte le porte che hanno accesso a una data subnet e consentire a ONTAP di selezionare una LIF di failover appropriata. Questo approccio può essere utilizzato da alcuni clienti, ma presenta limitazioni in un ambiente di rete di storage ad alta velocità a causa della mancanza di prevedibilità. Ad esempio, un ambiente può includere sia porte 1Gb GbE per l'accesso di routine al file system sia porte 10Gb GbE per l'i/o del file dati Se nello stesso dominio di broadcast sono presenti entrambi i tipi di porte, il failover LIF può spostare l'i/o del file dati da una porta 10Gb a una porta 1Gb.

In sintesi, prendere in considerazione le seguenti pratiche:

1. Configurare un gruppo di failover come definito dall'utente.
2. Popola il gruppo di failover con le porte sul partner controller di failover dello storage (SFO), in modo che le

LIF seguano gli aggregati durante un failover dello storage. In questo modo si evita di creare traffico indiretto.

3. Utilizza porte di failover con caratteristiche di performance corrispondenti alla LIF originale. Ad esempio, una LIF su una singola porta fisica di 10Gb deve includere un gruppo di failover con una singola porta 10Gb. Un LIF LACP a quattro porte deve eseguire il failover in un altro LIF LACP a quattro porte. Queste porte sono un sottoinsieme delle porte definite nel dominio di broadcast.
4. Impostare la policy di failover solo su partner SFO. Questo assicura che la LIF segua l'aggregato durante il failover.

Ripristino automatico

Impostare `auto-revert` parametro come desiderato. La maggior parte dei clienti preferisce impostare questo parametro su `true` Di ripristinare la porta home della LIF. Tuttavia, in alcuni casi, i clienti hanno impostato questo valore su `false' per poter esaminare un failover imprevisto prima di restituire una LIF alla porta home.

Rapporto LIF-volume

Un equivoco comune consiste nella necessità di una relazione 1:1:1 tra volumi e LIF NFS. Sebbene questa configurazione sia necessaria per spostare un volume ovunque in un cluster senza creare mai traffico di interconnessione aggiuntivo, non si tratta di un requisito categoricamente importante. Occorre considerare il traffico intercluster, ma la semplice presenza di traffico intercluster non crea problemi. Molti dei benchmark pubblicati per ONTAP includono principalmente l'i/o indiretto

Ad esempio, un progetto di database contenente un numero relativamente contenuto di database critici per le performance, che richiedevano solo un totale di 40 volumi, potrebbe giustificare un volume da 1:1 GB per la strategia LIF, una disposizione che richiederebbe 40 indirizzi IP. Quindi, è possibile spostare un qualsiasi volume nel cluster insieme alla LIF associata e il traffico sarebbe sempre diretto, minimizzando ogni origine di latenza anche a livelli di microsecondi.

Ad esempio, è possibile gestire più facilmente un ambiente di grandi dimensioni in hosting con una relazione di 1:1:1 tra clienti e LIF. Con il passare del tempo, potrebbe essere necessario migrare un volume su un nodo diverso, causando traffico indiretto. Tuttavia, l'effetto sulle prestazioni non dovrebbe essere rilevabile a meno che le porte di rete sullo switch di interconnessione non siano saturanti. In caso di problemi, è possibile stabilire una nuova LIF sui nodi aggiuntivi e l'host può essere aggiornato nella successiva finestra di manutenzione per rimuovere il traffico indiretto dalla configurazione.

Configurazione TCP/IP ed ethernet

Molti clienti di Oracle su ONTAP utilizzano ethernet, il protocollo di rete di NFS, iSCSI, NVMe/TCP e specialmente il cloud.

Impostazioni del sistema operativo host

La maggior parte della documentazione del fornitore di applicazioni include impostazioni TCP ed ethernet specifiche per garantire il funzionamento ottimale dell'applicazione. Queste stesse impostazioni sono in genere sufficienti per fornire anche prestazioni ottimali dello storage basato su IP.

Controllo di flusso Ethernet

Questa tecnologia consente a un client di richiedere che un mittente interrompa temporaneamente la trasmissione dei dati. Questa operazione viene solitamente eseguita perché il ricevitore non è in grado di elaborare i dati in ingresso abbastanza rapidamente. Una volta, la richiesta che un mittente cessi la trasmissione era meno disgregativa di avere pacchetti di scarto del destinatario perché i buffer erano pieni.

Questo non è più il caso degli stack TCP utilizzati oggi nei sistemi operativi. Infatti, il controllo di flusso causa più problemi di quanti ne risolva.

Negli ultimi anni sono aumentati i problemi di prestazioni causati dal controllo di flusso Ethernet. Questo perché il controllo di flusso Ethernet opera al livello fisico. Se una configurazione di rete consente a qualsiasi sistema operativo host di inviare una richiesta di controllo di flusso Ethernet a un sistema di storage, il risultato è una pausa in i/o per tutti i client connessi. Poiché un numero crescente di client viene servito da un singolo storage controller, aumenta la probabilità che uno o più client inviino richieste di controllo di flusso. Il problema è stato riscontrato frequentemente presso le sedi dei clienti con un'ampia virtualizzazione del sistema operativo.

Una scheda NIC su un sistema NetApp non dovrebbe ricevere richieste di controllo di flusso. Il metodo utilizzato per ottenere questo risultato varia in base al produttore dello switch di rete. Nella maggior parte dei casi, il controllo di flusso su uno switch Ethernet può essere impostato su `receive desired` oppure `receive on`, il che significa che una richiesta di controllo di flusso non viene inoltrata al controller di memorizzazione. In altri casi, la connessione di rete sul controller di storage potrebbe non consentire la disattivazione del controllo di flusso. In questi casi, i client devono essere configurati in modo da non inviare mai richieste di controllo di flusso, modificando la configurazione NIC sul server host stesso o le porte switch a cui è connesso il server host.

NetApp consiglia assicurarsi che i controller di archiviazione NetApp non ricevano pacchetti di controllo di flusso Ethernet. In genere, è possibile eseguire questa operazione impostando le porte dello switch a cui è collegato il controller, ma alcuni hardware dello switch presentano dei limiti che potrebbero richiedere modifiche sul lato client.

Dimensioni MTU

È stato dimostrato che l'utilizzo dei frame jumbo offre un certo miglioramento delle performance nelle reti 1Gb, riducendo l'overhead della CPU e della rete, ma i benefici non sono solitamente significativi.

NetApp consiglia l'implementazione di frame jumbo quando possibile, sia per ottenere potenziali vantaggi in termini di prestazioni sia per rendere la soluzione a prova di futuro.

L'utilizzo di frame jumbo in una rete 10Gb è quasi obbligatorio. Questo perché la maggior parte delle implementazioni 10Gb raggiungono un limite di pacchetti al secondo senza frame jumbo prima che raggiungano il contrassegno 10Gb. L'utilizzo di frame jumbo migliora l'efficienza dell'elaborazione TCP/IP, poiché consente a sistema operativo, server, schede di rete e sistema di storage di elaborare un numero inferiore di pacchetti, anche se di dimensioni maggiori. Il miglioramento delle prestazioni varia da scheda di rete a scheda di rete, ma è significativo.

Per le implementazioni jumbo-frame, esiste la convinzione comune, ma non corretta, che tutti i dispositivi connessi debbano supportare frame jumbo e che le dimensioni MTU debbano corrispondere end-to-end Al contrario, i due endpoint di rete negoziano la dimensione del frame più elevata reciprocamente accettabile quando si stabilisce una connessione. In un ambiente tipico, uno switch di rete è impostato su una dimensione MTU di 9216, il controller NetApp è impostato su 9000 e i client sono impostati su una combinazione di 9000 e 1514. I client in grado di supportare un valore MTU di 9000 possono utilizzare frame jumbo, mentre i client in grado di supportare solo 1514 possono negoziare un valore inferiore.

I problemi con questa disposizione sono rari in un ambiente completamente commutato. Tuttavia, in un ambiente con routing occorre assicurarsi che nessun router intermedio sia costretto a frammentare frame jumbo.

NetApp consiglia di configurare quanto segue:

- I frame jumbo sono desiderabili ma non necessari con 1Gb Ethernet (GbE).
- I frame jumbo sono necessari per ottenere le massime prestazioni con 10GbE e velocità.

Parametri TCP

Tre impostazioni spesso non sono configurate correttamente: Timestamp TCP, riconoscimento selettivo (SACK) e ridimensionamento finestra TCP. Molti documenti obsoleti su Internet consigliano di disabilitare uno o più di questi parametri per migliorare le prestazioni. Molti anni fa, questa raccomandazione ha avuto un certo merito quando le capacità della CPU erano molto inferiori e, quando possibile, vi era un vantaggio nel ridurre il sovraccarico sull'elaborazione TCP.

Tuttavia, con i sistemi operativi moderni, la disattivazione di una qualsiasi di queste funzioni TCP in genere non comporta alcun vantaggio rilevabile e, allo stesso tempo, può danneggiare le prestazioni. In ambienti di rete virtualizzati, i danni alle prestazioni sono particolarmente probabili, poiché queste funzioni sono necessarie per gestire in modo efficiente la perdita di pacchetti e le modifiche della qualità della rete.

NetApp consiglia di abilitare timestamp TCP, SACCO e ridimensionamento finestra TCP sull'host, e tutti e tre questi parametri dovrebbero essere attivi per impostazione predefinita in qualsiasi sistema operativo corrente.

Configurazione FC SAN

La configurazione di FC SAN per database Oracle riguarda principalmente le seguenti Best practice quotidiane SAN.

Sono incluse misure di pianificazione tipiche, quali la garanzia della presenza di una larghezza di banda sufficiente sulla SAN tra l'host e il sistema di storage, la verifica della presenza di tutti i percorsi SAN tra i dispositivi richiesti, l'utilizzo delle impostazioni della porta FC richieste dal fornitore dello switch FC, evitando conflitti ISL, e utilizzando un adeguato monitoraggio del fabric SAN.

Suddivisione in zone

Una zona FC non deve mai contenere più di un iniziatore. Una tale disposizione potrebbe sembrare funzionare inizialmente, ma la diafonia tra gli iniziatori interferisce eventualmente con le prestazioni e la stabilità.

Le zone MultiTarget sono generalmente considerate sicure, anche se in rare circostanze il comportamento delle porte target FC di fornitori diversi ha causato problemi. Ad esempio, evita di includere nella stessa zona le porte di destinazione di uno storage array NetApp e non NetApp. Inoltre, l'inserimento di un sistema di storage NetApp e di un dispositivo a nastro nella stessa zona è ancora più probabile che causino problemi.

Connessione di rete diretta

Gli amministratori dello storage a volte preferiscono semplificare le loro infrastrutture rimuovendo gli switch di rete dalla configurazione. Questo può essere supportato in alcuni scenari.

ISCSI e NVMe/TCP

Un host che utilizza iSCSI o NVMe/TCP può essere collegato direttamente a un sistema storage e funzionare

normalmente. La ragione è la pedata. Le connessioni dirette a due storage controller differenti offrono due percorsi indipendenti per il flusso di dati. La perdita di percorso, porta o controller non impedisce l'utilizzo dell'altro percorso.

NFS

È possibile utilizzare lo storage NFS con connessione diretta, ma con una limitazione significativa: Il failover non funzionerà senza una significativa attività di scripting, che sarà responsabilità del cliente.

Il motivo per cui il failover senza interruzioni è complicato con lo storage NFS connesso direttamente è il routing che si verifica sul sistema operativo locale. Ad esempio, si supponga che un host abbia un indirizzo IP 192.168.1.1/24 e che sia collegato direttamente a un controller ONTAP con un indirizzo IP 192.168.1.50/24. Durante il failover, l'indirizzo 192.168.1.50 può eseguire il failover sull'altro controller e sarà disponibile per l'host, ma in che modo l'host rileva la sua presenza? L'indirizzo 192.168.1.1 originale esiste ancora sulla scheda di rete host che non si connette più a un sistema operativo. Il traffico destinato a 192.168.1.50 continuerebbe ad essere inviato a una porta di rete inutilizzabile.

La seconda scheda NIC del sistema operativo potrebbe essere configurata come 19 2.168.1.2 e sarebbe in grado di comunicare con l'indirizzo 192.168.1.50 non riuscito, ma le tabelle di routing locali avrebbero un valore predefinito di utilizzo di un solo indirizzo **e di un solo indirizzo** per comunicare con la subnet 192.168.1.0/24. Un amministratore di sistema potrebbe creare un framework di script che rilevi una connessione di rete non riuscita e alteri le tabelle di routing locali o che porti le interfacce verso l'alto e verso il basso. La procedura esatta dipende dal sistema operativo in uso.

In pratica, i clienti NetApp dispongono di NFS con connessione diretta, ma in genere solo per i workload in cui le pause io durante i failover sono accettabili. Quando si utilizzano i supporti rigidi, non devono verificarsi errori di i/o durante tali pause. L'io dovrebbe bloccarsi finché i servizi non vengono ripristinati, mediante failback o intervento manuale, per spostare gli indirizzi IP tra le schede NIC dell'host.

Connessione diretta FC

Non è possibile connettere direttamente un host a un sistema storage ONTAP utilizzando il protocollo FC. Il motivo è l'uso di NPIV. Il WWN che identifica una porta FC ONTAP per la rete FC utilizza un tipo di virtualizzazione chiamato NPIV. Qualsiasi dispositivo collegato a un sistema ONTAP deve essere in grado di riconoscere un WWN NPIV. Attualmente non vi sono fornitori di HBA che offrono un HBA che può essere installato in un host in grado di supportare un target NPIV.

Configurazione dello storage

SAN FC

Allineamento delle LUN

L'allineamento delle LUN si riferisce all'ottimizzazione dell'i/o in relazione al layout del file system sottostante.

Direttiva non risolta in <stdin> - include::lun-alignment.adoc[]

Vedere anche la discussione sull'allineamento dei blocchi di compressione nella sezione "Efficienza". Qualsiasi layout allineato ai limiti del blocco di compressione 8KB è allineato ai limiti 4KB.

Avvertenze di disallineamento

Direttiva non risolta in <stdin> - include::database-misalignment-warnings.adoc[]

L'allineamento negli ambienti Solaris è più complicato. Fare riferimento a. "Configurazione host SAN ONTAP" per ulteriori informazioni.

Attenzione
Negli ambienti Solaris x86, prestare ulteriore attenzione al corretto allineamento poiché la maggior parte delle configurazioni prevede diversi livelli di partizioni. Le sezioni di partizione di Solaris x86 si trovano solitamente in cima a una tabella di partizioni del record di avvio master standard.

Dimensionamento e numero di LUN

La scelta delle dimensioni ottimali e del numero di LUN da utilizzare è un elemento critico per ottenere performance e gestibilità ottimali dei database Oracle.

Un LUN è un oggetto virtualizzato in ONTAP presente in tutti i dischi dell'aggregato di hosting. Di conseguenza, le performance della LUN non sono influenzate dalle sue dimensioni, perché la LUN sfrutta al massimo il potenziale in termini di performance dell'aggregato, indipendentemente dalle dimensioni scelte.

Per comodità, i clienti potrebbero desiderare di utilizzare un LUN di particolari dimensioni. Ad esempio, se un database è costruito su un gruppo di dischi LVM o Oracle ASM composto da due LUN da 1TB GB ciascuno, tale gruppo di dischi deve essere aumentato in incrementi di 1TB TB. Potrebbe essere preferibile costruire il gruppo di dischi da otto LUN da 500GB ciascuno in modo che il gruppo di dischi possa essere aumentato con incrementi più piccoli.

La pratica di stabilire una dimensione LUN standard universale è scoraggiata perché ciò può complicare la gestibilità. Ad esempio, è possibile che una dimensione LUN standard di 100GB TB sia ottimale quando un database o un datastore è compreso nell'intervallo da 1TB a 2TB TB, ma un database o un datastore di 20TB GB richiederebbe 200 LUN. Ciò significa che i tempi di riavvio del server sono più lunghi, che vi sono più oggetti da gestire nelle varie interfacce utente e che prodotti come SnapCenter devono eseguire la ricerca su molti oggetti. Utilizzando un numero inferiore di LUN di dimensioni maggiori è possibile evitare questi problemi.

- Il numero di LUN è più importante delle dimensioni delle LUN.
- Le dimensioni dei LUN sono principalmente controllate dai requisiti di numero di LUN.
- Evitare di creare più LUN del necessario.

Numero di LUN

A differenza delle dimensioni delle LUN, il numero di LUN influisce sulle performance. Spesso le prestazioni delle applicazioni dipendono dalla capacità di eseguire i/o paralleli attraverso il livello SCSI. Di conseguenza, due LUN offrono performance migliori rispetto a una singola LUN. Utilizzare un LVM come Veritas VxVM, Linux LVM2 o Oracle ASM è il metodo più semplice per aumentare il parallelismo.

I clienti di NetApp hanno in genere ottenuto il minimo beneficio dall'aumento del numero di LUN oltre i sedici, sebbene i test degli ambienti con dischi a stato solido al 100% con i/o casuali molto intensi abbiano dimostrato un ulteriore miglioramento fino a 64 LUN.

NetApp consiglia quanto segue:

In generale, da quattro a sedici LUN sono sufficienti per supportare le esigenze di i/o di qualsiasi carico di lavoro del database. Meno di quattro LUN potrebbero creare limiti di performance a causa delle limitazioni nelle implementazioni SCSI host.

Posizionamento delle LUN

Il posizionamento ottimale delle LUN del database all'interno dei volumi ONTAP dipende principalmente dalle diverse funzionalità di ONTAP.

Volumi

Un punto comune di confusione tra i clienti che non conoscono ONTAP è l'utilizzo di FlexVol, comunemente denominati semplicemente "volumi".

Un volume non è un LUN. Questi termini vengono utilizzati in maniera anonima con molti prodotti di altri vendor, inclusi i cloud provider. ONTAP Volumes sono semplicemente container di gestione. Non forniscono dati da soli, né occupano spazio. Sono container per file o LUN e esistono per migliorare e semplificare la gestibilità, in particolare su larga scala.

Volumi e LUN

I LUN correlati sono normalmente collocati in una stessa posizione in un singolo volume. Ad esempio, un database che richiede 10 LUN solitamente conterrà tutte le 10 LUN dello stesso volume.

- L'utilizzo di un rapporto di 1:1:1 tra LUN e volumi, vale a dire un LUN per volume, non è * una Best practice formale.
- I volumi dovrebbero invece essere visti come container per i carichi di lavoro o i set di dati. È possibile che sia presente un singolo LUN per volume o che ve ne siano molti. La risposta giusta dipende dai requisiti di gestibilità.
- La dispersione dei LUN in un numero non necessario di volumi può causare overhead e problemi di pianificazione aggiuntivi per operazioni quali operazioni di snapshot, un numero eccessivo di oggetti visualizzati nell'interfaccia utente e il raggiungimento dei limiti di volume della piattaforma prima del raggiungimento del limite LUN.

Volumi, LUN e snapshot

I criteri e le pianificazioni degli Snapshot vengono posizionati sul volume, non sul LUN. Un set di dati composto da 10 LUN richiederebbe una singola policy di snapshot quando le LUN sono collocate contemporaneamente nello stesso volume.

Inoltre, la co-localizzazione di tutti i LUN correlati per un dato dataset in un singolo volume consente di eseguire operazioni di snapshot atomiche. Ad esempio, un database di 10 LUN o un ambiente applicativo basato su VMware composto da 10 diversi sistemi operativi possono essere protetti come un singolo oggetto coerente se le LUN sottostanti vengono tutte collocate in un singolo volume. Se vengono posizionati su volumi diversi, gli snapshot possono essere o meno sincronizzati al 100%, anche se pianificati allo stesso tempo.

In alcuni casi, potrebbe essere necessario suddividere una serie di LUN correlata in due volumi diversi a causa dei requisiti di recovery. Ad esempio, un database potrebbe avere quattro LUN per i file di dati e due LUN per i log. In questo caso, un volume di file dati con 4 LUN e un volume di registro con 2 LUN potrebbe essere l'opzione migliore. Il motivo è la possibilità di recupero indipendente. Ad esempio, è possibile ripristinare in

maniera selettiva il volume di file dati a uno stato precedente, vale a dire che le quattro LUN vengono riportate allo stato della snapshot, senza influire sul volume di log con i dati critici.

Volumi, LUN e SnapMirror

Le policy e le operazioni di SnapMirror, come le operazioni di Snapshot, vengono eseguite sul volume, non sul LUN.

La co-localizzazione dei LUN correlati in un singolo volume consente di creare una singola relazione di SnapMirror e di aggiornare tutti i dati contenuti con un singolo update. Come per gli snapshot, l'aggiornamento sarà anche un'operazione atomica. La destinazione SnapMirror avrà una replica point-in-time singola delle LUN di origine. Se le LUN sono state distribuite su più volumi, le repliche possono essere o meno coerenti l'una con l'altra.

Volumi, LUN e QoS

Mentre la qualità del servizio può essere applicata in modo selettivo alle singole LUN, in genere è più semplice impostarla a livello di volume. Ad esempio, tutte le LUN utilizzate dai guest di un determinato server ESX possono essere collocate su un singolo volume e successivamente può essere applicata una policy di QoS adattiva di ONTAP. In questo modo si ottiene un limite di IOPS per TB autoscalabile valido per tutte le LUN.

Analogamente, se un database richiedeva 100K IOPS e occupava 10 LUN, sarebbe più semplice impostare un singolo limite di 100K IOPS su un singolo volume piuttosto che impostare 10 limiti individuali di 10K IOPS, uno per ogni LUN.

Layout a più volumi

Vi sono alcuni casi in cui la distribuzione delle LUN su più volumi può essere vantaggiosa. Il motivo principale è lo striping dei controller. Ad esempio, un sistema storage ha potrebbe ospitare un singolo database in cui è richiesto il potenziale completo di elaborazione e caching di ogni controller. In questo caso, una progettazione tipica sarebbe quella di collocare metà dei LUN in un singolo volume sul controller 1, e l'altra metà dei LUN in un singolo volume sul controller 2.

Analogamente, lo striping dei controller potrebbe essere utilizzato per il bilanciamento del carico. Un sistema ha che ospitava 100 database da 10 LUN ciascuno potrebbe essere progettato dove ogni database riceve un volume da 5 LUN su ciascuno dei due controller. Il risultato è garantito il caricamento simmetrico di ogni controller quando vengono forniti database aggiuntivi.

Tuttavia, nessuno di questi esempi riguarda un rapporto volume-LUN di 1:1:1. L'obiettivo resta quello di ottimizzare la gestibilità mediante la co-localizzazione dei LUN correlati in volumi.

Un esempio se un rapporto da 1:1 LUN a volume è sensato è la containerizzazione, laddove ogni LUN potrebbe rappresentare davvero un singolo carico di lavoro e deve essere gestita singolarmente. In questi casi, un rapporto 1:1:1 può essere ottimale.

Ridimensionamento LUN e ridimensionamento LVM

Quando un file system basato su SAN ha raggiunto il limite di capacità, sono disponibili due opzioni per aumentare lo spazio disponibile:

- Aumentare la dimensione dei LUN
- Aggiungere un LUN a un gruppo di volumi esistente e aumentare il volume logico contenuto

Sebbene il ridimensionamento delle LUN sia un'opzione per aumentare la capacità, in genere è preferibile

utilizzare un LVM, incluso Oracle ASM. Uno dei motivi principali per cui esistono le LVM è evitare la necessità di ridimensionare le LUN. Con un LVM, più LUN sono unite in un pool virtuale di storage. I volumi logici scavati da questo pool sono gestiti da LVM e possono essere facilmente ridimensionati. Un ulteriore vantaggio è l'eliminazione degli hotspot su una determinata unità distribuendo un determinato volume logico su tutte le LUN disponibili. Di solito, la migrazione trasparente può essere eseguita utilizzando il volume manager per spostare le estensioni sottostanti di un volume logico su nuovi LUN.

Striping LVM

Lo striping LVM si riferisce alla distribuzione dei dati su più LUN. Il risultato è un significativo miglioramento delle performance per molti database.

Prima dell'era dei dischi flash, era stato utilizzato lo striping per superare i limiti di performance dei dischi rotanti. Ad esempio, se un sistema operativo deve eseguire un'operazione di lettura a 1MB bit, la lettura di 1MB GB di dati da un'unica unità richiederebbe un'ampia ricerca e lettura della testina dell'unità poiché il sistema 1MB viene trasferito lentamente. Se quei 1MB TB di dati sono stati suddivisi in 8 LUN, il sistema operativo potrebbe emettere otto operazioni di lettura 128K in parallelo, riducendo il tempo necessario per completare il trasferimento da 1MB GB.

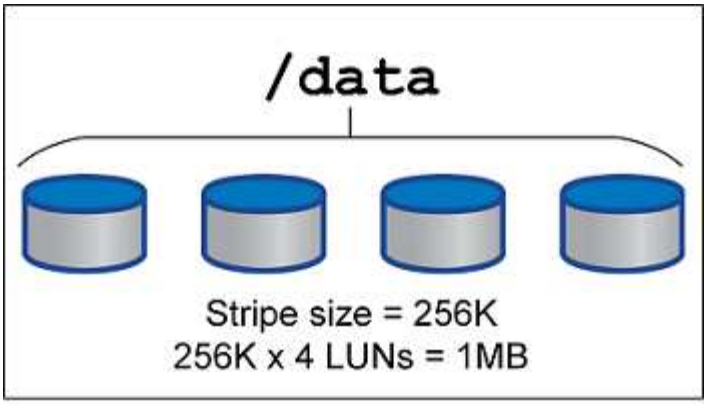
Lo striping con dischi rotanti era più difficile perché lo schema di i/o doveva essere noto in anticipo. Se lo striping non è stato regolato correttamente per i modelli i/o reali, le configurazioni con striping potrebbero danneggiare le prestazioni. Con i database Oracle, e in particolare con le configurazioni all-flash, lo striping è molto più semplice da configurare ed è stato dimostrato che le performance risultano notevolmente migliorate.

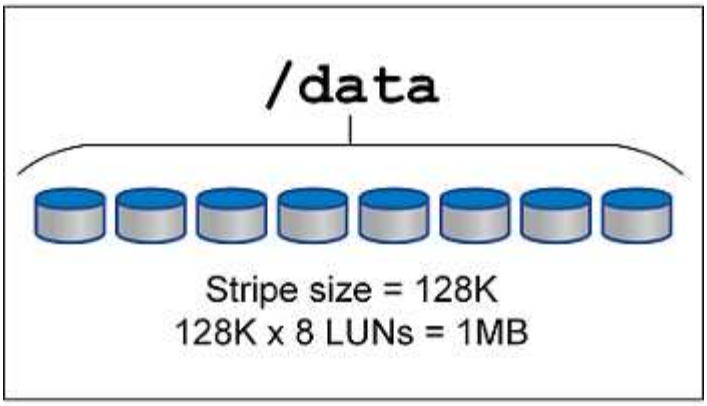
Per impostazione predefinita, i gestori di volume logici, come lo stripe di Oracle ASM, ma il sistema operativo LVM nativo non lo fanno. Alcune di esse collegano più LUN insieme come un dispositivo concatenato, il che comporta file di dati che esistono su un solo dispositivo LUN. Ciò causa punti caldi. Le altre implementazioni LVM sono impostate per impostazione predefinita su estensioni distribuite. Questo è simile allo striping, ma è più grossolano. I LUN nel gruppo di volumi vengono suddivisi in porzioni di grandi dimensioni, chiamate estensioni e generalmente misurati in molti megabyte, e i volumi logici vengono quindi distribuiti tra tali estensioni. Il risultato è un i/o casuale per un file dovrebbe essere ben distribuito tra i LUN, ma le operazioni i/o sequenziali non sono così efficienti come potrebbero essere.

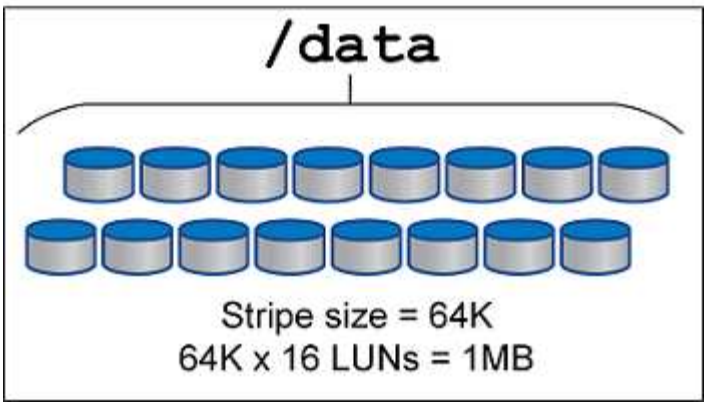
L'i/o delle applicazioni che richiedono elevate performance è quasi sempre (a) in unità delle dimensioni dei blocchi di base o (b) un megabyte.

L'obiettivo principale di una configurazione con striping è quello di garantire che l'i/o a file singolo possa essere eseguito come una singola unità, mentre l'i/o a blocchi multipli, di dimensioni pari a 1MB GB, può essere parallelizzato in modo uniforme tra tutti i LUN del volume con striping. Ciò significa che la dimensione dello stripe non deve essere inferiore alla dimensione del blocco del database e che la dimensione dello stripe moltiplicata per il numero di LUN deve essere 1MB.

La figura seguente mostra tre possibili opzioni per la regolazione delle dimensioni dello stripe e della larghezza. Il numero di LUN viene selezionato per soddisfare i requisiti di prestazioni come descritto sopra, ma in tutti i casi i dati totali all'interno di uno stripe singolo sono 1MB.







NFS

Panoramica

NetApp offre storage NFS Enterprise da oltre 30 anni e il suo utilizzo cresce insieme alla spinta verso infrastrutture basate sul cloud grazie alla sua semplicità.

Il protocollo NFS include diverse versioni con diversi requisiti. Per una descrizione completa della configurazione NFS con ONTAP, vedere "Best practice NFS su ONTAP TR-4067". Le sezioni seguenti descrivono alcuni dei requisiti più critici e gli errori comuni degli utenti.

Versioni di NFS

Il client NFS del sistema operativo deve essere supportato da NetApp.

- NFSv3 è supportato con sistemi operativi che seguono lo standard NFSv3.

- NFSv3 è supportato con il client Oracle DNFS.
- NFSv4 è supportato con tutti i sistemi operativi che seguono lo standard NFSv4.
- I sistemi NFSv4,1 e NFSv4,2 richiedono supporto specifico per il sistema operativo. Consultare "NetApp IMT" Per i sistemi operativi supportati.
- Il supporto di Oracle DNFS per NFSv4,1 richiede Oracle 12.2.0.2 o versione successiva.

> Il "Matrice di supporto di NetApp" Per NFSv3 e NFSv4 non sono inclusi sistemi operativi specifici. Tutti i sistemi operativi che rispettano la RFC sono generalmente supportati. Quando si cerca il supporto NFSv3 o NFSv4 nel IMT online, non selezionare un sistema operativo specifico perché non verranno visualizzate corrispondenze. Tutti i sistemi operativi sono implicitamente supportati dalla policy generale.

Tabelle degli slot TCP per Linux NFSv3

Le tabelle degli slot TCP sono l'equivalente di NFSv3 della profondità della coda degli HBA (host Bus Adapter). Queste tabelle controllano il numero di operazioni NFS che possono essere in sospeso in qualsiasi momento. Il valore predefinito è di solito 16, che è troppo basso per ottenere prestazioni ottimali. Il problema opposto si verifica sui kernel Linux più recenti, che possono aumentare automaticamente il limite della tabella degli slot TCP a un livello che satura il server NFS con le richieste.

Per prestazioni ottimali e per evitare problemi di prestazioni, regolare i parametri del kernel che controllano le tabelle degli slot TCP.

Eseguire `sysctl -a | grep tcp.*.slot_table` e osservare i seguenti parametri:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tutti i sistemi Linux dovrebbero includere `sunrpc.tcp_slot_table_entries`, ma solo alcuni includono `sunrpc.tcp_max_slot_table_entries`. Entrambi devono essere impostati su 128.

> La mancata impostazione di questi parametri può avere effetti significativi sulle prestazioni. In alcuni casi, le prestazioni sono limitate poiché il sistema operativo linux non fornisce i/o sufficienti In altri casi, le latenze i/o aumentano quando il sistema operativo linux tenta di emettere più i/o di quanto possa essere gestito.

ADR e NFS

Alcuni clienti hanno segnalato problemi di prestazioni derivanti da una quantità eccessiva di i/o sui dati in `ADR` posizione. Il problema generalmente non si verifica finché non si sono accumulati molti dati sulle prestazioni. Il motivo dell'eccessivo i/o è sconosciuto, ma questo problema sembra essere dovuto ai processi Oracle che eseguono ripetutamente la scansione della directory di destinazione per rilevare eventuali modifiche.

Smontaggio del `noac` e/o. `actimeo=0` Le opzioni di montaggio consentono il caching del sistema operativo host e riducono i livelli di i/o dello storage.

NetApp consiglia di non piazzare `ADR` dati su un file system con `noac` oppure `actimeo=0` perché è probabile che si verifichino problemi di prestazioni. Separare `ADR` dati in un punto di montaggio diverso, se necessario.

nfs-rootonly e mount-rootonly

ONTAP include un'opzione NFS denominata `nfs-rootonly` Che controlla se il server accetta connessioni di traffico NFS da porte elevate. Come misura di sicurezza, solo l'utente root è autorizzato ad aprire connessioni TCP/IP utilizzando una porta di origine inferiore a 1024, poiché tali porte sono normalmente riservate all'uso del sistema operativo, non ai processi utente. Questa restrizione aiuta a garantire che il traffico NFS provenga da un client NFS del sistema operativo effettivo e non da un processo dannoso che emula un client NFS. Il client Oracle DNFS è un driver userspace, ma il processo viene eseguito come root, quindi in genere non è necessario modificare il valore di `nfs-rootonly`. I collegamenti sono costituiti da porte basse.

Il `mount-rootonly` L'opzione è valida solo per NFSv3. Controlla se la chiamata di MONTAGGIO RPC può essere accettata dalle porte superiori a 1024. Quando si utilizza DNFS, il client viene nuovamente eseguito come root, in modo da poter aprire le porte al di sotto di 1024. Questo parametro non ha alcun effetto.

I processi che aprono connessioni con DNFS su NFS versione 4,0 e successive non vengono eseguiti come root e quindi richiedono porte su 1024. Il `nfs-rootonly` Il parametro deve essere impostato su disabilitato affinché DNFS completi la connessione.

Se `nfs-rootonly` È attivato, il risultato è un blocco durante la fase di mount che apre le connessioni DNFS. L'output di sqlplus è simile a questo:

```
SQL>startup
ORACLE instance started.
Total System Global Area 4294963272 bytes
Fixed Size                  8904776 bytes
Variable Size             822083584 bytes
Database Buffers         3456106496 bytes
Redo Buffers                7868416 bytes
```

Il parametro può essere modificato come segue:

```
Cluster01::> nfs server modify -nfs-rootonly disabled
```

In situazioni rare, potrebbe essere necessario modificare sia nfs-rootonly che mount-rootonly in disabled. Se un server gestisce un numero estremamente elevato di connessioni TCP, è possibile che non siano disponibili porte al di sotto di 1024 e che il sistema operativo sia costretto a utilizzare porte più elevate. Questi due parametri ONTAP devono essere modificati per consentire il completamento della connessione.

Policy di esportazione NFS: Superuser e setuid

Se i file binari Oracle si trovano in una condivisione NFS, la policy di esportazione deve includere autorizzazioni superser e setuid.

Le esportazioni NFS condivise utilizzate per servizi file generici come le home directory dell'utente spesso

fanno uso dell'utente root. Ciò significa che una richiesta da parte dell'utente root su un host che ha montato un filesystem viene rimappata come un altro utente con privilegi inferiori. In questo modo è possibile proteggere i dati impedendo a un utente root di un determinato server di accedere ai dati del server condiviso. Il bit setuid può anche essere un rischio per la protezione in un ambiente condiviso. Il bit setuid consente di eseguire un processo come un utente diverso da quello che richiama il comando. Ad esempio, uno script della shell di proprietà di root con il bit setuid viene eseguito come root. Se lo script della shell potrebbe essere modificato da altri utenti, qualsiasi utente non root potrebbe eseguire un comando come root aggiornando lo script.

I file binari di Oracle includono file di proprietà di root e utilizzano il bit setuid. Se i file binari Oracle sono installati su una condivisione NFS, la policy di esportazione deve includere le autorizzazioni appropriate per superutente e setuid. Nell'esempio seguente, la regola include entrambi `allow-suid` e permessi `superuser` Accesso root per client NFS utilizzando l'autenticazione di sistema.

```
Cluster01::> export-policy rule show -vserver vserver1 -policyname orabin
-fields allow-suid,superuser
vserver   policyname ruleindex superuser allow-suid
--------- ---------- --------- --------- ----------
vserver1  orabin     1         sys       true
```

Configurazione NFSv4/4,1

Per la maggior parte delle applicazioni, la differenza tra NFSv3 e NFSv4 è minima. L'i/o delle applicazioni è di solito un i/o molto semplice e non trae alcun vantaggio significativo da alcune delle funzionalità avanzate disponibili in NFSv4. Le versioni più elevate di NFS non devono essere considerate come un "aggiornamento" dal punto di vista dello storage dei database, ma come versioni di NFS che includono funzionalità aggiuntive. Ad esempio, se è richiesta la protezione end-to-end della modalità di privacy Kerberos (krb5p), è necessario NFSv4.

NetApp consiglia di utilizzare NFSv4,1 se sono necessarie funzionalità NFSv4. Sono stati apportati alcuni miglioramenti funzionali al protocollo NFSv4 di NFSv4,1 che migliorano la resilienza in alcuni casi edge.

Il passaggio a NFSv4 è più complicato che cambiare semplicemente le opzioni di montaggio da vers=3 a vers=4,1. Una spiegazione più completa della configurazione NFSv4 con ONTAP, incluse le istruzioni sulla configurazione del sistema operativo, vedere "Best practice TR-4067 NFS su ONTAP". Le seguenti sezioni di questo TR spiegano alcuni dei requisiti di base per l'utilizzo di NFSv4.

Dominio NFSv4

Una spiegazione completa della configurazione NFSv4/4,1 esula dall'ambito di questo documento, ma un problema comunemente riscontrato è una mancata corrispondenza nella mappatura del dominio. Dal punto di vista di sysadmin, i file system NFS sembrano comportarsi normalmente, ma le applicazioni segnalano errori relativi ai permessi e/o setuid su determinati file. In alcuni casi, gli amministratori hanno concluso erroneamente che le autorizzazioni dei binari dell'applicazione sono state danneggiate e hanno eseguito comandi chown o chmod quando il problema effettivo era il nome di dominio.

Il nome di dominio NFSv4 viene impostato sulla SVM ONTAP:

```
Cluster01::> nfs server show -fields v4-id-domain
vserver   v4-id-domain
--------- ------------
vserver1  my.lab
```

Il nome di dominio NFSv4 sull'host è impostato in `/etc/idmap.cfg`

```
[root@host1 etc]# head /etc/idmapd.conf
[General]
#Verbosity = 0
# The following should be set to the local NFSv4 domain name
# The default is the host's DNS domain name.
Domain = my.lab
```

I nomi di dominio devono corrispondere. In caso contrario, vengono visualizzati errori di mappatura simili a quelli riportati di seguito nella `/var/log/messages`:

```
Apr 12 11:43:08 host1 nfsidmap[16298]: nss_getpwnam: name 'root@my.lab'
does not map into domain 'default.com'
```

I file binari delle applicazioni, come i file binari dei database Oracle, includono i file di proprietà di root con il bit setuid, il che significa che una mancata corrispondenza nei nomi di dominio NFSv4 causa errori nell'avvio di Oracle e un avviso sulla proprietà o sulle autorizzazioni di un file chiamato `oradism`, che si trova nella `$ORACLE_HOME/bin` directory. Dovrebbe comparire come segue:

```
[root@host1 etc]# ls -l /orabin/product/19.3.0.0/dbhome_1/bin/oradism
-rwsr-x--- 1 root oinstall 147848 Apr 17  2019
/orabin/product/19.3.0.0/dbhome_1/bin/oradism
```

Se questo file viene visualizzato con proprietà di nessuno, potrebbe esserci un problema di mappatura del dominio NFSv4.

```
[root@host1 bin]# ls -l oradism
-rwsr-x--- 1 nobody oinstall 147848 Apr 17  2019 oradism
```

Per risolvere questo problema, controllare `/etc/idmap.cfg` Eseguire il file in base all'impostazione del dominio id v4 in ONTAP e assicurarsi che siano coerenti. In caso contrario, apportare le modifiche necessarie, eseguire `nfsidmap -c`, e attendere un momento per la propagazione delle modifiche. La proprietà del file dovrebbe quindi essere riconosciuta correttamente come root. Se un utente aveva tentato di eseguire `chown root` Su questo file prima che la configurazione dei domini NFS sia stata corretta, potrebbe essere necessario eseguire `chown root` di nuovo.

Oracle Direct NFS (DNFS)

I database Oracle possono utilizzare NFS in due modi.

In primo luogo, può usare un filesystem montato usando il client NFS nativo che fa parte del sistema operativo. Questo è talvolta chiamato kernel NFS, o kNFS. Il filesystem NFS è montato e usato dal database Oracle esattamente come qualsiasi altra applicazione userebbe un filesystem NFS.

Il secondo metodo è Oracle Direct NFS (DNFS). Si tratta di un'implementazione dello standard NFS nel software di database Oracle. Senza modificare le modalità di configurazione o gestione dei database Oracle da parte del DBA. Purché le impostazioni del sistema storage siano corrette, l'utilizzo del DNFS deve essere trasparente per il team DBA e gli utenti finali.

Un database con la funzione DNFS attivata ha ancora i consueti filesystem NFS montati. Una volta aperto il database, il database Oracle apre una serie di sessioni TCP/IP ed esegue direttamente le operazioni NFS.

NFS diretto

Il valore principale di Oracle Direct NFS è quello di ignorare il client NFS host ed eseguire operazioni di file NFS direttamente su un server NFS. Per abilitarla è sufficiente modificare la libreria Oracle Disk Manager (ODM). Le istruzioni per questo processo sono fornite nella documentazione di Oracle.

L'utilizzo di DNFS porta a un significativo miglioramento delle performance di i/o e riduce il carico sull'host e sul sistema storage poiché l'i/o viene eseguito nel modo più efficiente possibile.

Inoltre, Oracle DNFS include un'opzione **opzionale** per il multipathing e la fault tolerance dell'interfaccia di rete. Ad esempio, è possibile associare due interfacce 10Gb in modo da ottenere una larghezza di banda di 20Gb Gbps. Un errore di un'interfaccia provoca il tentativo di i/o sull'altra interfaccia. Il funzionamento complessivo è molto simile al multipathing FC. Il multipathing era comune anni fa quando ethernet a 1Gb GB rappresentava lo standard più comune. Una NIC 10Gb è sufficiente per la maggior parte dei carichi di lavoro Oracle, ma se ne richiede di più, è possibile collegare 10Gb NIC.

Quando si utilizza DNFS, è fondamentale che tutte le patch descritte in Oracle Doc 1495104,1 siano installate. Se non è possibile installare una patch, è necessario valutare l'ambiente per assicurarsi che i bug descritti in quel documento non causino problemi. In alcuni casi, l'impossibilità di installare le patch necessarie impedisce l'utilizzo di DNFS.

Non utilizzare DNFS con alcun tipo di risoluzione dei nomi round-robin, compresi DNS, DDNS, NIS o qualsiasi altro metodo. Ciò include la funzione di bilanciamento del carico DNS disponibile in ONTAP. Quando un database Oracle che utilizza DNFS risolve un nome host in un indirizzo IP, non deve cambiare nelle ricerche successive. Ciò può causare arresti anomali del database Oracle e possibili danni ai dati.

Attivazione di DNFS

Oracle DNFS può funzionare con NFSv3 senza necessità di configurazione oltre all'abilitazione della libreria DNFS (vedere la documentazione di Oracle per il comando specifico richiesto) ma se DNFS non è in grado di stabilire la connettività, può tornare automaticamente al client NFS del kernel. In questo caso, le prestazioni possono essere gravemente compromesse.

Se si desidera utilizzare la multiplazione DNFS su più interfacce, con NFSv4.X, o utilizzare la cifratura, è necessario configurare un file oranfstab. La sintassi è estremamente rigorosa. Piccoli errori nel file possono causare la sospensione dell'avvio o il bypass del file oranfstab.

Al momento della stesura del presente documento, il multipathing DNFS non funziona con NFSv4,1 con le versioni più recenti di Oracle Database. Un file oranfstab che specifica NFSv4,1 come protocollo può utilizzare

solo un'istruzione PATH singola per una determinata esportazione. Il motivo è che ONTAP non supporta il trunking clientID. Le patch dei database Oracle per risolvere questo limite potrebbero essere disponibili in futuro.

L'unico modo per essere certi che DNFS funzioni come previsto è eseguire una query sulle tabelle v$dnfs.

Di seguito è riportato un file oranfstab di esempio che si trova in /etc. Questa è una delle posizioni multiple un file oranfstab può essere posizionato.

```
[root@jfs11 trace]# cat /etc/oranfstab
server: NFSv3test
path: jfs_svmdr-nfs1
path: jfs_svmdr-nfs2
export: /dbf mount: /oradata
export: /logs mount: /logs
nfs_version: NFSv3
```

Il primo passo consiste nel verificare che DNFS sia operativo per i filesystem specificati:

```
SQL> select dirname,nfsversion from v$dnfs_servers;

DIRNAME
-----------------------------------
NFSVERSION
----------------
/logs
NFSv3.0

/dbf
NFSv3.0
```

Questo output indica che DNFS è in uso con questi due filesystem, ma **non** significa che oranfstab è operativo. Se fosse presente un errore, DNFS avrebbe scoperto automaticamente i filesystem NFS dell'host e si potrebbe comunque vedere lo stesso output da questo comando.

Il multipathing può essere controllato come segue:

```
SQL> select svrname,path,ch_id from v$dnfs_channels;

SVRNAME
-----------------------------------
PATH
-----------------------------------
     CH_ID
----------
NFSv3test
```

```
jfs_svmdr-nfs1
         0

NFSv3test
jfs_svmdr-nfs2
         1

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------

NFSv3test
jfs_svmdr-nfs1
         0

NFSv3test
jfs_svmdr-nfs2

[output truncated]

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------
NFSv3test
jfs_svmdr-nfs2
         1

NFSv3test
jfs_svmdr-nfs1
         0

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------

NFSv3test
jfs_svmdr-nfs2
```

```
          1


66 rows selected.
```

Di seguito sono riportate le connessioni utilizzate da DNFS. Per ogni voce SVRNAME sono visibili due percorsi e canali. Ciò significa che il multipathing funziona, il che significa che il file oranfstab è stato riconosciuto ed elaborato.

Accesso diretto NFS e file system host

L'utilizzo di DNFS può causare occasionalmente problemi per le applicazioni o le attività degli utenti che si basano sui file system visibili montati sull'host perché il client DNFS accede al file system fuori banda dal sistema operativo host. Il client DNFS può creare, eliminare e modificare i file senza conoscere il sistema operativo.

Quando vengono utilizzate le opzioni di montaggio per i database a istanza singola, consentono la memorizzazione nella cache degli attributi di file e directory, il che significa anche che il contenuto di una directory viene memorizzato nella cache. Pertanto, DNFS può creare un file, e c'è un breve ritardo prima che il sistema operativo rilegga il contenuto della directory e il file diventi visibile all'utente. Questo non è generalmente un problema, ma, in rare occasioni, utility come SAP BR*Tools potrebbero avere problemi. In questo caso, risolvere il problema modificando le opzioni di montaggio in modo da utilizzare le raccomandazioni per Oracle RAC. Questa modifica comporta la disabilitazione di tutta la cache dell'host.

Modificare le opzioni di montaggio solo quando (a) viene utilizzato DNFS e (b) un problema deriva da un ritardo nella visibilità dei file. Se DNFS non è in uso, l'utilizzo delle opzioni di montaggio di Oracle RAC su un database a singola istanza comporta un peggioramento delle prestazioni.

> Consultare la nota relativa a `nosharecache` in "Opzioni di montaggio NFS Linux" per un problema DNFS specifico di Linux che può produrre risultati insoliti.

Leasing e blocchi di NFS

NFSv3 è stateless. Ciò significa che il server NFS (ONTAP) non tiene traccia di quali file system sono montati, da chi, o quali blocchi sono realmente presenti.

ONTAP dispone di alcune funzionalità che registreranno i tentativi di mount, quindi si ha un'idea di quali client possono accedere ai dati e potrebbero essere presenti blocchi di avvisi, ma non è garantito che le informazioni siano complete al 100%. Non può essere completo, perché il tracciamento dello stato del client NFS non fa parte dello standard NFSv3.

NFSv4 statefulness

Al contrario, NFSv4 è stateful. Il server NFSv4 tiene traccia di quali client utilizzano i file system, quali file esistono, quali file e/o aree di file sono bloccati, ecc. Ciò significa che è necessaria una comunicazione regolare tra un server NFSv4 per mantenere aggiornati i dati di stato.

Gli stati più importanti gestiti dal server NFS sono NFSv4 Locks e NFSv4 Leasing, e sono molto interconnessi. Dovete capire come ognuno funziona da se stesso e come si relazionano l'uno con l'altro.

NFSv4 serrature

Con NFSv3, i blocchi sono indicativi. Un client NFS può comunque modificare o eliminare un file "bloccato". Un blocco NFSv3 non scade da solo, deve essere rimosso. Questo crea problemi. Ad esempio, se si dispone di un'applicazione in cluster che crea blocchi NFSv3 e uno dei nodi ha esito negativo, come procedere? È possibile codificare l'applicazione sui nodi sopravvissuti per rimuovere i blocchi, ma come si fa a sapere che questo è sicuro? Il nodo "guasto" potrebbe essere operativo, ma non comunica con il resto del cluster?

Con NFSv4, i blocchi hanno una durata limitata. Finché il client che mantiene i blocchi continua il check-in con il server NFSv4, nessun altro client è autorizzato ad acquisire tali blocchi. Se un client non riesce a eseguire il check in con NFSv4, i blocchi vengono revocati dal server e gli altri client potranno richiedere e ottenere i blocchi.

NFSv4 leasing

I blocchi NFSv4 sono associati a un lease NFSv4. Quando un client NFSv4 stabilisce una connessione con un server NFSv4, ottiene un lease. Se il client ottiene un blocco (ci sono molti tipi di blocchi) allora il blocco è associato al lease.

Questo lease ha un timeout definito. Per impostazione predefinita, ONTAP imposta il valore di timeout su 30 secondi:

```
Cluster01::*> nfs server show -vserver vserver1 -fields v4-lease-seconds

vserver   v4-lease-seconds
--------- ----------------
vserver1  30
```

Ciò significa che un client NFSv4 deve effettuare il check-in con il server NFSv4 ogni 30 secondi per rinnovare i propri leasing.

Il leasing viene rinnovato automaticamente da qualsiasi attività, quindi se il client sta lavorando non è necessario eseguire operazioni di aggiunta. Se un'applicazione diventa silenziosa e non sta svolgendo un lavoro reale, sarà necessario eseguire una sorta di operazione keep-alive (chiamata SEQUENZA). In sostanza, è solo dire "sono ancora qui, ti prego di rinnovare i miei contratti di leasing".

```
 *Question:* What happens if you lose network connectivity for 31 seconds?
NFSv3 è stateless. Non si aspetta la comunicazione dai clienti. NFSv4 è
stateful, e una volta trascorso il periodo di leasing, il lease scade, i
blocchi vengono revocati e i file bloccati vengono resi disponibili ad
altri client.
```

Con NFSv3, è possibile spostare i cavi di rete, riavviare gli switch di rete, apportare modifiche alla configurazione e assicurarsi che non si verifichi alcun problema. Normalmente, le applicazioni aspettavano solo pazientemente che la connessione di rete funzionasse di nuovo.

Con NFSv4, avete 30 secondi (a meno che non abbiate aumentato il valore di quel parametro all'interno di ONTAP) per completare il vostro lavoro. Se si supera questo limite, il tempo di leasing è scaduto. In genere si verificano arresti anomali delle applicazioni.

Ad esempio, se si dispone di un database Oracle e si verifica una perdita di connettività di rete (talvolta chiamata "partizione di rete") che supera il timeout del lease, il database verrà arrestato.

Di seguito viene riportato un esempio di ciò che accade nel registro degli avvisi di Oracle:

```
2022-10-11T15:52:55.206231-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00202: control file: '/redo0/NTAP/ctrl/control01.ctl'
ORA-27072: File I/O error
Linux-x86_64 Error: 5: Input/output error
Additional information: 4
Additional information: 1
Additional information: 4294967295
2022-10-11T15:52:59.842508-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00206: error in writing (block 3, # blocks 1) of control file
ORA-00202: control file: '/redo1/NTAP/ctrl/control02.ctl'
ORA-27061: waiting for async I/Os failed
```

Se si esaminano i syslogs, si dovrebbero vedere alcuni di questi errori:

```
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
```

I messaggi di registro sono in genere il primo segno di un problema, diverso dal blocco dell'applicazione. In genere, durante l'interruzione della rete non viene visualizzato nulla, poiché i processi e il sistema operativo stesso sono bloccati e tentano di accedere al file system NFS.

Gli errori vengono visualizzati dopo che la rete è nuovamente operativa. Nell'esempio precedente, una volta ristabilita la connettività, il sistema operativo tentava di riacquisire i blocchi, ma era troppo tardi. Il leasing era scaduto e i blocchi sono stati rimossi. Ciò genera un errore che si propaga fino al livello Oracle e causa il messaggio nel registro degli avvisi. È possibile che vengano visualizzate variazioni su questi modelli a seconda della versione e della configurazione del database.

Riassumendo, NFSv3 tollera l'interruzione di rete, ma NFSv4 è più sensibile e impone un periodo di leasing definito.

Cosa succede se un timeout di 30 secondi non è accettabile? Cosa succede se si gestisce una rete a variazione dinamica in cui gli switch vengono riavviati o i cavi vengono ricollocati e il risultato è un'interruzione occasionale della rete? È possibile scegliere di estendere il periodo di leasing, ma se si desidera farlo richiede una spiegazione di NFSv4 periodi di tolleranza.

NFSv4 periodi di grazia

Se un server NFSv3 viene riavviato, è pronto a servire i/o quasi istantaneamente. Non manteneva alcun tipo di stato sui client. Il risultato è che un'operazione di takeover della ONTAP spesso sembra quasi istantanea. Quando un controller è pronto a iniziare a servire i dati, invia un ARP alla rete che segnala la modifica della topologia. I client normalmente rilevano questo quasi istantaneamente e i dati riprendono a fluire.

NFSv4, tuttavia, produrrà una breve pausa. È solo una parte di come funziona NFSv4.

> Le sezioni seguenti sono aggiornate a partire da ONTAP 9.15,1, ma il comportamento lease e lock e le opzioni di ottimizzazione possono cambiare da versione a versione. Se è necessario ottimizzare i timeout di lease/lock NFSv4, consultare il supporto NetApp per le ultime informazioni.

I server NFSv4 devono tenere traccia dei lease, dei blocchi e di chi utilizza i dati. Se un server NFS si riavvia o perde potenza per un momento o viene riavviato durante l'attività di manutenzione, il risultato è il lease/lock e le altre informazioni del client vengono perse. Il server deve individuare quale client utilizza i dati prima di riprendere le operazioni. È qui che entra in gioco il periodo di grazia.

Se all'improvviso si spegne e riaccende il server NFSv4. Quando viene eseguito il backup, i client che tentano di riprendere io riceveranno una risposta che essenzialmente dice: "Ho perso le informazioni di lease/lock. Vuoi registrare nuovamente i blocchi?" Questo è l'inizio del periodo di grazia. Il valore predefinito è 45 secondi su ONTAP:

```
Cluster01::> nfs server show -vserver vserver1 -fields v4-grace-seconds

vserver   v4-grace-seconds
--------- ----------------
vserver1  45
```

Il risultato è che, dopo un riavvio, un controller sospenderà io mentre tutti i client recuperano i loro lease e blocchi. Al termine del periodo di prova, il server riprenderà le operazioni io.

Questo periodo di tolleranza controlla il recupero del leasing durante le modifiche all'interfaccia di rete, ma esiste un secondo periodo di tolleranza che controlla il recupero durante il failover dello storage, `locking.grace_lease_seconds`. Questa è un'opzione a livello di nodo.

```
cluster01::> node run [node names or *] options
locking.grace_lease_seconds
```

Ad esempio, se hai spesso bisogno di eseguire failover LIF ed è necessario ridurre il periodo di tolleranza, cambierai `v4-grace-seconds`. Se si desidera migliorare il tempo di ripresa io durante il failover del controller, è necessario modificare `locking.grace_lease_seconds`.

Alterare questi valori solo con cautela e dopo aver compreso appieno i rischi e le conseguenze. Le pause io coinvolte nelle operazioni di failover e migrazione con NFSv4.X non possono essere evitate del tutto. I periodi di blocco, leasing e grazia fanno parte della RFC NFS. Per molti clienti, NFSv3 è preferibile perché i tempi di failover sono più rapidi.

Timeout leasing vs periodi di grazia

Il periodo di tolleranza e il periodo di leasing sono collegati. Come menzionato sopra, il timeout di lease predefinito è di 30 secondi, il che significa che NFSv4 client devono effettuare il check-in con il server almeno ogni 30 secondi o perdere i lease e, a loro volta, i blocchi. Il periodo di tolleranza esiste per consentire a un server NFS di ricostruire i dati di lease/lock e il valore predefinito è 45 secondi. Il periodo di tolleranza deve essere più lungo del periodo di leasing. In questo modo, un ambiente client NFS progettato per rinnovare i lease almeno ogni 30 secondi avrà la possibilità di effettuare il check-in con il server dopo un riavvio. Un periodo di tolleranza di 45 secondi assicura che tutti quei clienti che si aspettano di rinnovare i loro leasing almeno ogni 30 secondi definitivamente hanno l'opportunità di farlo.

Se un timeout di 30 secondi non è accettabile, è possibile scegliere di prolungare il periodo di leasing.

Se si desidera aumentare il timeout del lease a 60 secondi per resistere a un'interruzione di rete di 60 secondi, sarà necessario aumentare anche il periodo di tolleranza. Ciò significa che si verificheranno pause di i/o più lunghe durante il failover del controller.

Normalmente questo non dovrebbe essere un problema. Gli utenti tipici aggiornano i controller ONTAP solo una o due volte all'anno e il failover non pianificato dovuto a guasti hardware è estremamente raro. Inoltre, se aveste una rete in cui un'interruzione di rete di 60 secondi fosse una possibilità preoccupante e aveste bisogno di un timeout del leasing di 60 secondi, probabilmente non vi opporreste a un raro failover del sistema storage con una pausa di 61 secondi. Hai già riconosciuto che la tua rete è in pausa per più di 60 secondi piuttosto frequentemente.

Caching di NFS

La presenza di una delle seguenti opzioni di montaggio causa la disattivazione della cache host:

```
cio, actimeo=0, noac, forcedirectio
```

Queste impostazioni possono avere un grave effetto negativo sulla velocità di installazione del software, l'applicazione di patch e le operazioni di backup/ripristino. In alcuni casi, in particolare con le applicazioni in cluster, queste opzioni sono necessarie come conseguenza inevitabile della necessità di garantire la coerenza della cache in tutti i nodi del cluster. In altri casi, i clienti utilizzano erroneamente questi parametri e il risultato è un inutile danno alle prestazioni.

Molti clienti rimuovono temporaneamente queste opzioni di montaggio durante l'installazione o l'applicazione di patch dei file binari. Questa rimozione può essere eseguita in modo sicuro se l'utente verifica che nessun altro processo stia utilizzando attivamente la directory di destinazione durante il processo di installazione o di applicazione delle patch.

Dimensioni trasferimento NFS

Per impostazione predefinita, ONTAP limita le dimensioni i/o NFS a 64K.

L'i/o casuale con la maggior parte delle applicazioni e dei database utilizza blocchi di dimensioni molto inferiori, ben al di sotto del limite massimo di 64K KB. L'i/o a blocchi di grandi dimensioni è solitamente a parallelismo, pertanto anche il massimo di 64K Gbps non costituisce un limite all'ottenimento della massima larghezza di banda.

Ci sono alcuni carichi di lavoro in cui il massimo di 64K crea un limite. In particolare, le operazioni single-threaded, come l'operazione di backup o ripristino o la scansione di un database completa della tabella,

vengono eseguite più velocemente e in modo più efficiente se il database è in grado di eseguire un numero di i/o inferiore ma maggiore. Le dimensioni ottimali per la gestione i/o per ONTAP sono 256K KB.

Le dimensioni massime di trasferimento per una SVM ONTAP possono essere modificate come segue:

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

> Non diminuire mai la dimensione massima di trasferimento consentita su ONTAP al di sotto del valore rsize/wsize dei file system NFS attualmente montati. In alcuni sistemi operativi, ciò può causare blocchi o addirittura danni ai dati. Ad esempio, se i client NFS sono attualmente impostati su un valore rsize/wsize di 65536, la dimensione massima di trasferimento ONTAP potrebbe essere regolata tra 65536 e 1048576 senza alcun effetto perché i client stessi sono limitati. La riduzione della dimensione massima di trasferimento inferiore a 65536 GB può danneggiare la disponibilità o i dati.

NVFAIL

NVFAIL è una funzionalità di ONTAP che garantisce l'integrità in scenari di failover catastrofici.

I database sono vulnerabili al danneggiamento durante gli eventi di failover dello storage perché mantengono grandi cache interne. Se un evento catastrofico richiede l'imposizione di un failover ONTAP o il forzamento dello switchover MetroCluster, a prescindere dallo stato di salute della configurazione complessiva, il risultato viene riconosciuto in precedenza che le modifiche potrebbero essere effettivamente scartate. Il contenuto dell'array di storage torna indietro nel tempo e lo stato della cache del database non riflette più lo stato dei dati su disco. Questa incoerenza provoca il danneggiamento dei dati.

La memorizzazione nella cache può avvenire a livello di applicazione o di server. Ad esempio, una configurazione Oracle Real Application Cluster (RAC) con server attivi sia su un sito primario che su un sito remoto memorizza nella cache i dati all'interno di Oracle SGA. Un'operazione di switchover forzata che comportava la perdita di dati rischierebbe di danneggiare il database poiché i blocchi archiviati nell'SGA potrebbero non corrispondere ai blocchi su disco.

Un uso meno ovvio della memorizzazione nella cache è a livello del file system del sistema operativo. I blocchi di un file system NFS montato possono essere memorizzati nella cache del sistema operativo. In alternativa, un file system in cluster basato su LUN che si trovano nel sito primario può essere montato sui server nel sito remoto e, ancora una volta, i dati possono essere memorizzati nella cache. In queste situazioni, un errore della NVRAM, un takeover forzato o uno switchover forzato possono danneggiare il file system.

ONTAP protegge i database e i sistemi operativi da questo scenario con NVFAIL e le relative impostazioni.

ASM Reclamation Utility (ASMRU)

ONTAP rimuove in modo efficiente i blocchi azzerati scritti su un file o LUN quando la

compressione inline è abilitata. Utility come Oracle ASM Reclamation Utility (ASRU) funzionano scrivendo zero in estensioni ASM non utilizzate.

In questo modo, gli amministratori di database possono recuperare spazio sull'array di storage dopo l'eliminazione dei dati. ONTAP intercetta gli zero e dealloca lo spazio dal LUN. Il processo di recupero dei dati è estremamente rapido, poiché non viene scritto alcun dato all'interno del sistema di storage.

Dal punto di vista del database, il gruppo di dischi ASM contiene zero; la lettura di tali aree del LUN produce un flusso di zero, tuttavia ONTAP non memorizza gli zero sui dischi. Vengono invece apportate semplici modifiche ai metadati che contrassegnano internamente le aree azzerate del LUN come vuote di qualsiasi dato.

Per motivi simili, il test delle performance che implica dati azzerati non è valido, in quanto i blocchi di zero non vengono effettivamente elaborati come scritture all'interno dello storage array.

Quando si utilizza ASRU, assicurarsi che tutte le patch consigliate da Oracle siano installate.

Virtualizzazione

La virtualizzazione dei database con VMware, Oracle OLVM o KVM è una scelta sempre più comune per i clienti NetApp che hanno scelto la virtualizzazione anche per i database mission-critical.

Supportabilità

Esistono numerosi preconcetti sui criteri di supporto Oracle per la virtualizzazione, in particolare per i prodotti VMware. Non è raro che Oracle Outright non supporti la virtualizzazione. Questa nozione non è corretta e porta alla perdita di opportunità per trarre vantaggio dalla virtualizzazione. Oracle Doc ID 249212,1 illustra i requisiti effettivi e raramente viene considerato un problema da parte dei clienti.

Se si verifica un problema su un server virtualizzato e il supporto di Oracle non lo conosce, al cliente potrebbe essere richiesto di riprodurre il problema sull'hardware fisico. Un cliente Oracle che utilizza una versione all'avanguardia di un prodotto potrebbe non voler utilizzare la virtualizzazione a causa di potenziali problemi di supportabilità, ma questa situazione non è stata un problema reale per i clienti che utilizzano versioni di prodotti Oracle generalmente disponibili.

Presentazione storage

I clienti che stanno considerando la virtualizzazione dei propri database devono basare le proprie decisioni di storage sulle esigenze aziendali. Sebbene questa affermazione sia generalmente vera per tutte le decisioni IT, è particolarmente importante per i progetti di database, poiché le dimensioni e l'ambito dei requisiti variano notevolmente.

Sono disponibili tre opzioni di base per la presentazione dello storage:

- LUN virtualizzate nei datastore di hypervisor
- LUN iSCSI gestite dall'iniziatore iSCSI sulla macchina virtuale, non dall'hypervisor
- File system NFS montati dalla macchina virtuale (non da un datastore basato su NFS)
- Mappatura diretta dei dispositivi. Gli RDM VMware sono svantaggiati dai clienti, ma spesso i dispositivi fisici sono ancora mappati in modo simile direttamente con la virtualizzazione KVM e OLVM.

Performance

Il metodo di presentazione dello storage a un guest virtualizzato non influisce in genere sulle prestazioni. I sistemi operativi host, i driver di rete virtualizzati e le implementazioni del datastore degli hypervisor sono tutti altamente ottimizzati e generalmente possono consumare tutta la larghezza di banda della rete FC o IP disponibile tra l'hypervisor e il sistema storage, purché vengano seguite le Best practice di base. In alcuni casi, ottenere prestazioni ottimali può essere leggermente più semplice utilizzando un approccio di presentazione dello storage rispetto a un altro, ma il risultato finale dovrebbe essere comparabile.

Gestibilità

Il fattore chiave nella scelta di come presentare lo storage a un guest virtualizzato è la manovrabilità. Non esiste un metodo giusto o sbagliato. L'approccio migliore dipende dalle esigenze operative, dalle competenze e dalle preferenze DELL'IT.

I fattori da prendere in considerazione includono:

- **Trasparenza.** quando una VM gestisce i propri file system, è più facile per un amministratore di database o un amministratore di sistema identificare l'origine dei file system per i propri dati. L'accesso ai file system e ai LUN non avviene in modo diverso rispetto a un server fisico.
- **Coerenza.** quando una VM è proprietaria dei file system, l'utilizzo o il mancato utilizzo di un livello di hypervisor influisce sulla gestibilità. È possibile utilizzare le stesse procedure per il provisioning, il monitoraggio, la protezione dei dati e così via nell'intero ambiente, inclusi ambienti virtualizzati e non.

  D'altra parte, in un data center altrimenti virtualizzato al 100% potrebbe essere preferibile utilizzare anche lo storage basato su datastore per l'intera impronta, sulla stessa logica sopra menzionata, la coerenza, la capacità di utilizzare le stesse procedure per il provisioning, la protezione, il montoring e la protezione dei dati.

- **Stabilità e risoluzione dei problemi.** quando una VM possiede i propri file system, fornire prestazioni buone e stabili e risolvere i problemi è più semplice perché l'intero stack di storage è presente sulla VM. L'unico ruolo dell'hypervisor è il trasporto di frame FC o IP. Quando un datastore è incluso in una configurazione, complica la configurazione introducendo un altro insieme di timeout, parametri, file di log e potenziali bug.
- **Portabilità.** quando una VM è proprietaria dei suoi file system, il processo di spostamento di un ambiente Oracle diventa molto più semplice. I file system possono essere spostati facilmente tra guest virtualizzati e non.
- **Vendor lock-in.** una volta posizionati i dati in un datastore, diventa difficile utilizzare un hypervisor diverso o estrarre i dati dall'ambiente virtualizzato.
- **Abilitazione snapshot.** le procedure di backup tradizionali in un ambiente virtualizzato possono diventare un problema a causa della larghezza di banda relativamente limitata. Ad esempio, un trunk 10GbE a quattro porte potrebbe essere sufficiente per supportare le esigenze quotidiane di prestazioni di molti database virtualizzati, ma tale trunk non sarebbe sufficiente per eseguire backup utilizzando RMAN o altri prodotti di backup che richiedono lo streaming di una copia di dimensioni complete dei dati. Il risultato è che un ambiente virtualizzato sempre più consolidato ha bisogno di eseguire backup tramite snapshot di storage. In questo modo si evita la necessità di sovrascrivere la configurazione dell'hypervisor solo per supportare i requisiti di larghezza di banda e CPU nella finestra di backup.

  L'utilizzo di file system guest facilita a volte l'utilizzo di backup e ripristini basati su snapshot, poiché gli oggetti storage che necessitano di protezione possono essere indirizzati più facilmente. Tuttavia, esiste un numero sempre maggiore di prodotti di data Protection di virtualizzazione che si integrano perfettamente con datastore e snapshot. La strategia di backup deve essere considerata attentamente prima di prendere una decisione su come presentare lo storage a un host virtualizzato.

Driver paravirtualizzati

Per prestazioni ottimali, l'uso di driver di rete paravirtualizzati è fondamentale. Quando si utilizza un datastore, è necessario un driver SCSI paravirtualizzato. Un driver di dispositivo paravirtualizzato consente a un guest di integrarsi più profondamente nell'hypervisor, invece di un driver emulato in cui l'hypervisor spende più tempo CPU che imita il comportamento dell'hardware fisico.

Overcommit RAM

L'overcommit della RAM implica la configurazione di una quantità di RAM virtualizzata su vari host superiore a quella presente sull'hardware fisico. In caso contrario, si potrebbero verificare problemi di prestazioni imprevisti. Quando si virtualizza un database, i blocchi sottostanti di Oracle SGA non devono essere sostituiti con lo storage dall'hypervisor. Ciò causa risultati di prestazioni altamente instabili.

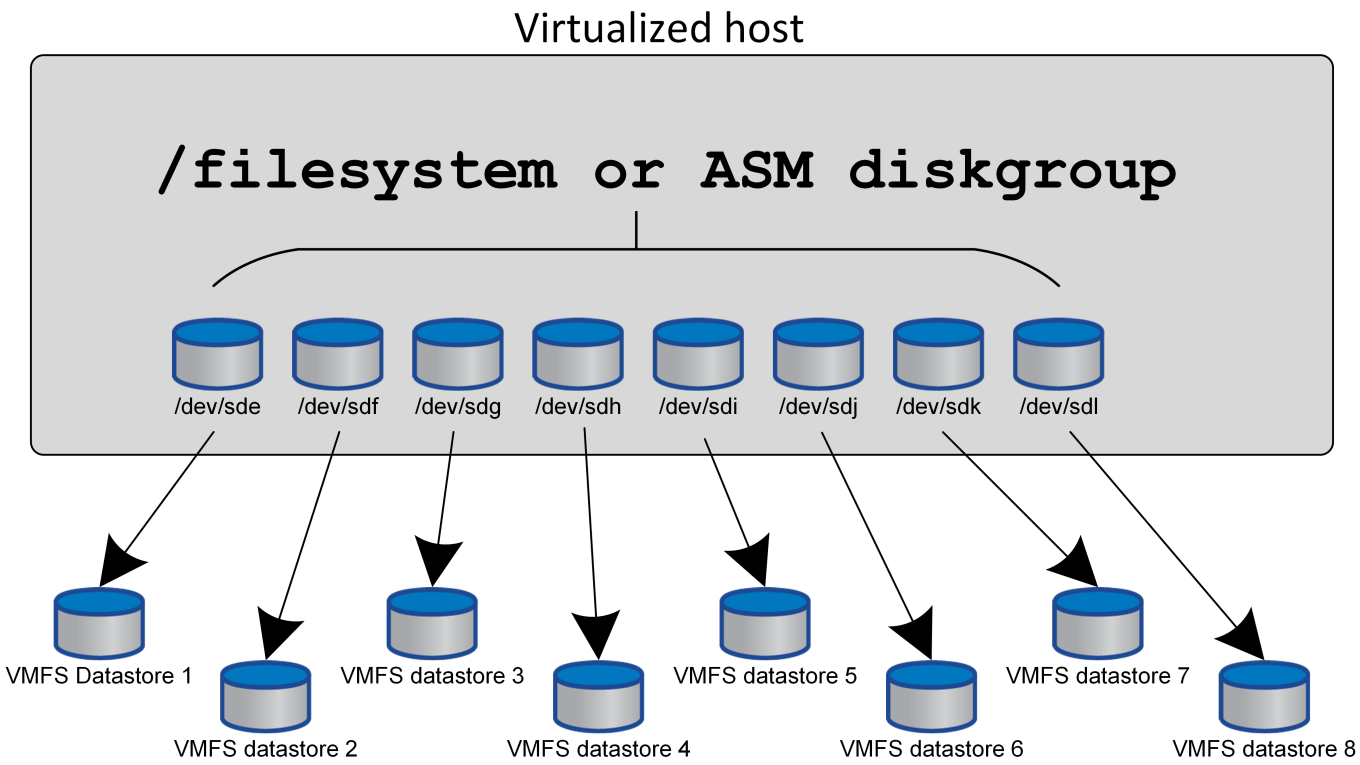
Striping dei datastore

Quando si utilizzano database con datastore, c'è un fattore critico da considerare in relazione allo striping delle performance.

Le tecnologie dei datastore come VMFS sono in grado di estendersi su più LUN, ma non su dispositivi suddivisi in striping. I LUN sono concatenati. Il risultato finale può essere costituito da hot spot LUN. Ad esempio, un database Oracle tipico potrebbe avere un gruppo di dischi ASM di 8 LUN. È possibile eseguire il provisioning di tutte le 8 LUN virtualizzate in un datastore VMFS da 8 LUN, senza tuttavia alcuna garanzia su quali LUN risiedono i dati. La configurazione risultante potrebbe essere tutta una LUN virtualizzata da 8 GB che occupa una singola LUN nel datastore VMFS. Ciò si traduce in un collo di bottiglia per le prestazioni.

Lo striping è in genere necessario. Con alcuni hypervisor, incluso KVM, è possibile creare un datastore utilizzando lo striping LVM come descritto "qui". Con VMware, l'architettura appare un po' diversa. Ogni LUN virtualizzata deve essere posizionata in un datastore VMFS diverso.

Ad esempio:



Il driver principale di questo approccio non è ONTAP, ma è dovuto alla limitazione intrinseca del numero di operazioni che una singola VM o LUN dell'hypervisor può eseguire in parallelo. In genere, un singolo LUN ONTAP può supportare un maggior numero di IOPS rispetto a quello richiesto da un host. Il limite di prestazioni di un singolo LUN è quasi universalmente il risultato del sistema operativo host. Il risultato è che per soddisfare le esigenze di performance della maggior parte dei database sono necessarie LUN comprese tra 4 e 8 GB.

Le architetture VMware devono pianificare con attenzione le proprie architetture per garantire che questo approccio non soddisfi i massimi di datastore e/o percorso LUN. Inoltre, non è necessario un set univoco di datastore VMFS per ogni database. L'esigenza principale consiste nel garantire che ogni host disponga di un set pulito di percorsi io da 4-8 GB dalle LUN virtualizzate alle LUN di backend sul sistema storage stesso. In rare occasioni, anche un numero maggiore di datatores può rivelarsi vantaggioso per richieste di performance veramente estreme, ma le LUN da 4-8 GB sono in genere sufficienti per il 95% di tutti i database. Un singolo volume ONTAP contenente 8 LUN può supportare fino a 250.000 IOPS casuali con blocchi Oracle con una tipica configurazione di sistema operativo/ONTAP/rete.

Tiering

Panoramica

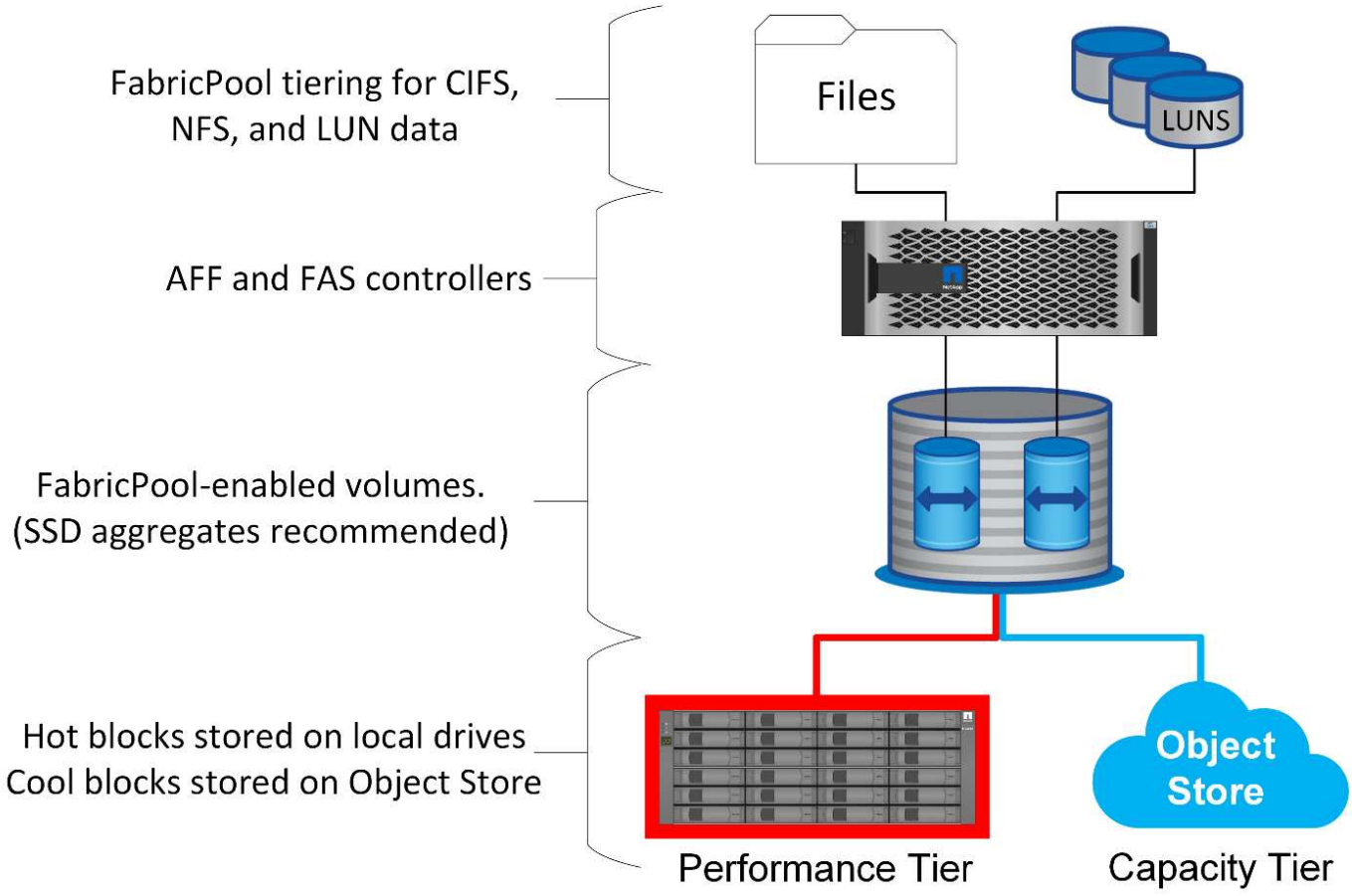
La comprensione dell'impatto del tiering FabricPool su Oracle e altri database richiede una conoscenza dell'architettura FabricPool di basso livello.

Architettura

FabricPool è una tecnologia di tiering che classifica i blocchi come "hot" o "cool" e li colloca nel Tier di storage più appropriato. Il Tier di performance è nella maggior parte dei casi collocato nello storage SSD e ospita i blocchi di dati "hot". Il Tier di capacità si trova in un archivio di oggetti e ospita i blocchi di dati "cool". Il supporto per lo storage a oggetti include NetApp StorageGRID, ONTAP S3, archiviazione BLOB di Microsoft

Azure, il servizio di storage a oggetti Alibaba Cloud, archiviazione a oggetti IBM Cloud, archiviazione Google Cloud e Amazon AWS S3.

Sono disponibili più policy di tiering che controllano le modalità di classificazione dei blocchi come "hot" o "cool", che possono essere impostate in base al volume e modificate secondo necessità. Solo i blocchi di dati vengono spostati tra i Tier di performance e capacità. I metadati che definiscono la struttura LUN e del file system rimangono sempre sul Tier di performance. Di conseguenza, la gestione è centralizzata su ONTAP. I file e le LUN non appaiono diversi dai dati memorizzati in qualsiasi altra configurazione ONTAP. Il controller NetApp AFF o FAS applica le policy definite per spostare i dati nel Tier appropriato.



Provider di archivi di oggetti

I protocolli di storage a oggetti utilizzano semplici richieste HTTP o HTTPS per la memorizzazione di un grande numero di oggetti dati. L'accesso allo storage a oggetti deve essere affidabile, poiché l'accesso ai dati da parte di ONTAP dipende dalla puntuale manutenzione delle richieste. Le opzioni includono le opzioni Amazon S3 Standard e accesso poco frequente, Microsoft Azure Hot e Cool Blob Storage, IBM Cloud e Google Cloud. Le opzioni di archiviazione come Amazon Glacier e Amazon Archive non sono supportate, perché il tempo necessario per recuperare i dati può superare le tolleranze dei sistemi operativi e delle applicazioni host.

NetApp StorageGRID è anche supportato e rappresenta una soluzione di livello Enterprise ottimale. Si tratta di un sistema storage a oggetti dalle performance elevate, scalabile e altamente sicuro, in grado di fornire ridondanza geografica per i dati FabricPool nonché per altre applicazioni di archivi di oggetti che hanno sempre più probabilità di far parte di ambienti applicativi Enterprise.

StorageGRID può anche ridurre i costi evitando le spese di uscita imposte da molti provider di cloud pubblici per la lettura dei dati di nuovo dai propri servizi.

Dati e metadati

Si noti che il termine "dati" si applica in questo caso ai blocchi di dati effettivi, non ai metadati. Viene eseguito il tiering solo dei blocchi di dati, mentre i metadati rimangono nel Tier di performance. Inoltre, lo stato di un blocco come caldo o freddo è influenzato solo dalla lettura del blocco di dati effettivo. La semplice lettura del nome, dell'indicatore data e ora o dei metadati di proprietà di un file non influisce sulla posizione dei blocchi di dati sottostanti.

Backup

Anche se FabricPool può ridurre significativamente l'impatto dello storage, non rappresenta di per sé una soluzione di backup. I metadati NetApp WAFL rimangono sempre nel Tier di performance. Se un disastro catastrofico distrugge il Tier di performance, non è possibile creare un nuovo ambiente utilizzando i dati sul Tier di capacità perché non contiene metadati WAFL.

FabricPool, tuttavia, può entrare a far parte di una strategia di backup. Ad esempio, FabricPool può essere configurato con la tecnologia di replica NetApp SnapMirror. Ciascuna metà del mirror può avere la propria connessione a una destinazione dello storage a oggetti. Il risultato sono due copie indipendenti dei dati. La copia primaria è costituita dai blocchi sul Tier di performance e dai blocchi associati nel Tier di capacità, mentre la replica è un secondo set di blocchi di performance e capacità.

Policy di tiering

Policy di tiering

In ONTAP sono disponibili quattro criteri che controllano il modo in cui i dati Oracle sul livello di prestazioni diventano candidati per il trasferimento al livello di capacità.

Solo snapshot

Il `snapshot-only` `tiering-policy` si applica solo ai blocchi non condivisi con il file system attivo. Essenzialmente si traduce in tiering dei backup del database. I blocchi diventano candidati per il tiering dopo la creazione di uno snapshot e il blocco viene quindi sovrascritto, generando un blocco presente solo all'interno dello snapshot. Il ritardo prima di un `snapshot-only` il blocco è considerato freddo e controllato da `tiering-minimum-cooling-days` impostazione del volume. L'intervallo a partire da ONTAP 9,8 è compreso tra 2 e 183 giorni.

Molti set di dati hanno tassi di cambiamento bassi, con conseguenti risparmi minimi derivanti da questa policy. Ad esempio, un database tipico osservato con ONTAP ha un tasso di variazione inferiore al 5% alla settimana. I log di archivio dei database possono occupare spazio esteso, ma in genere continuano a esistere nel file system attivo e pertanto non possono essere candidati per il tiering in base a questa policy.

Automatico

Il `auto` la policy di tiering estende il tiering sia a blocchi specifici di snapshot che a blocchi nel file system attivo. Il ritardo prima che un blocco venga considerato freddo è controllato dall' `tiering-minimum-cooling-days` impostazione del volume. L'intervallo a partire da ONTAP 9,8 è compreso tra 2 e 183 giorni.

Questo approccio abilita opzioni di tiering che non sono disponibili con `snapshot-only` policy. Ad esempio, un criterio di protezione dei dati potrebbe richiedere la conservazione di 90 giorni di determinati file di registro. L'impostazione di un periodo di raffreddamento di 3 giorni comporta il tiering di tutti i file di registro precedenti a 3 giorni dal livello delle prestazioni. Questa azione libera spazio sostanziale sul Tier delle performance, consentendoti comunque di visualizzare e gestire tutti e 90 i giorni di dati.

Nessuno

Il `none` la policy di tiering impedisce il tiering di blocchi aggiuntivi dal layer di storage, ma i dati ancora presenti nel tier di capacità rimangono nel tier di capacità fino a quando non vengono letti. Se quindi il blocco viene letto, viene tirato indietro e posizionato nel Tier di performance.

Il motivo principale per cui si utilizza `none` la policy di tiering impedisce il tiering dei blocchi, ma nel tempo potrebbe risultare utile modificarli. Ad esempio, supponiamo che un set di dati specifico venga suddiviso in Tier per il livello di capacità, ma sorge un'esigenza inaspettata di funzionalità di performance complete. La policy può essere modificata per impedire qualsiasi tiering aggiuntivo e per confermare che i blocchi letti nuovamente quando l'io aumenta rimangono nel Tier di performance.

Tutto

Il `all` la policy di tiering sostituisce `backup` Policy in data ONTAP 9,6. Il `backup` Policy applicata solo ai volumi di data Protection, che significa destinazione SnapMirror o NetApp SnapVault. Il `all` le funzioni dei criteri sono identiche, ma non si limitano ai volumi di protezione dei dati.

Grazie a questa policy, i blocchi vengono immediatamente considerati COOL e possono essere immediatamente suddivisi in Tier nel livello di capacità.

Questo criterio è particolarmente appropriato per i backup a lungo termine. Può anche essere utilizzato come forma di gestione gerarchica dello storage (HSM, Hierarchical Storage Management). In passato, HSM veniva comunemente utilizzato per eseguire il tiering dei blocchi di dati di un file su nastro, mantenendo il file stesso visibile nel file system. Un volume FabricPool con `all` il criterio consente di archiviare i file in un archivio visibile e gestibile pur non occupando quasi nessuno spazio nel livello di storage locale.

Criteri di recupero

I criteri di tiering controllano i blocchi di database Oracle sottoposti a tiering dal Tier di performance al Tier di capacità. I criteri di recupero controllano ciò che accade quando viene letto un blocco a cui è stato eseguito il tiering.

Predefinito

Tutti i volumi FabricPool sono inizialmente impostati su `default`, il che significa che il comportamento è controllato da `cloud-retrieval-policy. `il comportamento esatto dipende dal criterio di tiering utilizzato.

- `auto`consente di recuperare solo dati letti in modo casuale
- `snapshot-only`- consente di recuperare tutti i dati letti in modo sequenziale o casuale
- `none`- consente di recuperare tutti i dati letti in modo sequenziale o casuale
- `all`- non recuperare i dati dal tier di capacità

A lettura

Impostazione `cloud-retrieval-policy` in lettura sovrascrive il comportamento predefinito, in modo che una lettura di dati a livelli determini il ritorno dei dati al livello di prestazioni.

Ad esempio, un volume potrebbe essere stato leggermente utilizzato per un lungo periodo sotto il `auto` la policy di tiering e la maggior parte dei blocchi ora vengono suddivisi in livelli.

Se una modifica imprevista delle esigenze aziendali richiedeva la scansione ripetuta di alcuni dati per

preparare un determinato rapporto, potrebbe essere opportuno modificare `cloud-retrieval-policy` a. `on-read` per garantire che tutti i dati letti vengano restituiti al livello delle prestazioni, inclusi i dati letti in modo sequenziale e casuale. In questo modo si migliorano le prestazioni dell'i/o sequenziale rispetto al volume.

Promuovi

Il comportamento della policy di promozione dipende dalla policy di tiering. Se la policy di tiering è `auto`, quindi impostare `cloud-retrieval-policy `to `promote` riporta tutti i blocchi dal tier di capacità nella successiva scansione del tiering.

Se la policy di tiering è `snapshot-only`, gli unici blocchi restituiti sono i blocchi associati al file system attivo. Normalmente questo non avrebbe alcun effetto perché gli unici blocchi suddivisi in livelli sotto `snapshot-only` la policy dovrebbe essere costituita da blocchi associati esclusivamente agli snapshot. Nel file system attivo non sono presenti blocchi a livelli.

Se, tuttavia, i dati di un volume sono stati ripristinati da un'operazione SnapRestore di volume o di file-clone da una snapshot, alcuni dei blocchi suddivisi in Tier perché associati solo a snapshot potrebbero ora essere richiesti dal file system attivo. Potrebbe essere opportuno modificare temporaneamente `cloud-retrieval-policy` policy to `promote` per recuperare rapidamente tutti i blocchi richiesti localmente.

Mai

Non recuperare i blocchi dal Tier di capacità.

Strategie di tiering

Tiering completo dei file

Anche se il tiering FabricPool opera a livello di blocco, in alcuni casi può essere utilizzato per fornire un tiering a livello di file.

Molti set di dati delle applicazioni sono organizzati per data e tali dati hanno generalmente sempre meno probabilità di accedere man mano che invecchiano. Ad esempio, una banca potrebbe avere un archivio di file PDF che contengono cinque anni di dichiarazioni dei clienti, ma solo gli ultimi mesi sono attivi. FabricPool può essere utilizzato per spostare i file di dati meno recenti nel Tier di capacità. Un periodo di raffreddamento di 14 giorni garantirebbe che i 14 giorni più recenti di file PDF rimangano sul livello di prestazioni. Inoltre, i file letti almeno ogni 14 giorni resterebbero hot e quindi nel Tier di performance.

Policy

Per implementare un approccio di tiering basato su file, è necessario disporre di file scritti e non modificati successivamente. Il `tiering-minimum-cooling-days` i criteri devono essere impostati su un livello sufficientemente alto da mantenere i file di cui potresti aver bisogno nel tier di performance. Ad esempio, un set di dati per il quale sono necessari gli ultimi 60 giorni di dati con performance ottimali garantisce la definizione di `tiering-minimum-cooling-days` periodo a 60. Risultati simili possono essere ottenuti anche in base ai modelli di accesso ai file. Ad esempio, se sono necessari gli ultimi 90 giorni di dati e l'applicazione sta accedendo a quell'arco di dati di 90 giorni, i dati resteranno sul Tier di performance. Impostando `tiering-minimum-cooling-days` a 2, si ottiene un tiering prompt dopo che i dati sono meno attivi.

Il `auto` la policy è necessaria per gestire il tiering di questi blocchi perché solo l' `auto` il criterio influisce sui blocchi che si trovano nel file system attivo.

Qualsiasi tipo di accesso ai dati ripristina i dati della mappa termica. La scansione virus, l'indicizzazione e persino le attività di backup in grado di leggere i file di origine impediscono il tiering perché è necessario `tiering-minimum-cooling-days` la soglia non viene mai raggiunta.

Tiering parziale dei file

Poiché FabricPool opera a livello di blocchi, i file soggetti a modifiche possono essere parzialmente suddivisi in Tier nello storage a oggetti e rimanere parzialmente anche nel Tier di performance.

Ciò è comune con i database. Anche i database che contengono blocchi inattivi sono candidati per il tiering FabricPool. Ad esempio, un database di gestione della catena logistica potrebbe contenere informazioni cronologiche che devono essere disponibili se necessario ma non accessibili durante le normali operazioni. La funzione FabricPool può essere utilizzata per spostare selettivamente i blocchi inattivi.

Ad esempio, i file di dati in esecuzione su un volume FabricPool con un `tiering-minimum-cooling-days` il periodo di 90 giorni conserva i blocchi a cui si accede nei 90 giorni precedenti nel tier di performance. Tuttavia, qualsiasi elemento a cui non si accede per 90 giorni viene ricollocato nel Tier di capacità. In altri casi, la normale attività applicativa preserva i blocchi corretti sul livello corretto. Ad esempio, se un database viene normalmente utilizzato per elaborare regolarmente i 60 giorni precedenti di dati, è molto più basso `tiering-minimum-cooling-days` il periodo può essere impostato perché l'attività naturale dell'applicazione garantisce che i blocchi non vengano spostati prematuramente.

Il `auto` i criteri devono essere utilizzati con attenzione per i database. Numerosi database prevedono attività periodiche come la fine del quarter o la reindicizzazione delle operazioni. Se il periodo di queste operazioni è superiore a. `tiering-minimum-cooling-days` possono verificarsi problemi di prestazioni. Ad esempio, se l'elaborazione a fine quarter richiede 1TB TB di dati che non vengono intatti, è possibile che tali dati siano presenti nel Tier di capacità. Le letture dal Tier di capacità sono spesso estremamente veloci e potrebbero non causare problemi di performance, ma i risultati esatti dipendono dalla configurazione dell'archivio di oggetti.

Policy

Il `tiering-minimum-cooling-days` il criterio deve essere impostato su un livello sufficientemente alto da conservare i file che potrebbero essere necessari nel livello di prestazioni. Ad esempio, un database in cui potrebbero essere necessari gli ultimi 60 giorni di dati con prestazioni ottimali giustificherebbe l'impostazione di `tiering-minimum-cooling-days` periodo a 60 giorni. Risultati simili possono essere ottenuti anche in base ai modelli di accesso dei file. Ad esempio, se sono necessari gli ultimi 90 giorni di dati e l'applicazione sta accedendo a quell'arco di dati di 90 giorni, i dati resteranno sul Tier di performance. Impostazione di `tiering-minimum-cooling-days` un periodo di 2 giorni eseguirebbe il tiering dei dati non appena i dati diventano meno attivi.

Il `auto` la policy è necessaria per gestire il tiering di questi blocchi perché solo l' `auto` il criterio influisce sui blocchi che si trovano nel file system attivo.

Qualsiasi tipo di accesso ai dati ripristina i dati della mappa termica. Pertanto, le scansioni delle tabelle complete dei database e persino le attività di backup in grado di leggere i file di origine impediscono il tiering perché necessario `tiering-minimum-cooling-days` la soglia non viene mai raggiunta.

Tiering dei log di archivio

Forse l'utilizzo più importante per FabricPool è il miglioramento dell'efficienza dei dati cold noti, come i log delle transazioni dei database.

La maggior parte dei database relazionali opera in modalità di archiviazione dei log delle transazioni per fornire un ripristino point-in-time. Le modifiche apportate ai database vengono salvate registrando le modifiche nei registri delle transazioni e il registro delle transazioni viene conservato senza essere sovrascritto. Il risultato può essere la necessità di conservare un enorme volume di registri delle transazioni archiviati. Esempi simili esistono con molti altri flussi di lavoro delle applicazioni che generano dati che devono essere conservati, ma con molte probabilità di accesso.

FabricPool risolve questi problemi offrendo una singola soluzione con tiering integrato. I file vengono memorizzati e rimangono accessibili nella loro posizione abituale, ma non occupano praticamente spazio nell'array primario.

Policy

Utilizzare un `tiering-minimum-cooling-days` la policy di pochi giorni comporta la conservazione dei blocchi nei file creati di recente (che sono i file più probabilmente necessari a breve termine) nel tier di performance. I blocchi di dati dei file meno recenti vengono quindi spostati nel Tier di capacità.

Il `auto` applica il tiering prompt quando viene raggiunta la soglia di raffreddamento, indipendentemente dal fatto che i log siano stati eliminati o continuino a esistere nel file system primario. Inoltre, l'archiviazione di tutti i log potenzialmente necessari in un'unica posizione nel file system attivo semplifica la gestione. Non c'è motivo di cercare tra gli snapshot per individuare un file che deve essere ripristinato.

Alcune applicazioni, come Microsoft SQL Server, troncano i file di log delle transazioni durante le operazioni di backup in modo che i log non si trovino più nel file system attivo. È possibile risparmiare capacità utilizzando `snapshot-only` tiering delle policy, ma `auto` il criterio non è utile per i dati di log perché raramente dovrebbero essere raffreddati i dati di log nel file system attivo.

Tiering delle Snapshot

La release iniziale di FabricPool era rivolta a un caso di utilizzo di backup. L'unico tipo di blocchi che è possibile eseguire il tiering era costituito da blocchi che non erano più associati a dati nel file system attivo. Pertanto, solo i blocchi di dati Snapshot possono essere spostati nel Tier di capacità. Questa rimane una delle opzioni di tiering più sicure quando occorre, in modo da garantire che le performance non subiscano alcun impatto.

Criteri - istantanee locali

Esistono due opzioni per il tiering di blocchi di snapshot inattivi nel Tier di capacità. Innanzitutto, la `snapshot-only` la politica riguarda solo i blocchi di snapshot. Anche se il `auto` il criterio include `snapshot-only` ed esegue il tiering dei blocchi dal file system attivo. Ciò potrebbe non essere desiderabile.

Il `tiering-minimum-cooling-days` valore deve essere impostato su un periodo di tempo in cui i dati che potrebbero essere necessari durante un ripristino sono disponibili sul livello di prestazioni. Ad esempio, la maggior parte degli scenari di ripristino di un database di produzione critico include un punto di ripristino in un determinato momento dei giorni precedenti. Impostazione a. `tiering-minimum-cooling-days` il valore 3 garantisce che qualsiasi ripristino del file porti a un file che offre immediatamente le massime prestazioni. Tutti i blocchi dei file attivi sono ancora presenti sullo storage veloce senza dover ripristinarli dal livello di capacità.

Criteri - istantanee replicate

Di norma, uno snapshot replicato con SnapMirror o SnapVault utilizzato solo per il ripristino deve utilizzare FabricPool `all` policy. Con questa policy, i metadati vengono replicati, ma tutti i blocchi di dati vengono inviati immediatamente al Tier di capacità, ottenendo il massimo delle performance. La maggior parte dei processi di recovery implica un i/o sequenziale, che è intrinsecamente efficiente. È necessario valutare il tempo di ripristino dalla destinazione dell'archivio oggetti, ma in un'architettura ben progettata questo processo di ripristino non deve essere significativamente più lento del ripristino da dati locali.

Se per il cloning è prevista anche l'utilizzo dei dati replicati, l' `auto` la politica è più appropriata, con un `tiering-minimum-cooling-days` valore che comprende i dati che si prevede vengano utilizzati regolarmente in un ambiente di clonazione. Ad esempio, il working set attivo di un database potrebbe includere dati letti o scritti nei tre giorni precedenti, ma potrebbe includere anche altri 6 mesi di dati storici. In tal caso, il `auto` La policy nella destinazione di SnapMirror rende disponibile il working set nel Tier di performance.

Tiering del backup

I backup delle applicazioni tradizionali includono prodotti come Oracle Recovery Manager, che creano backup basati su file al di fuori della posizione del database originale.

```
 `tiering-minimum-cooling-days` policy of a few days preserves the most
recent backups, and therefore the backups most likely to be required for
an urgent recovery situation, on the performance tier. The data blocks of
the older files are then moved to the capacity tier.
Il `auto` il criterio è il criterio più appropriato per i dati di backup.
In questo modo si garantisce un tiering rapido quando la soglia di
raffreddamento è stata raggiunta, indipendentemente dal fatto che i file
siano stati eliminati o continuino a esistere nel file system primario.
Inoltre, l'archiviazione di tutti i file potenzialmente necessari in
un'unica posizione nel file system attivo semplifica la gestione. Non c'è
motivo di cercare tra gli snapshot per individuare un file che deve essere
ripristinato.
```

Il `snapshot-only` i criteri potrebbero funzionare, ma si applicano solo ai blocchi che non si trovano più nel file system attivo. Pertanto, i file presenti in una condivisione NFS o SMB devono essere eliminati prima del tiering dei dati.

Questa policy risulterebbe ancora meno efficiente con la configurazione LUN, poiché l'eliminazione di un file da una LUN rimuove solo i riferimenti dei file dai metadati del file system. I blocchi effettivi sui LUN restano in posizione fino a quando non vengono sovrascritti. Questa situazione può creare un lungo ritardo tra il tempo di eliminazione di un file e il tempo in cui i blocchi vengono sovrascritti e candidati per il tiering. Lo spostamento dell' comporta alcuni vantaggi `snapshot-only` Dei blocchi nel Tier di capacità, ma, nel complesso, la gestione FabricPool dei dati di backup funziona meglio con l' `auto` policy.

Questo approccio aiuta gli utenti a gestire lo spazio richiesto per i backup in modo più efficiente, ma FabricPool non è una tecnologia di backup. Il tiering dei file di backup nell'archivio di oggetti semplifica la gestione perché i file sono ancora visibili nel sistema di storage originale, ma i blocchi di dati nella destinazione dell'archivio di oggetti dipendono dal sistema di storage originale. Se il volume di origine viene perso, i dati dell'archivio di oggetti non sono più utilizzabili.

Interruzioni di accesso agli archivi di oggetti

Il tiering di un set di dati con FabricPool determina una dipendenza tra lo storage array primario e il Tier dell'archivio di oggetti. Sono disponibili molte opzioni di storage a oggetti che offrono livelli di disponibilità variabili. È importante comprendere l'impatto di una possibile perdita di connettività tra lo storage array primario e il Tier dello storage a oggetti.

Se un i/o emesso a ONTAP richiede dati dal Tier di capacità e ONTAP non riesce a raggiungere il Tier di capacità per recuperare i blocchi, l'i/o finisce il time-out. L'effetto di questo timeout dipende dal protocollo utilizzato. In un ambiente NFS, ONTAP risponde con una risposta EJUKEBOX o EDELAY, a seconda del protocollo. Alcuni sistemi operativi meno recenti potrebbero interpretare questo come un errore, ma i sistemi operativi attuali e i livelli di patch correnti del client Oracle Direct NFS considerano questo come un errore recuperabile e continuano ad attendere il completamento dell'i/O.

Un timeout più breve si applica agli ambienti SAN. Se un blocco nell'ambiente dell'archivio oggetti è necessario e rimane irraggiungibile per due minuti, viene restituito un errore di lettura all'host. Il volume e i LUN di ONTAP rimangono online, ma il sistema operativo host potrebbe segnalare il file system come in uno stato di errore.

Problemi di connettività dello storage a oggetti `snapshot-only` i criteri sono meno preoccupanti, perché vengono suddivisi in livelli solo i dati di backup. I problemi di comunicazione rallenterebbero il recupero dei dati, ma non influenzerebbero altrimenti l'utilizzo attivo dei dati. Il `auto` e. `all` Le policy consentono il tiering dei dati cold dal LUN attivo, il che significa che un errore durante il recupero dei dati dell'archivio oggetti può influire sulla disponibilità del database. Un'implementazione SAN con queste policy deve essere utilizzata solo con storage a oggetti di classe Enterprise e connessioni di rete progettate per l'alta disponibilità. NetApp StorageGRID è l'opzione superiore.

Data Protection Oracle

Data Protection con ONTAP

NetApp sa che i dati più mission-critical sono presenti nei database.

Un'azienda non può operare senza accesso ai propri dati, e a volte i dati definiscono l'azienda. Questi dati devono essere protetti; tuttavia, la protezione dei dati non è solo garanzia di un backup utilizzabile, ma consiste nell'eseguire i backup in modo rapido e affidabile, oltre a memorizzarli in modo sicuro.

L'altro lato della protezione dei dati è la recovery. Quando i dati sono inaccessibili, l'azienda ne è interessata e potrebbe non funzionare fino a quando i dati non vengono ripristinati. Questo processo deve essere rapido e affidabile. Infine, la maggior parte dei database deve essere protetta dai disastri, il che significa mantenere una replica del database. La replica deve essere sufficientemente aggiornata. Rendere la replica un database completamente operativo deve anche essere semplice e veloce.

Questa documentazione sostituisce il report tecnico precedentemente pubblicato *TR-4591: Data Protection di Oracle: Backup, recovery e replica.*

Pianificazione

La corretta architettura di protezione dei dati aziendali dipende dai requisiti di business correlati a conservazione dei dati, ripristinabilità e tolleranza per le interruzioni durante i vari eventi.

Ad esempio, consideriamo il numero di applicazioni, database e set di dati importanti inclusi nell'ambito della fornitura. La costruzione di una strategia di backup per un singolo set di dati che garantisca la conformità con gli SLA tipici è piuttosto semplice, perché non ci sono molti oggetti da gestire. Con l'aumento del numero di set di dati, il monitoraggio diventa più complicato e gli amministratori potrebbero essere costretti a spendere una crescente quantità di tempo nella risoluzione degli errori di backup. Quando un ambiente raggiunge il cloud e scala un service provider, è necessario un approccio completamente diverso.

Anche le dimensioni del set di dati influiscono sulla strategia. Ad esempio, esistono molte opzioni per il backup e il ripristino con un database da 100GB TB perché il set di dati è così piccolo. La semplice copia dei dati dai supporti di backup con gli strumenti tradizionali in genere offre un RTO sufficiente per il recovery. Un database 100TB ha normalmente bisogno di una strategia completamente diversa, a meno che l'RTO non consenta un'interruzione di più giorni, nel qual caso una tradizionale procedura di backup e ripristino basata sulla copia potrebbe essere accettabile.

Infine, vi sono alcuni fattori che esulano dal processo di backup e ripristino stesso. Ad esempio, esistono database che supportano attività di produzione critiche e che rendono il ripristino una rara eventualità che viene eseguita solo da DBA esperti? In alternativa, i database fanno parte di un grande ambiente di sviluppo in cui il ripristino è un evento frequente e gestito da un team IT generico?

RTO, RPO e pianificazione SLA

ONTAP ti consente di personalizzare facilmente una strategia di protezione dei dati dei database di Oracle in base ai tuoi requisiti di business.

Questi requisiti includono fattori quali la velocità del recovery, la perdita massima consentita di dati e le esigenze di conservazione del backup. Il piano di protezione dei dati deve anche tenere in considerazione i vari requisiti normativi per la conservazione e il ripristino dei dati. Infine, è necessario considerare diversi scenari di ripristino dei dati, che vanno dal recupero tipico e prevedibile derivante da errori di utenti o applicazioni fino a scenari di ripristino di emergenza che includono la perdita completa di un sito.

Piccole modifiche alle policy di protezione e ripristino dei dati possono avere un effetto significativo sull'architettura generale dello storage, del backup e del ripristino. È fondamentale definire e documentare gli standard prima di iniziare il lavoro di progettazione, per evitare di complicare un'architettura di protezione dati. Le funzioni o i livelli di protezione non necessari comportano costi e costi di gestione inutili, mentre un requisito inizialmente trascurato può condurre un progetto nella direzione sbagliata o richiedere modifiche di progettazione dell'ultimo minuto.

Recovery time objective

L'obiettivo RTO (Recovery Time Objective) definisce il tempo massimo consentito per il ripristino di un servizio. Ad esempio, un database di risorse umane potrebbe avere un RTO di 24 ore perché, sebbene sarebbe molto scomodo perdere l'accesso a questi dati durante la giornata lavorativa, l'azienda può comunque operare. Al contrario, un database che supporta la contabilità generale di una banca avrebbe un RTO misurato in minuti o anche secondi. Un RTO di zero non è possibile, perché deve esserci un modo per distinguere tra un'effettiva interruzione del servizio e un evento di routine, come un pacchetto di rete perso. Tuttavia, un RTO prossimo

allo zero è un requisito tipico.

Obiettivo RPO

Il recovery point objective (RPO) definisce la massima perdita di dati tollerabile. In molti casi, l'RPO è determinato unicamente dalla frequenza delle snapshot o degli aggiornamenti di snapmirror.

In alcuni casi, l'RPO può essere reso più aggressivo proteggendo determinati dati con maggiore frequenza. In un contesto di database, l'RPO è in genere una questione di quanti dati di registro possono essere persi in una situazione specifica. In uno scenario di ripristino tipico in cui un database viene danneggiato a causa di un bug del prodotto o di un errore dell'utente, l'RPO deve essere pari a zero, il che significa che non ci devono essere perdite di dati. La procedura di ripristino prevede il ripristino di una copia precedente dei file di database e la riproduzione dei file di registro per portare lo stato del database al momento desiderato. I file di registro necessari per questa operazione dovrebbero essere già presenti nella posizione originale.

In scenari insoliti, i dati del registro potrebbero andare persi. Ad esempio, un accidentale o dannoso `rm -rf *` di file di database potrebbe causare l'eliminazione di tutti i dati. L'unica opzione sarebbe il ripristino dal backup, inclusi i file di registro, e alcuni dati andrebbero inevitabilmente persi. L'unica opzione per migliorare gli RPO in un ambiente di backup tradizionale sarebbe l'esecuzione di backup ripetuti dei dati di log. Questo comporta dei limiti, tuttavia, a causa dello spostamento costante dei dati e della difficoltà di mantenere un sistema di backup come servizio in esecuzione costante. Uno dei benefici dei sistemi storage avanzati è la capacità di proteggere i dati da danni accidentali o dannosi ai file e garantire quindi un RPO migliore senza spostamento dei dati.

Disaster recovery

Il ripristino di emergenza include l'architettura IT, i criteri e le procedure necessarie per il ripristino di un servizio in caso di emergenza fisica. Tra questi, inondazioni, incendi o persone che agiscono con intento doloso o negligente.

Il disaster recovery non è solo una serie di procedure di ripristino. Si tratta del processo completo di identificazione dei vari rischi, definizione dei requisiti di ripristino dei dati e continuità del servizio e realizzazione della giusta architettura con le relative procedure.

Durante la definizione dei requisiti di protezione dei dati, è fondamentale differenziare tra i requisiti tipici di RPO e RTO e quelli di RPO e RTO necessari per il disaster recovery. Alcuni ambienti applicativi richiedono un RPO pari a zero e un RTO prossimo allo zero per situazioni di perdita di dati che vanno da errori utente relativamente normali a incendi che distruggono un data center. Tuttavia, vi sono conseguenze amministrative e di costo per questi elevati livelli di protezione.

In generale, i requisiti di ripristino dei dati non di emergenza devono essere rigorosi per due motivi. Innanzitutto, i bug applicativi e gli errori degli utenti che danneggiano i dati sono prevedibili al punto che sono quasi inevitabili. In secondo luogo, non è difficile progettare una strategia di backup in grado di offrire un RPO pari a zero e un RTO basso finché il sistema storage non viene distrutto. Non c'è motivo di non affrontare un rischio significativo che sia facilmente risolvibile, motivo per cui gli obiettivi di RPO e RTO per il ripristino locale dovrebbero essere aggressivi.

I requisiti di RTO e RPO per il disaster recovery variano in modo più ampio in base alla probabilità di un disastro e alle conseguenze della perdita di dati associata o dell'interruzione di un business. I requisiti di RPO e RTO devono essere basati sulle effettive esigenze di business e non su principi generali. Devono tenere conto di più scenari di emergenza logici e fisici.

Disastri logici

I disastri logici includono la corruzione dei dati causata da utenti, bug delle applicazioni o del sistema operativo e malfunzionamenti del software. I disastri logici possono includere anche attacchi dannosi da parte di terzi

con virus o worm o sfruttando le vulnerabilità delle applicazioni. In questi casi, l'infrastruttura fisica rimane intatta, ma i dati sottostanti non sono più validi.

Un tipo sempre più comune di disastro logico è noto come ransomware, in cui un vettore di attacco viene utilizzato per crittografare i dati. La crittografia non danneggia i dati, ma li rende non disponibili fino a quando non viene effettuato il pagamento a terzi. Un numero sempre crescente di aziende è specificatamente preso di mira con gli hack ransomware. A causa di questa minaccia, NetApp offre snapshot a prova di manomissione, in cui nemmeno l'amministratore dello storage può modificare i dati protetti prima della data di scadenza configurata.

Disastri fisici

I disastri fisici includono l'errore di componenti di un'infrastruttura che superano le sue capacità di ridondanza e causano una perdita di dati o un'estesa perdita di servizio. Ad esempio, la protezione RAID fornisce la ridondanza dell'unità disco e l'utilizzo di HBA fornisce la ridondanza di porte FC e cavi FC. I guasti hardware di tali componenti sono prevedibili e non influiscono sulla disponibilità.

In un ambiente aziendale, è generalmente possibile proteggere l'infrastruttura di un intero sito con componenti ridondanti fino al punto in cui l'unico scenario di emergenza fisica prevedibile è la perdita completa del sito. Quindi, il piano di disaster recovery dipende dalla replica sito-sito.

Protezione dei dati sincrona e asincrona

In un mondo ideale, tutti i dati verrebbero replicati in modo sincrono tra siti dispersi geograficamente. Tale replicazione non è sempre fattibile o addirittura possibile per diversi motivi:

- La replica sincrona aumenta inevitabilmente la latenza di scrittura, perché tutte le modifiche devono essere replicate in entrambe le posizioni prima che l'applicazione/database possa procedere con l'elaborazione. L'effetto sulle prestazioni risultante è talvolta inaccettabile, escludendo l'uso del mirroring sincrono.
- La maggiore adozione di storage SSD al 100% implica maggiore probabilità di ottenere una latenza di scrittura aggiuntiva, poiché le aspettative di performance includono centinaia di migliaia di IOPS e latenza sotto al millisecondo. Ottenere tutti i benefici dell'utilizzo di SSD al 100% può richiedere la revisione della strategia di disaster recovery.
- I set di dati continuano a crescere in termini di byte, creando difficoltà per garantire una larghezza di banda sufficiente a sostenere la replica sincrona.
- I set di dati crescono anche in termini di complessità, creando problemi con la gestione della replica sincrona su larga scala.
- Le strategie basate sul cloud spesso implicano maggiori distanze di replica e latenza, precludendo ulteriormente l'utilizzo di mirroring sincrono.

NetApp offre soluzioni che includono replica sincrona per le più esigenti richieste di recovery di dati e soluzioni asincrone che consentono performance e flessibilità migliori. Inoltre, la tecnologia NetApp si integra perfettamente con molte soluzioni di replica di terze parti, come Oracle DataGuard

Tempo di conservazione

L'ultimo aspetto di una strategia di protezione dei dati è il tempo di conservazione dei dati, che può variare drasticamente.

- Un requisito tipico è rappresentato da 14 giorni di backup notturni sul sito primario e 90 giorni di backup memorizzati su un sito secondario.
- Molti clienti creano archivi trimestrali autonomi archiviati su supporti diversi.

- Un database costantemente aggiornato potrebbe non richiedere i dati storici e i backup devono essere conservati solo per alcuni giorni.
- I requisiti normativi potrebbero richiedere la possibilità di recupero fino al punto in cui avviene una transazione arbitraria nell'arco di 365 giorni.

Disponibilità del database

ONTAP è progettato per garantire la massima disponibilità dei database Oracle. Una descrizione completa delle funzioni di alta disponibilità di ONTAP esula dall'ambito di questo documento. Tuttavia, come per la protezione dei dati, una conoscenza di base di questa funzionalità è importante quando si progetta un'infrastruttura di database.

Coppie HA

L'unità di base dell'alta disponibilità è la coppia ha. Ciascuna coppia contiene collegamenti ridondanti per supportare la replica dei dati nella NVRAM. NVRAM non è una cache di scrittura. La RAM all'interno del controller funge da cache di scrittura. Lo scopo della NVRAM è quello di memorizzare temporaneamente i dati come salvaguardia da errori di sistema imprevisti. A questo proposito, è simile a un log di ripristino del database.

Sia la NVRAM che il redo log del database consentono di memorizzare i dati rapidamente, consentendo il commit delle modifiche ai dati il più rapidamente possibile. L'aggiornamento ai dati persistenti sulle unità (o file di dati) viene eseguito solo in un secondo momento durante un processo chiamato checkpoint sulle piattaforme ONTAP e sulla maggior parte dei database. Durante le normali operazioni, non vengono letti i dati della NVRAM né i log di ripristino del database.

Se un controller si guasta bruscamente, è probabile che vi siano modifiche in sospeso memorizzate nella NVRAM che non sono ancora state scritte sulle unità. Il partner controller rileva il guasto, assume il controllo dei dischi e applica le modifiche richieste che sono state memorizzate nella NVRAM.

Takeover e giveback

Il takeover e il giveback fanno riferimento al processo di trasferimento della responsabilità delle risorse di storage fra i nodi di una coppia ha. L'acquisizione e il giveback presentano due aspetti:

- Gestione della connettività di rete che consente l'accesso alle unità
- Gestione delle unità stesse

Le interfacce di rete che supportano il traffico CIFS e NFS sono configurate sia con una posizione home che di failover. Un takeover include lo spostamento delle interfacce di rete nella loro abitazione temporanea su un'interfaccia fisica situata sulla stessa subnet della posizione originale. Un giveback prevede lo spostamento delle interfacce di rete nelle posizioni originali. Il comportamento esatto può essere regolato come richiesto.

Le interfacce di rete che supportano i protocolli a blocchi SAN, come iSCSI e FC, non vengono ricollocate durante il takeover e lo giveback. È invece necessario eseguire il provisioning delle LUN attraverso percorsi che includano una coppia ha completa che si traduce in un percorso primario e un percorso secondario.

È possibile configurare anche percorsi aggiuntivi per controller aggiuntivi in modo da supportare la riallocazione dei dati tra i nodi di un cluster più grande, non facente parte del processo di ha.

Il secondo aspetto del takeover e dello sconto è il trasferimento della proprietà del disco. Il processo esatto dipende da diversi fattori, tra cui il motivo del takeover/giveback e le opzioni della riga di comando emesse.

L'obiettivo è quello di eseguire l'operazione nel modo più efficiente possibile. Anche se il processo complessivo potrebbe richiedere diversi minuti, il momento effettivo in cui la proprietà dell'unità viene trasferita da nodo a nodo può generalmente essere misurato in secondi.

Tempo di takeover

L'i/o dell'host subisce una breve pausa in i/o durante le operazioni di takeover e giveback, senza tuttavia alcuna interruzione dell'applicazione in un ambiente configurato correttamente. L'effettivo processo di transizione in cui l'i/o subisce un ritardo viene generalmente misurato in secondi, ma l'host potrebbe richiedere tempo aggiuntivo per riconoscere la modifica nei percorsi di dati e inviare di nuovo le operazioni i/O.

La natura dell'interruzione dipende dal protocollo:

- Un'interfaccia di rete che supporta il traffico NFS e CIFS emette una richiesta ARP (Address Resolution Protocol) alla rete dopo la transizione a una nuova posizione fisica. Ciò fa sì che gli switch di rete aggiornino le tabelle degli indirizzi MAC (Media Access Control) e riprendano l'elaborazione i/O. Le interruzioni nel caso di takeover e giveback pianificati vengono di solito misurate in secondi e in molti casi non sono rilevabili. Alcune reti potrebbero essere più lente a riconoscere completamente la modifica del percorso di rete e alcuni sistemi operativi potrebbero mettere in coda molti i/o in un breve periodo di tempo che deve essere rieseguito. Ciò può estendere il tempo necessario per riprendere l'i/O.
- Un'interfaccia di rete che supporta i protocolli SAN non passa a una nuova posizione. Un sistema operativo host deve modificare il percorso o i percorsi in uso. La pausa in i/o osservata dall'host dipende da diversi fattori. Dal punto di vista del sistema storage, il periodo in cui non è possibile fornire i/o è di pochi secondi. Tuttavia, sistemi operativi host diversi potrebbero richiedere tempo aggiuntivo per consentire un timeout i/o prima di riprovare. I sistemi operativi più recenti sono in grado di riconoscere un cambiamento di percorso molto più rapidamente, ma i sistemi operativi più vecchi in genere richiedono fino a 30 secondi per riconoscere un cambiamento.

La seguente tabella illustra i tempi di takeover previsti durante i quali il sistema storage non può fornire i dati a un ambiente applicativo. Non dovrebbero esserci errori in alcun ambiente applicativo, il takeover dovrebbe invece apparire come una breve pausa nell'elaborazione io.

	NFS	AFF	ASA
Takeover pianificato	15 sec.	6-10 sec.	2-3 sec.
Takeover non pianificato	30 sec.	6-10 sec.	2-3 sec.

Checksum e integrità dei dati

ONTAP e i protocolli supportati includono svariate funzionalità che proteggono l'integrità del database Oracle, inclusi dati a riposo e dati trasmessi sulla rete.

La protezione dei dati logici all'interno di ONTAP è costituita da tre requisiti principali:

- I dati devono essere protetti dalla corruzione.
- I dati devono essere protetti da guasti al disco.
- Le modifiche ai dati devono essere protette dalla perdita.

Queste tre esigenze sono discusse nelle sezioni seguenti.

Corruzione della rete: Checksum

Il livello più basilare di protezione dei dati è il checksum, che è uno speciale codice di rilevamento degli errori memorizzato insieme ai dati. La corruzione dei dati durante la trasmissione di rete viene rilevata con l'utilizzo di un checksum e, in alcuni casi, di checksum multipli.

Ad esempio, un frame FC include una forma di checksum chiamata CRC (Cyclic Redundancy Check) per assicurarsi che il payload non sia corrotto durante il transito. Il trasmettitore invia sia i dati che il CRC dei dati. Il ricevitore di un frame FC ricalcola il CRC dei dati ricevuti per assicurarsi che corrisponda al CRC trasmesso. Se il CRC appena calcolato non corrisponde al CRC collegato al frame, i dati sono corrotti e il frame FC viene scartato o rifiutato. Un'operazione i/o iSCSI include checksum ai livelli TCP/IP ed Ethernet e, per una maggiore protezione, può anche includere la protezione CRC opzionale al livello SCSI. Qualsiasi corruzione di bit sul filo viene rilevata dal livello TCP o IP, che porta alla ritrasmissione del pacchetto. Come nel caso di FC, gli errori nel CRC SCSI determinano un'eliminazione o un rifiuto dell'operazione.

Corruzione dei dischi: Checksum

I checksum vengono utilizzati anche per verificare l'integrità dei dati memorizzati sui dischi. I blocchi di dati scritti sui dischi vengono memorizzati con una funzione di checksum che produce un numero imprevedibile e legato ai dati originali. Quando i dati vengono letti dall'unità, il checksum viene ricalcolato e confrontato con il checksum memorizzato. Se non corrisponde, i dati sono corrotti e devono essere recuperati dal livello RAID.

Corruzione dei dati: Scritture perse

Uno dei tipi più difficili di corruzione da rilevare è una scrittura persa o posizionata erroneamente. Quando una scrittura viene confermata, deve essere scritta sul supporto nella posizione corretta. La corruzione dei dati sul posto è relativamente semplice da rilevare utilizzando un semplice checksum memorizzato con i dati. Tuttavia, se la scrittura viene semplicemente persa, la versione precedente dei dati potrebbe ancora esistere e il checksum sarebbe corretto. Se la scrittura viene posizionata nella posizione fisica errata, il checksum associato sarebbe ancora una volta valido per i dati memorizzati, anche se la scrittura ha distrutto altri dati.

La soluzione a questa sfida è la seguente:

- Un'operazione di scrittura deve includere metadati che indicano la posizione in cui dovrebbe essere trovata la scrittura.
- Un'operazione di scrittura deve includere un tipo di identificatore di versione.

Quando ONTAP scrive un blocco, include i dati sulla posizione di appartenenza del blocco. Se una lettura successiva identifica un blocco, ma i metadati indicano che esso appartiene alla posizione 123 quando è stato trovato nella posizione 456, allora la scrittura è stata erroneamente posizionata.

Rilevare una scrittura totalmente persa è più difficile. La spiegazione è molto complicata, ma essenzialmente ONTAP memorizza i metadati in modo che un'operazione di scrittura determini aggiornamenti a due posizioni diverse sulle unità. Se una scrittura viene persa, una successiva lettura dei dati e dei metadati associati mostra due diverse identità di versione. Ciò indica che la scrittura non è stata completata dall'unità.

La corruzione in scrittura persa e posizionata erroneamente è estremamente rara, ma con il continuo aumento dei dischi e la diminuzione dei set di dati nella scala di exabyte, il rischio aumenta. Il rilevamento delle operazioni di scrittura perse deve essere incluso in qualsiasi sistema storage che supporti i carichi di lavoro del database.

Guasti del disco: RAID, RAID DP e RAID-TEC

Se un blocco di dati su un'unità viene rilevato come danneggiato o se l'intera unità si guasta e non è completamente disponibile, i dati devono essere ricostituiti. Questo viene fatto in ONTAP utilizzando unità di

parità. Lo striping dei dati viene eseguito su più unità dati, quindi vengono generati i dati di parità. I dati vengono memorizzati separatamente dai dati originali.

ONTAP utilizzava in origine RAID 4, che utilizza un singolo disco di parità per ciascun gruppo di unità dati. Il risultato è che un'unità del gruppo potrebbe guastarsi senza causare una perdita di dati. Se l'unità di parità non funziona correttamente, non sono stati danneggiati dati ed è stato possibile costruire una nuova unità di parità. Se si è verificato un errore in una singola unità dati, è possibile utilizzare le unità rimanenti con l'unità di parità per rigenerare i dati mancanti.

Quando le unità erano di piccole dimensioni, la possibilità statistica di due unità che si guastavano contemporaneamente era trascurabile. Con la progressiva crescita della capacità del disco aumentano anche il tempo necessario per ricostruire i dati in seguito a un guasto al disco. Ciò ha aumentato la finestra in cui un guasto di una seconda unità causerebbe la perdita di dati. Inoltre, il processo di ricostruzione crea numerosi i/o aggiuntivi sui dischi ancora in uso. Man mano che i dischi diventano obsoleti, aumenta anche il rischio di carico aggiuntivo che potrebbe causare un guasto al secondo disco. Infine, anche se il rischio di perdita di dati non aumentasse con il continuo utilizzo di RAID 4, le conseguenze della perdita di dati diventerebbero più gravi. Maggiore è la quantità di dati che andrebbero persi in caso di guasto a un gruppo RAID, più tempo occorrerebbe per ripristinare i dati, prolungando l'interruzione del business.

Questi problemi hanno portato NetApp a sviluppare la tecnologia NetApp RAID DP, una variante di RAID 6. Questa soluzione include due unità di parità, il che significa che due unità in un gruppo RAID possono guastarsi senza creare perdite di dati. Le dimensioni dei dischi hanno continuato a crescere, portando infine NetApp a sviluppare la tecnologia NetApp RAID-TEC, che introduce un disco a terza parità.

Alcune procedure consigliate per i database storici consigliano l'uso di RAID-10, noto anche come mirroring con striping. Ciò offre una protezione dei dati inferiore rispetto a quella dei sistemi RAID DP, in quanto vi sono più scenari di guasto a due dischi, mentre in RAID DP non ve ne sono nessuno.

Esistono inoltre alcune procedure consigliate per i database storici che indicano che le opzioni RAID-10 sono preferite a quelle RAID-4/5/6 a causa di problemi di prestazioni. Queste raccomandazioni a volte fanno riferimento a una penalizzazione RAID. Sebbene queste raccomandazioni siano generalmente corrette, non sono applicabili alle implementazioni di RAID all'interno di ONTAP. Il problema di prestazioni è relativo alla rigenerazione di parità. Con le implementazioni RAID tradizionali, l'elaborazione delle random write di routine eseguite da un database richiede letture multiple del disco per rigenerare i dati di parità e completare la scrittura. La penalità viene definita come gli IOPS in lettura aggiuntivi necessari per eseguire le operazioni di scrittura.

ONTAP non subisce alcuna penalizzazione RAID perché le scritture vengono organizzate in memoria dove la parità viene generata e quindi scritta su disco come singolo stripe RAID. Non sono richieste letture per completare l'operazione di scrittura.

In sintesi, rispetto al RAID 10, RAID DP e RAID-TEC offrono una capacità utilizzabile molto maggiore, una migliore protezione contro i guasti ai dischi e nessun compromesso in termini di performance.

Protezione da errori hardware: NVRAM

Qualsiasi storage array che gestisce un carico di lavoro del database deve eseguire le operazioni di scrittura il più rapidamente possibile. Inoltre, un'operazione di scrittura deve essere protetta dalla perdita da un evento imprevisto, come un'interruzione dell'alimentazione. Ciò significa che qualsiasi operazione di scrittura deve essere conservata in modo sicuro in almeno due posizioni.

I sistemi AFF e FAS si affidano alla NVRAM per soddisfare questi requisiti. Il processo di scrittura funziona come segue:

1. I dati di scrittura in entrata sono memorizzati nella RAM.

2. Le modifiche che devono essere apportate ai dati sul disco vengono registrate nella NVRAM sia sul nodo locale che sul nodo partner. NVRAM non è una cache di scrittura, ma un journal simile a un log di ripristino dei database. In condizioni normali, non viene letta. Viene utilizzata solo per il ripristino, ad esempio in seguito a un'interruzione dell'alimentazione durante l'elaborazione i/O.
3. La scrittura viene quindi riconosciuta all'host.

Il processo di scrittura in questa fase è completo dal punto di vista dell'applicazione e i dati sono protetti dalla perdita, perché vengono memorizzati in due posizioni diverse. Alla fine, le modifiche vengono scritte su disco, ma il processo risulta fuori banda dal punto di vista dell'applicazione perché si verifica dopo il riconoscimento della scrittura e quindi non influisce sulla latenza. Questo processo è ancora una volta simile alla registrazione del database. Una modifica al database viene registrata nei registri di ripristino il più rapidamente possibile e la modifica viene quindi riconosciuta come confermata. Gli aggiornamenti ai file di dati avvengono molto più tardi e non influenzano direttamente la velocità di elaborazione.

In caso di guasto a un controller, il partner controller assume la proprietà dei dischi richiesti e riproduce i dati registrati nella NVRAM per ripristinare le operazioni di i/o in corso quando si è verificato il guasto.

Protezione da errori hardware: NVFAIL

Come discusso in precedenza, una scrittura non viene riconosciuta fino a quando non è stata registrata nella NVRAM locale e nella NVRAM su almeno un altro controller. Questo approccio garantisce che un guasto dell'hardware o un'interruzione di corrente non comporti la perdita dell'i/o in-flight In caso di guasto della NVRAM locale o di guasto della connettività al partner di ha, i dati in-flight non verranno più mirrorati.

Se la NVRAM locale riporta un errore, il nodo si arresta. Questo arresto determina il failover su un controller partner ha. Nessun dato viene perso perché il controller che presenta il guasto non ha confermato l'operazione di scrittura.

ONTAP non consente un failover quando i dati non sono sincronizzati, a meno che il failover non sia forzato. La forzatura di una modifica delle condizioni in questo modo riconosce che i dati potrebbero essere lasciati indietro nel controllore originale e che la perdita di dati è accettabile.

I database sono particolarmente vulnerabili al danneggiamento se un failover viene forzato perché mantengono grandi cache interne di dati su disco. In caso di failover forzato, le modifiche precedentemente riconosciute vengono effettivamente eliminate. Il contenuto dell'array di storage torna indietro nel tempo e lo stato della cache del database non riflette più lo stato dei dati su disco.

Per proteggere i dati da questa situazione, ONTAP consente di configurare i volumi per una protezione speciale contro gli errori della NVRAM. Quando attivato, questo meccanismo di protezione determina l'ingresso di un volume nello stato chiamato NVFAIL. Questo stato causa errori di i/o che causano l'arresto di un'applicazione in modo che non utilizzino dati obsoleti. I dati non devono essere persi perché qualsiasi scrittura riconosciuta deve essere presente sull'array di storage.

Solitamente, gli amministratori dovranno arrestare completamente gli host prima di riportare manualmente LUN e volumi in linea. Sebbene queste fasi possano comportare un certo lavoro, questo approccio è il modo più sicuro per garantire l'integrità dei dati. Non tutti i dati richiedono questa protezione, motivo per cui il comportamento di NVFAIL può essere configurato in base al volume.

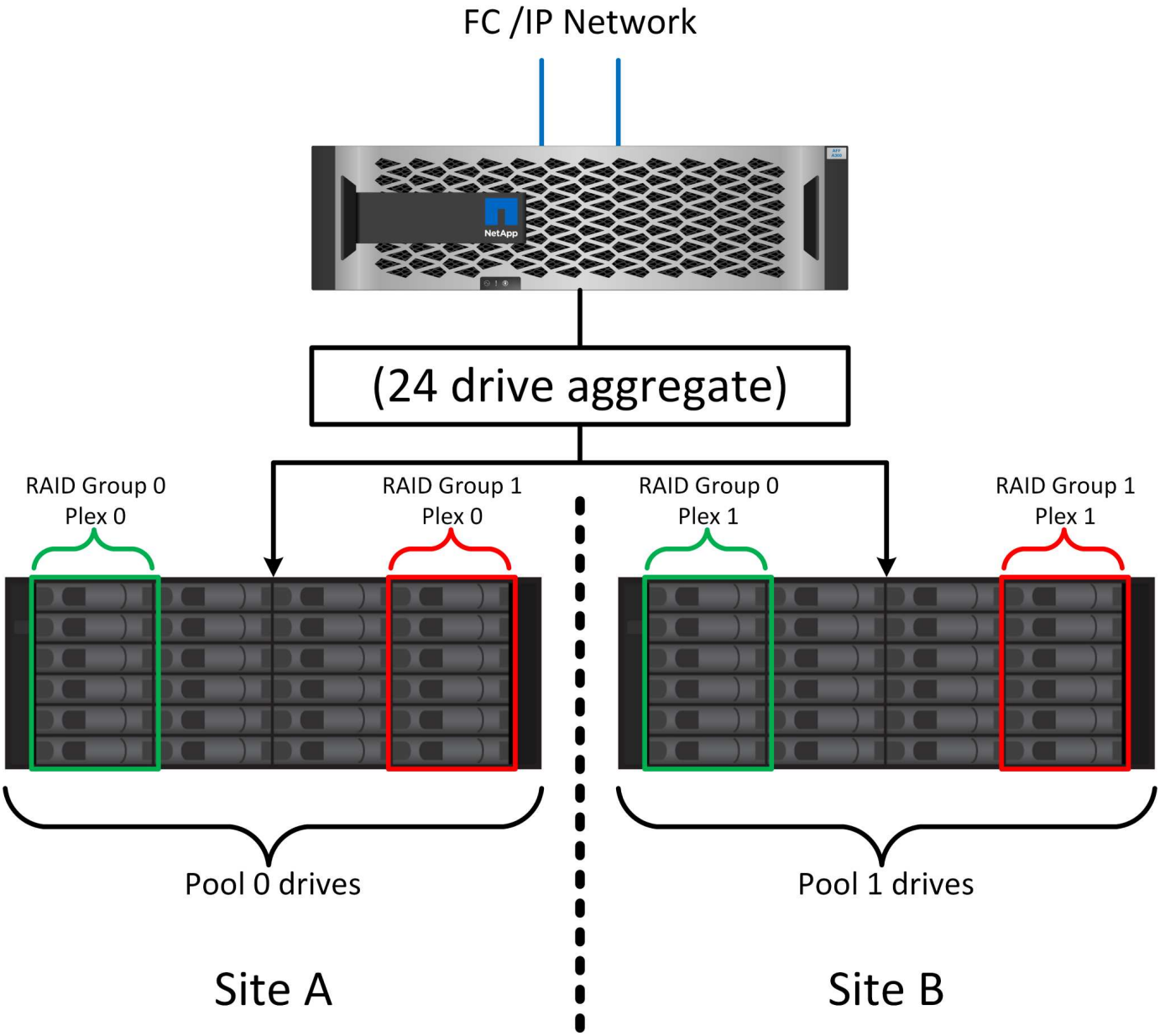
Protezione dai guasti di shelf e siti: SyncMirror e plessi

SyncMirror è una tecnologia di mirroring che migliora, ma non sostituisce, RAID DP o RAID-TEC. Esegue il mirroring del contenuto di due gruppi RAID indipendenti. La configurazione logica è la seguente:

- I dischi sono configurati in due pool in base alla posizione. Un pool è composto da tutti i dischi sul sito A,

mentre il secondo è composto da tutti i dischi sul sito B.

- Viene quindi creato un pool di storage comune, detto aggregato, in base a set di gruppi RAID con mirroring. Viene ottenuto lo stesso numero di unità per ciascun sito. Ad esempio, un aggregato SyncMirror da 20 dischi sarebbe composto da 10 dischi del sito A e 10 dischi del sito B.
- Ogni set di unità su un dato sito viene configurato automaticamente come uno o più gruppi RAID-DP o RAID-TEC completamente ridondanti, indipendentemente dall'utilizzo del mirroring. In questo modo si garantisce una protezione dei dati continua, anche dopo la perdita di un sito.



La figura precedente illustra una configurazione SyncMirror di esempio. È stato creato un aggregato di 24 dischi sul controller con 12 dischi da uno shelf allocato sul sito A e 12 dischi da uno shelf allocato sul sito B. I dischi sono stati raggruppati in due gruppi RAID con mirroring. Il gruppo RAID 0 include un plesso A 6 unità sul sito A con mirroring su un plesso A 6 unità sul sito B. Analogamente, il gruppo RAID 1 include un plesso A 6 unità sul sito A con mirroring su un plesso A 6 unità sul sito B.

Di norma, SyncMirror viene utilizzato per fornire il mirroring remoto con i sistemi MetroCluster, con una copia dei dati in ciascun sito. A volte, è stato utilizzato per fornire un livello di ridondanza extra in un unico sistema. In

particolare, fornisce ridondanza a livello di shelf. Uno shelf di dischi contiene già doppi controller e alimentatori e nel complesso è poco più di una lamiera, ma in alcuni casi è consigliabile garantire una protezione extra. Ad esempio, un cliente NetApp ha implementato SyncMirror per una piattaforma mobile di analytics in tempo reale utilizzata durante i test nel settore automobilistico. Il sistema è stato separato in due rack fisici forniti da alimentatori indipendenti da sistemi UPS indipendenti.

Checksum

L'argomento dei checksum è di particolare interesse per i DBA abituati all'utilizzo dei backup in streaming Oracle RMAN che migrano a backup basati su snapshot. Una caratteristica di RMAN è che esegue controlli di integrità durante le operazioni di backup. Sebbene questa funzionalità offra un certo valore, il suo vantaggio principale è quello di un database non utilizzato su uno storage array moderno. Quando si utilizzano dischi fisici per un database Oracle, è quasi certo che il danneggiamento si verifica anche in caso di invecchiamento dei dischi, un problema che viene risolto dai checksum basati su array negli storage array reali.

Con un vero storage array, l'integrità dei dati è protetta utilizzando checksum a livelli multipli. Se i dati sono corrotti in una rete basata su IP, il livello TCP (Transmission Control Protocol) rifiuta i dati a pacchetto e richiede la ritrasmissione. Il protocollo FC include i checksum, così come i dati SCSI incapsulati. Dopo essere stato inserito nell'array, ONTAP dispone della protezione RAID e checksum. Il danneggiamento può verificarsi, ma, come nella maggior parte degli array Enterprise, viene rilevato e corretto. In genere, si verifica un guasto di un intero disco, che richiede una ricostruzione RAID e l'integrità del database rimane inalterata. È ancora possibile che i singoli byte su un'unità siano danneggiati dalla radiazione cosmica o da celle flash difettose. In questo caso, il controllo della parità non viene eseguito correttamente, l'unità viene chiusa in errore e viene avviata la ricostruzione RAID. Ancora una volta, l'integrità dei dati non viene influenzata. L'ultima linea di difesa è l'uso di checksum. Se, ad esempio, un errore catastrofico del firmware su un'unità ha danneggiato i dati in un modo che in qualche modo non è stato rilevato da un controllo di parità RAID, il checksum non corrisponderebbe e ONTAP impedirebbe il trasferimento di un blocco danneggiato prima che il database Oracle potesse riceverlo.

L'architettura dei log di ripristino e file dati di Oracle è inoltre progettata per offrire il massimo livello di integrità dei dati possibile, anche in circostanze estreme. A livello massimo, i blocchi Oracle includono il checksum e controlli logici di base con quasi ogni i/O. Se Oracle non è in crash o non ha portato offline uno spazio di tabella, i dati saranno intatti. Il grado di controllo dell'integrità dei dati è regolabile e Oracle può anche essere configurato per confermare le operazioni di scrittura. Di conseguenza, è possibile ripristinare quasi tutti gli scenari di crash e di guasto e, nel caso estremamente raro di una situazione irreversibile, viene immediatamente rilevata la corruzione.

La maggior parte dei clienti NetApp che utilizzano database Oracle interrompe l'utilizzo di RMAN e di altri prodotti di backup dopo la migrazione a backup basati su snapshot. Esistono ancora opzioni in cui RMAN può essere utilizzato per eseguire un ripristino a livello di blocco con SnapCenter. Tuttavia, ogni giorno, RMAN, NetBackup e altri prodotti vengono utilizzati solo occasionalmente per creare copie di archivio mensili o trimestrali.

Alcuni clienti scelgono di eseguire `dbv` eseguire periodicamente controlli di integrità dei database esistenti. NetApp scoraggia questa pratica perché crea un carico i/o non necessario. Come illustrato in precedenza, se il database non presentava problemi, la possibilità di `dbv` Il rilevamento di un problema è prossimo allo zero e questa utility crea un carico i/o sequenziale molto elevato sulla rete e sul sistema di storage. A meno che non vi sia motivo di ritenere che esista una corruzione, come l'esposizione a un bug Oracle noto, non c'è motivo di eseguire `dbv`.

Elementi di base di backup e recovery

Backup basati su snapshot

La base della protezione dei dati dei database Oracle su ONTAP è la tecnologia Snapshot di NetApp.

I valori chiave sono i seguenti:

- **Semplicità.** Uno snapshot è una copia di sola lettura del contenuto di un contenitore di dati in un determinato momento.
- **Efficienza.** le istantanee non richiedono spazio al momento della creazione. Lo spazio viene occupato solo quando i dati vengono modificati.
- **Gestibilità.** Una strategia di backup basata sugli snapshot è facile da configurare e gestire perché gli snapshot sono parte nativa del sistema operativo di storage. Se il sistema di archiviazione è acceso, è pronto per creare dei backup.
- **Scalabilità.** è possibile conservare fino a 1024 backup di un singolo contenitore di file e LUN. Per set di dati complessi, più container di dati possono essere protetti da un singolo set coerente di snapshot.
- Le prestazioni non sono influenzate, indipendentemente dal fatto che un volume contenga 1024 snapshot o nessuno.

Sebbene molti vendor di soluzioni storage offrano la tecnologia Snapshot, la tecnologia Snapshot all'interno di ONTAP è unica e offre benefici significativi per gli ambienti applicativi aziendali e di database:

- Le copie snapshot fanno parte del layout file WAFL (Write-Anywhere file Layout) sottostante. Non si tratta di una tecnologia aggiuntiva o esterna. Questo semplifica la gestione, perché il sistema storage è il sistema di backup.
- Le copie Snapshot non influiscono sulle prestazioni, ad eccezione di alcuni casi edge, come ad esempio quando una quantità così elevata di dati viene memorizzata nelle snapshot che il sistema storage sottostante si riempie.
- Il termine "gruppo di coerenza" viene spesso utilizzato per fare riferimento a un raggruppamento di oggetti di storage che vengono gestiti come una raccolta coerente di dati. Uno snapshot di un particolare volume ONTAP costituisce il backup del gruppo di coerenza.

Le snapshot ONTAP offrono anche una scalabilità migliore rispetto alle tecnologie della concorrenza. I clienti possono memorizzare 5, 50 o 500 snapshot senza influire sulle performance. Il numero massimo di snapshot attualmente consentiti in un volume è 1024. Se è necessaria una conservazione aggiuntiva degli snapshot, sono disponibili opzioni per trasferire gli snapshot in cascata ad altri volumi.

Di conseguenza, la protezione di un set di dati ospitato su ONTAP è semplice e altamente scalabile. I backup non richiedono lo spostamento dei dati, pertanto una strategia di backup può essere personalizzata in base alle esigenze dell'azienda piuttosto che alle limitazioni delle velocità di trasferimento di rete, del numero elevato di unità a nastro o delle aree di staging del disco.

Uno snapshot è un backup?

Una domanda comunemente posta sull'utilizzo delle istantanee come strategia di protezione dei dati è il fatto che i dati "reali" e i dati snapshot si trovano sulle stesse unità. La perdita di tali unità causerebbe la perdita sia dei dati primari che del backup.

Si tratta di un problema valido. Le snapshot locali vengono utilizzate per le esigenze di backup e ripristino quotidiane, e in questo senso la snapshot è un backup. Quasi il 99% di tutti gli scenari di ripristino in ambienti NetApp si affida alle snapshot per soddisfare anche i requisiti RTO più aggressivi.

Gli snapshot locali, tuttavia, non dovrebbero mai rappresentare l'unica strategia di backup, motivo per cui NetApp offre tecnologie come SnapMirror e la replica SnapVault per replicare in modo rapido ed efficiente le snapshot su un set indipendente di dischi. In una soluzione adeguatamente progettata con istantanee e replica snapshot, l'utilizzo del nastro può essere ridotto a icona in un archivio trimestrale o eliminato del tutto.

Backup basati su snapshot

Le copie Snapshot di ONTAP sono disponibili diverse opzioni per la protezione dei dati, mentre le snapshot sono alla base di molte altre funzionalità di ONTAP, tra cui replica, disaster recovery e cloning. Una descrizione completa della tecnologia snapshot non rientra nell'ambito di questo documento, ma le sezioni seguenti forniscono una panoramica generale.

Esistono due approcci principali per creare uno snapshot di un dataset:

- Backup coerenti con il crash
- Backup coerenti con le applicazioni

Un backup coerente con i crash di un set di dati si riferisce all'acquisizione dell'intera struttura di set di dati in un singolo point-in-time. Se il set di dati è memorizzato in un singolo volume, il processo è semplice ed è possibile creare una Snapshot in qualsiasi momento. Se un set di dati si estende tra i volumi, è necessario creare uno snapshot del gruppo di coerenza (CG). Esistono diverse opzioni per la creazione di snapshot CG, tra cui il software NetApp SnapCenter, le funzionalità native del gruppo di coerenza ONTAP e gli script gestiti dagli utenti.

I backup coerenti con i crash vengono utilizzati principalmente quando è sufficiente un ripristino point-of-the-backup. Quando è richiesto un ripristino più granulare, sono in genere necessari backup coerenti con l'applicazione.

La parola "coerente" in "coerente con l'applicazione" è spesso un nome scorretto. Ad esempio, l'inserimento di un database Oracle in modalità di backup viene definito backup coerente con l'applicazione, ma i dati non vengono resi coerenti o disattivati in alcun modo. I dati continuano a cambiare durante il backup. Al contrario, la maggior parte dei backup di MySQL e Microsoft SQL Server disattivano i dati prima di eseguire il backup. VMware può o non può rendere certi file coerenti.

Gruppi di coerenza

Il termine "gruppo di coerenza" si riferisce alla capacità di un array di archiviazione di gestire più risorse di archiviazione come una singola immagine. Ad esempio, un database può essere composto da 10 LUN. L'array deve essere in grado di eseguire il backup, il ripristino e la replica delle 10 LUN in modo coerente. Il ripristino non è possibile se le immagini dei LUN non erano coerenti nel punto di backup. La replica di queste 10 LUN richiede che tutte le repliche siano perfettamente sincronizzate l'una con l'altra.

Il termine "gruppo di coerenza" non viene spesso utilizzato quando si parla di ONTAP perché la coerenza è sempre stata una funzione di base dell'architettura di volumi e aggregati all'interno di ONTAP. Molti altri storage array gestiscono LUN o file system come unità singole. Possono quindi essere configurati facoltativamente come "gruppo di coerenza" ai fini della protezione dei dati, ma questo è un passaggio aggiuntivo nella configurazione.

ONTAP è sempre stata in grado di acquisire immagini di dati coerenti locali e replicate. Anche se i vari volumi su un sistema ONTAP non vengono in genere formalmente descritti come un gruppo di coerenza, è proprio questo lo sono. Una snapshot di tale volume è un'immagine del gruppo di coerenza, il ripristino di tale snapshot è un ripristino di un gruppo di coerenza e sia SnapMirror che SnapVault offrono la replica di un gruppo di coerenza.

Snapshot di gruppo di coerenza

Le snapshot di gruppo di coerenza (cg-Snapshot) sono un'estensione della tecnologia Snapshot di base di ONTAP. Un'operazione Snapshot standard crea un'immagine coerente di tutti i dati all'interno di un singolo volume, ma a volte è necessario creare un set coerente di Snapshot su più volumi e persino su sistemi di storage multipli. Ne risulta una serie di snapshot che possono essere utilizzate allo stesso modo di uno snapshot di un solo volume. Possono essere utilizzati per il recovery locale dei dati, replicati a scopo di disaster recovery o clonati come una singola unità coerente.

Il più grande utilizzo noto di cg-snapshot è per un ambiente di database di circa 1PB GB su 12 controller. Le cg-Snapshot create su questo sistema sono state utilizzate per il backup, il ripristino e il cloning.

Nella maggior parte dei casi, quando un set di dati copre i volumi e l'ordine di scrittura deve essere preservato, il software di gestione scelto utilizza automaticamente uno snapshot cg. In questi casi non è necessario comprendere i dettagli tecnici delle istantanee cg. Tuttavia, in alcune situazioni, i complessi requisiti di protezione dei dati richiedono un controllo dettagliato sul processo di protezione e replica dei dati. I flussi di lavoro di automazione o l'uso di script personalizzati per richiamare le API cg-snapshot sono alcune delle opzioni disponibili. La comprensione dell'opzione migliore e del ruolo di cg-snapshot richiede una spiegazione più dettagliata della tecnologia.

La creazione di una serie di istantanee cg è un processo in due fasi:

1. Stabilire il recencing in scrittura su tutti i volumi di destinazione.
2. Creare Snapshot di tali volumi nello stato fenced (fenced).

La recinzione in scrittura viene stabilita in serie. Ciò significa che, mentre il processo di scherma viene configurato su più volumi, l'i/o in scrittura viene bloccato sul primo volume della sequenza mentre continua ad essere assegnato ai volumi che compaiono in seguito. Questo potrebbe inizialmente sembrare una violazione del requisito per il mantenimento dell'ordine di scrittura, ma ciò si applica solo all'i/o emesso in modo asincrono sull'host e non dipende da altre scritture.

Ad esempio, un database potrebbe eseguire numerosi aggiornamenti asincroni del file dati, consentendo al sistema operativo di riordinare l'i/o e completarli in base alla propria configurazione dell'utilità di pianificazione. L'ordine di questo tipo di i/o non può essere garantito perché l'applicazione e il sistema operativo hanno già rilasciato il requisito di mantenere l'ordine di scrittura.

Come esempio di contatore, la maggior parte delle attività di registrazione del database è sincrona. Il database non procede con ulteriori scritture di registro fino a quando l'i/o non viene riconosciuto e l'ordine di tali scritture deve essere conservato. Se un i/o di registro arriva su un volume fenced, non viene riconosciuto e le applicazioni vengono bloccate in ulteriori scritture. Analogamente, l'i/o di metadati del file system è di solito sincrono. Ad esempio, un'operazione di eliminazione file non deve essere persa. Se un sistema operativo con un file system xfs eliminava un file e l'i/o che aggiornava i metadati del file system xfs per rimuovere il riferimento a quel file apposto su un volume recintato, l'attività del file system si interrompeva. Ciò garantisce l'integrità del file system durante le operazioni cg-snapshot.

Dopo aver configurato la funzionalità write fencing nei volumi di destinazione, sono pronti per la creazione di snapshot. Non è necessario creare esattamente gli snapshot contemporaneamente, perché lo stato dei volumi è bloccato da un punto di vista di scrittura dipendente. Per evitare un difetto nell'applicazione che crea le istantanee cg, la recinzione iniziale include un timeout configurabile in cui ONTAP rilascia automaticamente la recinzione e riprende l'elaborazione di scrittura dopo un numero definito di secondi. Se tutte le istantanee vengono create prima dello scadere del periodo di timeout, il gruppo risultante di istantanee è un gruppo di coerenza valido.

Ordine di scrittura dipendente

Da un punto di vista tecnico, la chiave per un gruppo di coerenza è preservare l'ordine di scrittura e, nello specifico, l'ordine di scrittura dipendente. Ad esempio, un database in scrittura su 10 LUN scrive simultaneamente su tutte. Molte scritture vengono emesse in modo asincrono, il che significa che l'ordine in cui vengono completate non è importante e l'ordine effettivo in cui vengono completate varia in base al comportamento del sistema operativo e della rete.

Alcune operazioni di scrittura devono essere presenti sul disco prima che il database possa procedere con operazioni di scrittura aggiuntive. Queste operazioni critiche di scrittura sono chiamate scritture dipendenti. I/o di scrittura successivi dipendono dalla presenza di queste scritture sul disco. Qualsiasi snapshot, recovery o replica di queste 10 LUN deve garantire l'ordine di scrittura dipendente. Gli aggiornamenti del file system sono un altro esempio di scritture dipendenti dall'ordine di scrittura. L'ordine in cui vengono apportate le modifiche al file system deve essere mantenuto o l'intero file system potrebbe danneggiarsi.

Strategie

Esistono due approcci principali ai backup basati su snapshot:

- Backup coerenti con il crash
- Backup a caldo protetti dagli snapshot

Un backup coerente con i crash di un database si riferisce all'acquisizione dell'intera struttura del database, inclusi i file di dati, i log di ripristino e i file di controllo, in un singolo momento. Se il database è memorizzato in un singolo volume, il processo è semplice ed è possibile creare una Snapshot in qualsiasi momento. Se un database si estende su volumi, è necessario creare uno snapshot del gruppo di coerenza (CG). Esistono diverse opzioni per la creazione di snapshot CG, tra cui il software NetApp SnapCenter, le funzionalità native del gruppo di coerenza ONTAP e gli script gestiti dagli utenti.

I backup Snapshot coerenti con i crash vengono utilizzati principalmente quando è sufficiente un recovery point-of-the-backup. In alcune circostanze è possibile applicare i registri di archivio, ma quando è necessario un ripristino point-in-time più granulare, è preferibile un backup online.

La procedura di base per un backup online basato su snapshot è la seguente:

1. Inserire il database in `backup` modalità.
2. Creare una snapshot di tutti i volumi che ospitano file di dati.
3. Esci `backup` modalità.
4. Eseguire il comando `alter system archive log current` per forzare l'archiviazione del registro.
5. Creare snapshot di tutti i volumi che ospitano i log di archivio.

Questa procedura produce una serie di istantanee contenenti file di dati in modalità backup e i registri di archivio critici generati in modalità backup. Questi sono i due requisiti per il ripristino di un database. I file come i file di controllo dovrebbero essere protetti per comodità, ma l'unico requisito assoluto è la protezione per i file di dati e i registri di archivio.

Sebbene i diversi clienti possano avere strategie molto diverse, quasi tutte queste strategie si basano in ultima analisi sugli stessi principi delineati di seguito.

Recovery basato su Snapshot

Quando si progettano layout di volumi per database Oracle, la prima decisione è se utilizzare la tecnologia VBSR (Volume-Based NetApp SnapRestore).

La funzione SnapRestore basata su volume consente di ripristinare quasi istantaneamente un volume in un point-in-time precedente. Poiché tutti i dati sul volume vengono ripristinati, VBSR potrebbe non essere appropriato per tutti i casi di utilizzo. Ad esempio, se un intero database, inclusi file di dati, log di ripristino e log di archivio, viene memorizzato in un singolo volume e questo volume viene ripristinato con VBSR, i dati vengono persi perché i log di archivio e i dati di ripristino più recenti vengono scartati.

VBSR non è necessario per il ripristino. Molti database possono essere ripristinati utilizzando SFSR (Single-file SnapRestore) basato su file o semplicemente copiando i file dalla snapshot nel file system attivo.

VBSR è preferibile quando un database è molto grande o quando deve essere recuperato il più rapidamente possibile, e l'uso di VBSR richiede l'isolamento dei file di dati. In un ambiente NFS, i file di dati di un dato database devono essere archiviati in volumi dedicati che non sono contaminati da alcun altro tipo di file. In un ambiente SAN, i file di dati devono essere memorizzati in LUN dedicate su volumi dedicati. Se viene utilizzato un volume manager (incluso Oracle Automatic Storage Management [ASM]), il gruppo di dischi deve essere dedicato anche ai file di dati.

L'isolamento dei file di dati in questo modo consente loro di tornare a uno stato precedente senza danneggiare altri file system.

Riserva di Snapshot

Per ogni volume con i dati Oracle in un ambiente SAN, il `percent-snapshot-space` Dovrebbe essere impostato su zero perché non è utile riservare spazio per uno snapshot in un ambiente LUN. Se la riserva frazionaria è impostata su 100, uno snapshot di un volume con LUN richiede spazio libero sufficiente nel volume, esclusa la riserva snapshot, per assorbire il 100% di turnover di tutti i dati. Se la riserva frazionaria è impostata su un valore inferiore, è necessaria una quantità di spazio libero corrispondente inferiore, ma esclude sempre la riserva istantanea. Ciò significa che viene sprecato lo spazio di riserva di Snapshot in un ambiente LUN.

In un ambiente NFS, esistono due opzioni:

- Impostare `percent-snapshot-space` in base al consumo di spazio snapshot previsto.
- Impostare `percent-snapshot-space` a zero e gestire collettivamente il consumo di spazio attivo e snapshot.

Con la prima opzione, `percent-snapshot-space` è impostato su un valore diverso da zero, in genere intorno al 20%. Questo spazio viene quindi nascosto all'utente. Tuttavia, questo valore non crea un limite di utilizzo. Se un database con una prenotazione del 20% registra un fatturato del 30%, lo spazio snapshot può crescere oltre i limiti della riserva del 20% e occupare spazio non riservato.

Il vantaggio principale dell'impostazione di una riserva a un valore come 20% è verificare che una parte di spazio sia sempre disponibile per gli snapshot. Ad esempio, un volume da 1TB TB con una riserva del 20% consentirebbe all'amministratore di database (DBA) di memorizzare 800GB TB di dati. Questa configurazione garantisce almeno 200GB GB di spazio per il consumo di snapshot.

Quando `percent-snapshot-space` è impostato su zero, tutto lo spazio nel volume è disponibile per l'utente finale, il che garantisce una migliore visibilità. Un DBA deve capire che, se rileva un volume di 1TB GB che sfrutta le snapshot, questo 1TB GB di spazio viene condiviso tra i dati attivi e il turnover di Snapshot.

Non esiste una chiara preferenza tra l'opzione 1 e l'opzione 2 tra gli utenti finali.

ONTAP e snapshot di terze parti

Oracle Doc ID 604683,1 illustra i requisiti per il supporto di snapshot di terze parti e le varie opzioni disponibili

per le operazioni di backup e ripristino.

Il fornitore di terze parti deve garantire che le istantanee dell'azienda siano conformi ai seguenti requisiti:

- Gli snapshot devono integrarsi con le operazioni di ripristino e ripristino consigliate da Oracle.
- Gli snapshot devono essere coerenti con il crash del database nel punto dello snapshot.
- L'ordine di scrittura viene mantenuto per ogni file all'interno di uno snapshot.

I prodotti di gestione ONTAP e NetApp di Oracle sono conformi a questi requisiti.

SnapRestore

La tecnologia NetApp SnapRestore offre il ripristino rapido dei dati in ONTAP a partire da una snapshot.

Quando un set di dati critico non è disponibile, le operazioni di business critiche non sono attive. I nastri possono interrompersi e persino i ripristini da backup basati su disco possono essere lenti da trasferire sulla rete. SnapRestore consente di evitare questi problemi grazie al ripristino quasi istantaneo dei set di dati. Anche i database di diversi petabyte possono essere ripristinati completamente con pochi minuti di lavoro.

Esistono due forme di SnapRestore: Basata su file/LUN e basata su volume.

- Singoli file o LUN possono essere ripristinati in pochi secondi, sia in una LUN da 2TB GB che in un file da 4KB GB.
- Il container di file o LUN può essere ripristinato in pochi secondi, siano essi 10GB o 100TB TB di dati.

Un "contenitore di file o LUN" generalmente si riferisce a un volume FlexVol. Ad esempio, potresti avere 10 LUN che costituiscono un gruppo di dischi LVM in un singolo volume, oppure un volume potrebbe archiviare le home directory NFS di 1000 utenti. Invece di eseguire un'operazione di ripristino per ogni singolo file o LUN, è possibile ripristinare l'intero volume come un'unica operazione. Questo processo funziona anche con container scale-out che includono volumi multipli, come FlexGroup o un gruppo di coerenza ONTAP.

Il motivo per cui SnapRestore funziona in modo così rapido ed efficiente è dovuto alla natura di uno snapshot, che è essenzialmente una vista parallela di sola lettura del contenuto di un volume in uno specifico momento. I blocchi attivi sono i blocchi reali che è possibile modificare, mentre lo snapshot è una vista di sola lettura dello stato dei blocchi che costituiscono i file e le LUN al momento della creazione dello snapshot.

ONTAP consente solo l'accesso in sola lettura ai dati snapshot, ma i dati possono essere riattivati con SnapRestore. Lo snapshot viene riabilitato come visualizzazione lettura-scrittura dei dati, riportando i dati allo stato precedente. SnapRestore può operare a livello di volume o di file. La tecnologia è essenzialmente la stessa con alcune differenze minori nel comportamento.

SnapRestore volume

La SnapRestore basata su volume riporta l'intero volume di dati a uno stato precedente. Questa operazione non richiede lo spostamento dei dati, il che significa che il processo di ripristino è essenzialmente istantaneo, sebbene l'elaborazione delle operazioni API o CLI possa richiedere alcuni secondi. Il ripristino di 1GB TB di dati non è più complicato o richiede molto tempo rispetto al ripristino di 1PB TB di dati. Questa funzionalità è il motivo principale per cui molti clienti aziendali migrano ai sistemi storage ONTAP. Offre un RTO misurato in secondi anche per i set di dati più grandi.

Uno svantaggio di SnapRestore basato su volumi è causato dal fatto che le modifiche all'interno di un volume sono cumulative nel tempo. Pertanto, ogni snapshot e i dati del file attivo dipendono dalle modifiche che hanno

portato a quel punto. Ripristinare uno stato precedente di un volume significa ignorare tutte le modifiche successive apportate ai dati. Ciò che è meno ovvio, tuttavia, è che questo include gli snapshot creati successivamente. Ciò non è sempre desiderabile.

Ad esempio, uno SLA di conservazione dei dati può specificare 30 giorni di backup notturni. Il ripristino di un set di dati in uno snapshot creato cinque giorni fa con Volume SnapRestore scaricherebbe tutti gli snapshot creati nei cinque giorni precedenti, violando lo SLA.

Sono disponibili diverse opzioni per risolvere questo limite:

1. I dati possono essere copiati da una snapshot precedente, invece di eseguire un SnapRestore dell'intero volume. Questo metodo funziona meglio con set di dati più piccoli.
2. È possibile clonare una snapshot invece di ripristinarla. Il limite a questo approccio è che lo snapshot di origine è una dipendenza del clone. Pertanto, non può essere eliminato a meno che il clone non venga anch'esso eliminato o diviso in un volume indipendente.
3. Utilizzo di SnapRestore basati su file.

File SnapRestore (Stato file)

SnapRestore basato su file è un processo di ripristino più granulare e basato su snapshot. Invece di ripristinare lo stato di un intero volume, viene ripristinato lo stato di un singolo file o LUN. Non è necessario eliminare gli snapshot, né questa operazione crea alcuna dipendenza da uno snapshot precedente. Il file o LUN diventa immediatamente disponibile nel volume attivo.

Durante il ripristino di SnapRestore di un file o LUN non è necessario alcuno spostamento dei dati. Tuttavia, alcuni aggiornamenti dei metadati interni sono necessari per riflettere il fatto che i blocchi sottostanti in un file o LUN ora esistono sia in una snapshot che nel volume attivo. Non dovrebbe avere alcun effetto sulle prestazioni, ma questo processo blocca la creazione di snapshot fino al completamento. La velocità di elaborazione è di circa 5Gbps MB (18TB MB/ora) in base alla dimensione totale dei file ripristinati.

Backup in linea

Per proteggere e ripristinare un database Oracle in modalità backup sono richiesti due set di dati. Si noti che questa non è l'unica opzione di backup di Oracle, ma è la più comune.

- Un'istantanea dei file di dati in modalità di backup
- I registri di archivio creati mentre i file di dati erano in modalità backup

Se è richiesto il recupero completo, comprese tutte le transazioni impegnate, è necessario un terzo elemento:

- Una serie di registri di ripristino correnti

Esistono diversi modi per eseguire il ripristino di un backup online. Molti clienti ripristinano le snapshot utilizzando l'interfaccia CLI di ONTAP e quindi Oracle RMAN o sqlplus per completare il ripristino. Ciò è particolarmente comune negli ambienti di produzione di grandi dimensioni, in cui la probabilità e la frequenza dei ripristini dei database sono estremamente ridotte e qualsiasi procedura di ripristino viene gestita da un DBA esperto. Per un'automazione completa, soluzioni come NetApp SnapCenter includono un plug-in Oracle con interfacce sia a riga di comando che grafiche.

Alcuni clienti su larga scala hanno adottato un approccio più semplice configurando script di base sugli host per impostare i database in modalità di backup in un momento specifico in preparazione a uno snapshot pianificato. Ad esempio, pianificare il comando `alter database begin backup` alle 23:58, `alter`

`database end backup` alle 00:02, quindi programmare le snapshot direttamente sul sistema storage a mezzanotte. Il risultato è una strategia di backup semplice e altamente scalabile che non richiede licenze o software esterni.

Layout dei dati

Il layout più semplice consiste nell'isolare i file di dati in uno o più volumi dedicati. Non devono essere contaminati da alcun altro tipo di file. In questo modo si garantisce che i volumi dei file dati possano essere ripristinati rapidamente tramite un'operazione SnapRestore senza distruggere un log di ripristino, controlfile o un log di archivio importante.

LE SAN hanno requisiti simili per l'isolamento dei file dati all'interno di volumi dedicati. Con un sistema operativo come Microsoft Windows, un singolo volume potrebbe contenere più LUN di file dati, ciascuno con un file system NTFS. Con altri sistemi operativi, in genere esiste un volume manager logico. Ad esempio, con Oracle ASM, l'opzione più semplice sarebbe limitare i LUN di un gruppo di dischi ASM a un singolo volume che può essere sottoposto a backup e ripristinato come unità. Se per motivi di gestione delle performance o della capacità sono necessari volumi aggiuntivi, la creazione di un gruppo di dischi aggiuntivo sul nuovo volume semplifica la gestione.

Se vengono seguite queste linee guida, le snapshot possono essere pianificate direttamente sul sistema di storage, senza che sia necessario eseguire uno snapshot del gruppo di coerenza. Il motivo è che i backup Oracle non richiedono il backup dei file di dati contemporaneamente. La procedura di backup online è stata progettata per consentire ai file di dati di continuare ad essere aggiornati, poiché vengono lentamente trasmessi su nastro nel corso delle ore.

Una complicazione si verifica in situazioni come l'utilizzo di un gruppo di dischi ASM distribuito tra i volumi. In questi casi, è necessario eseguire uno snapshot cg per assicurarsi che i metadati ASM siano coerenti in tutti i volumi costituenti.

Attenzione: verificare che l'ASM `spfile` e. `passwd` i file non si trovano nel gruppo di dischi che ospita i file di dati. Ciò interferisce con la capacità di ripristinare selettivamente i dati e solo i file di dati.

Procedura di ripristino locale: NFS

Questa procedura può essere gestita manualmente o tramite un'applicazione come SnapCenter. La procedura di base è la seguente:

1. Arrestare il database.
2. Recuperare i volumi di file dati nello snapshot immediatamente prima del punto di ripristino desiderato.
3. Riprodurre i log di archivio nel punto desiderato.
4. Se si desidera completare il ripristino, riprodurre i registri di ripristino correnti.

Questa procedura presuppone che i log di archivio desiderati siano ancora presenti nel file system attivo. In caso contrario, è necessario ripristinare i log di archivio oppure è possibile indirizzare rman/sqlplus ai dati nella directory snapshot.

Inoltre, per i database di dimensioni inferiori, i file di dati possono essere recuperati da un utente finale direttamente da `.snapshot` directory senza l'assistenza di tool di automazione o amministratori dello storage per eseguire una `snaprestore` comando.

Procedura di ripristino locale: SAN

Questa procedura può essere gestita manualmente o tramite un'applicazione come SnapCenter. La procedura di base è la seguente:

1. Arrestare il database.
2. Chiudere i gruppi di dischi che ospitano i file di dati. La procedura varia a seconda del volume manager logico scelto. Con ASM, il processo richiede lo smontaggio del gruppo di dischi. Con Linux, i file system devono essere smontati e i volumi logici e i gruppi di volumi devono essere disattivati. L'obiettivo è quello di interrompere tutti gli aggiornamenti del gruppo di volumi di destinazione da ripristinare.
3. Ripristinare i gruppi di dischi del file dati nello snapshot immediatamente prima del punto di ripristino desiderato.
4. Riattivare i gruppi di dischi appena ripristinati.
5. Riprodurre i log di archivio nel punto desiderato.
6. Se si desidera eseguire il ripristino completo, riprodurre tutti i registri di ripristino.

Questa procedura presuppone che i log di archivio desiderati siano ancora presenti nel file system attivo. In caso contrario, è necessario ripristinare i registri di archivio portando i LUN del registro di archivio offline ed eseguendo un ripristino. Questo è anche un esempio in cui è utile dividere i log di archivio in volumi dedicati. Se i log dell'archivio condividono un gruppo di volumi con log di ripristino, i log di ripristino devono essere copiati in un altro punto prima di ripristinare il set complessivo di LUN. Questa fase impedisce la perdita di tali transazioni finali registrate.

Backup ottimizzati per le snapshot di storage

Il backup e il ripristino basati su Snapshot sono diventati ancora più semplici quando è stato rilasciato Oracle 12c perché non è necessario collocare un database in modalità hot backup. Il risultato è la possibilità di pianificare backup basati su snapshot direttamente in un sistema storage, preservando comunque la capacità di eseguire ripristini completi o point-in-time.

Sebbene la procedura di ripristino con backup a caldo sia più familiare per gli amministratori di database, da molto tempo è stato possibile utilizzare istantanee che non sono state create mentre il database era in modalità di backup a caldo. Per rendere il database coerente, sono stati necessari ulteriori passaggi manuali con Oracle 10g e 11g durante il ripristino. Con Oracle 12c, `sqlplus` e. `rman` contenere la logica aggiuntiva per riprodurre i log di archivio sui backup dei file dati che non erano in modalità hot backup.

Come indicato in precedenza, il ripristino di un backup a caldo basato su snapshot richiede due set di dati:

- Un'istantanea dei file di dati creati in modalità backup
- I log di archivio generati mentre i file di dati erano in modalità hot backup

Durante il ripristino, il database legge i metadati dai file di dati per selezionare i log di archivio richiesti per il ripristino.

Per ottenere gli stessi risultati, il recovery ottimizzato per le snapshot di storage richiede set di dati leggermente diversi:

- Un'istantanea dei file di dati, più un metodo per identificare l'ora in cui è stata creata l'istantanea
- Archiviare i log dall'ora del checkpoint del file dati più recente all'ora esatta dello snapshot

Durante il ripristino, il database legge i metadati dai file di dati per identificare il registro di archivio più recente richiesto. È possibile eseguire il ripristino completo o point-in-time. Quando si esegue un ripristino point-in-time, è fondamentale conoscere l'ora dello snapshot dei file di dati. Il punto di ripristino specificato deve essere successivo all'ora di creazione degli snapshot. NetApp consiglia di aggiungere almeno alcuni minuti all'ora

dello snapshot per tenere conto della variazione dell'orologio.

Per informazioni dettagliate, vedere la documentazione di Oracle sull'argomento "Recovery Using Storage Snapshot Optimization" disponibile in varie versioni della documentazione di Oracle 12c. Inoltre, consultare l'ID documento Oracle Doc ID 604683,1 relativo al supporto per le istantanee di terze parti di Oracle.

Layout dei dati

Il layout più semplice consiste nell'isolare i file di dati in uno o più volumi dedicati. Non devono essere contaminati da alcun altro tipo di file. In questo modo si garantisce che i volumi dei file dati possano essere ripristinati rapidamente con un'operazione SnapRestore senza distruggere un log di ripristino, controlfile o un log di archivio importante.

LE SAN hanno requisiti simili per l'isolamento dei file dati all'interno di volumi dedicati. Con un sistema operativo come Microsoft Windows, un singolo volume potrebbe contenere più LUN di file dati, ciascuno con un file system NTFS. Con altri sistemi operativi, esiste in genere anche un volume manager logico. Ad esempio, con Oracle ASM, l'opzione più semplice sarebbe quella di limitare i gruppi di dischi a un singolo volume di cui è possibile eseguire il backup e il ripristino come unità. Se per motivi di gestione delle performance o della capacità sono necessari volumi aggiuntivi, la creazione di un gruppo di dischi aggiuntivo sul nuovo volume semplifica la gestione.

Se si seguono queste linee guida, gli snapshot possono essere pianificati direttamente su ONTAP senza che sia necessario eseguire uno snapshot del gruppo di coerenza. Il motivo è che i backup ottimizzati per le istantanee non richiedono che venga eseguito contemporaneamente il backup dei file di dati.

Una complicazione si verifica in situazioni come un gruppo di dischi ASM distribuito tra i volumi. In questi casi, è necessario eseguire uno snapshot cg per assicurarsi che i metadati ASM siano coerenti in tutti i volumi costituenti.

[Note]verificare che i file ASM spfile e passwd non siano nel gruppo di dischi che ospita i file di dati. Ciò interferisce con la capacità di ripristinare selettivamente i dati e solo i file di dati.

Procedura di ripristino locale: NFS

Questa procedura può essere gestita manualmente o tramite un'applicazione come SnapCenter. La procedura di base è la seguente:

1. Arrestare il database.
2. Recuperare i volumi di file dati nello snapshot immediatamente prima del punto di ripristino desiderato.
3. Riprodurre i log di archivio nel punto desiderato.

Questa procedura presuppone che i log di archivio desiderati siano ancora presenti nel file system attivo. In caso contrario, è necessario ripristinare i registri di archivio, o. `rman` oppure `sqlplus` può essere indirizzato ai dati in `.snapshot` directory.

Inoltre, per i database di dimensioni inferiori, i file di dati possono essere recuperati da un utente finale direttamente da `.snapshot` Senza l'assistenza di tool di automazione o di un amministratore dello storage per eseguire un comando SnapRestore.

Procedura di ripristino locale: SAN

Questa procedura può essere gestita manualmente o tramite un'applicazione come SnapCenter. La procedura di base è la seguente:

1. Arrestare il database.
2. Chiudere i gruppi di dischi che ospitano i file di dati. La procedura varia a seconda del volume manager logico scelto. Con ASM, il processo richiede lo smontaggio del gruppo di dischi. Con Linux, i file system devono essere smontati e i volumi logici e i gruppi di volumi sono disattivati. L'obiettivo è quello di interrompere tutti gli aggiornamenti del gruppo di volumi di destinazione da ripristinare.
3. Ripristinare i gruppi di dischi del file dati nello snapshot immediatamente prima del punto di ripristino desiderato.
4. Riattivare i gruppi di dischi appena ripristinati.
5. Riprodurre i log di archivio nel punto desiderato.

Questa procedura presuppone che i log di archivio desiderati siano ancora presenti nel file system attivo. In caso contrario, è necessario ripristinare i registri di archivio portando i LUN del registro di archivio offline ed eseguendo un ripristino. Questo è anche un esempio in cui è utile dividere i log di archivio in volumi dedicati. Se i log dell'archivio condividono un gruppo di volumi con i log di ripristino, i log di ripristino devono essere copiati in un altro punto prima del ripristino del set complessivo di LUN, per evitare di perdere le transazioni finali registrate.

Esempio di recupero completo

Si supponga che i file di dati siano stati corrotti o distrutti e che sia necessario un ripristino completo. La procedura da seguire è la seguente:

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover automatic;
Media recovery complete.
SQL> alter database open;
Database altered.
SQL>
```

Esempio di recupero point-in-time

L'intera procedura di ripristino è un singolo comando: `recover automatic`.

Se è necessario un ripristino point-in-time, l'indicatore data e ora degli snapshot deve essere noto e può essere identificato come segue:

```
Cluster01::> snapshot show -vserver vserver1 -volume NTAP_oradata -fields
create-time
vserver   volume        snapshot   create-time
--------  ------------  ---------  ------------------------
vserver1  NTAP_oradata  my-backup  Thu Mar 09 10:10:06 2017
```

L'ora di creazione dell'istantanea è indicata come marzo 9th e 10:10:06. Per essere sicuri, viene aggiunto un minuto all'ora dell'istantanea:

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover database until time '09-MAR-2017 10:44:15' snapshot time '09-
MAR-2017 10:11:00';
```

Il ripristino viene avviato. È stato specificato un tempo di snapshot di 10:11:00, un minuto dopo il tempo registrato per tenere conto della possibile varianza dell'orologio e un tempo di recupero target di 10:44. Successivamente, sqlplus richiede i registri di archivio necessari per raggiungere il tempo di ripristino desiderato di 10:44.

```
ORA-00279: change 551760 generated at 03/09/2017 05:06:07 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_31_930813377.dbf
ORA-00280: change 551760 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 552566 generated at 03/09/2017 05:08:09 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_32_930813377.dbf
ORA-00280: change 552566 for thread 1 is in sequence #32
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 553045 generated at 03/09/2017 05:10:12 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_33_930813377.dbf
ORA-00280: change 553045 for thread 1 is in sequence #33
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 753229 generated at 03/09/2017 05:15:58 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_34_930813377.dbf
ORA-00280: change 753229 for thread 1 is in sequence #34
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
Log applied.
Media recovery complete.
SQL> alter database open resetlogs;
Database altered.
SQL>
```

Completare il ripristino di un database utilizzando gli snapshot utilizzando `recover automatic` command non richiede licenze specifiche, ma utilizza un ripristino point-in-time `snapshot time` Richiede la licenza Oracle Advanced Compression.

Tool di gestione e automazione del database

Il valore primario di ONTAP in un ambiente di database Oracle deriva dalle tecnologie principali di ONTAP, come copie Snapshot istantanee, semplice replica SnapMirror e creazione efficiente dei volumi FlexClone.

In alcuni casi, una semplice configurazione di queste funzionalità chiave direttamente su ONTAP soddisfa i requisiti, ma esigenze più complesse richiedono un livello di orchestrazione.

SnapCenter

SnapCenter è il prodotto di punta della protezione dei dati di NetApp. A un livello molto basso, è simile ai prodotti SnapManager in termini di modalità di esecuzione dei backup del database, ma è stato creato da zero per fornire un singolo pannello di controllo per la gestione della protezione dati sui sistemi di storage NetApp.

SnapCenter include le funzioni di base come backup e ripristini basati su snapshot, la replica SnapMirror e SnapVault e altre funzionalità necessarie per operare su larga scala per le grandi imprese. Queste funzionalità

avanzate includono una funzionalità estesa di controllo degli accessi in base al ruolo (RBAC), API RESTful per l'integrazione con prodotti di orchestrazione di terze parti, gestione centrale senza interruzioni dei plug-in SnapCenter sugli host di database e un'interfaccia utente progettata per ambienti cloud-scale.

RIPOSO

ONTAP contiene anche un ricco set di API RESTful. Questo consente a 3rd vendor di creare data Protection e altre applicazioni di gestione con una profonda integrazione con ONTAP. Inoltre, l'API RESTful è facile da utilizzare da parte dei clienti che desiderano creare i propri flussi di lavoro e utility di automazione.

Disaster recovery Oracle

Panoramica

Il disaster recovery si riferisce al ripristino dei servizi dati dopo un evento catastrofico, come un incendio che distrugge un sistema storage o persino un'intera sede.

Questa documentazione sostituisce i report tecnici precedentemente pubblicati *TR-4591: Oracle Data Protection* e *TR-4592: Oracle on MetroCluster.*

Il disaster recovery può essere eseguito mediante una semplice replica dei dati tramite SnapMirror, naturalmente, con molti clienti che aggiornano le repliche con mirroring ogni ora.

Per la maggior parte dei clienti, il disaster recovery non richiede solo una copia remota dei dati, ma anche la capacità di sfruttarli in maniera rapida. NetApp offre due tecnologie che soddisfano questa esigenza: MetroCluster e SnapMirror Active Sync

MetroCluster fa riferimento a ONTAP in una configurazione hardware che include storage con mirroring sincrono di basso livello e numerose funzionalità aggiuntive. Le soluzioni integrate come MetroCluster semplificano le complesse e scalabili infrastrutture di database, applicazioni e virtualizzazione. Sostituisce diversi prodotti e strategie di protezione dati esterni con un unico semplice storage array centrale. Fornisce inoltre backup, recovery, disaster recovery e alta disponibilità (ha) integrati in un singolo sistema storage in cluster.

La sincronizzazione attiva di SnapMirror (SM-AS) si basa sulla sincronizzazione sincrona di SnapMirror. Con MetroCluster, ogni controller ONTAP è responsabile della replica dei dati dell'unità in una posizione remota. Con la sincronizzazione attiva di SnapMirror, avrai essenzialmente due sistemi ONTAP diversi che mantengono copie indipendenti dei dati LUN, ma cooperano per presentare una singola istanza di tale LUN. Dal punto di vista dell'host, si tratta di una singola entità LUN.

Confronto SM-AS e MCC

SM-AS e MetroCluster sono simili per quanto riguarda le funzionalità generali, ma esistono importanti differenze nel modo in cui è stata implementata la replica RPO=0 e nel modo in cui viene gestita. Anche se è possibile utilizzare la modalità asincrona e sincrona di SnapMirror come parte di un piano di disaster recovery, non sono progettate come tecnologie di replica ha.

- Una configurazione MetroCluster è più simile a un cluster integrato con nodi distribuiti tra i siti. SM-AS si comporta come due cluster altrimenti indipendenti che stanno cooperando nel servire un RPO selezionato=0 LUN replicati in modo sincrono.
- I dati in una configurazione MetroCluster sono accessibili solo da un determinato sito alla volta. Una seconda copia dei dati è presente sul sito opposto, ma i dati sono passivi. Non è possibile accedervi senza un failover del sistema storage.

- MetroCluster e SM-as eseguono il mirroring a diversi livelli. Il mirroring MetroCluster viene eseguito al livello RAID. I dati di basso livello sono memorizzati in un formato di mirroring utilizzando SyncMirror. L'utilizzo del mirroring è praticamente invisibile ai livelli di LUN, volume e protocollo.
- Al contrario, il mirroring SM-AS avviene a livello di protocollo. I due cluster sono complessivamente cluster indipendenti. Una volta sincronizzate le due copie di dati, i due cluster devono solo eseguire il mirroring delle scritture. Quando si verifica una scrittura su un cluster, questa viene replicata nell'altro cluster. La scrittura viene riconosciuta all'host solo quando la scrittura è stata completata su entrambi i siti. A parte questo comportamento di suddivisione del protocollo, i due cluster sono altrimenti normali cluster ONTAP.
- Il ruolo principale di MetroCluster è la replica su larga scala. Puoi replicare un intero array con RPO=0 e RTO prossimo allo zero. Questo semplifica il processo di failover perché esiste un solo "problema" da eseguire e consente una scalabilità perfetta in termini di capacità e IOPS.
- Un caso d'utilizzo chiave per SM-AS è la replica granulare. A volte non vuoi replicare tutti i dati come una singola unità oppure devi eseguire il failover selettivo su alcuni carichi di lavoro.
- Un altro caso d'utilizzo chiave per SM-AS è per operazioni Active-Active, dove desideri che siano disponibili copie dei dati completamente utilizzabili su due cluster diversi situati in due posizioni diverse con caratteristiche di performance identiche e, se desiderato, non richiedere l'estensione della SAN tra i siti. Le applicazioni possono essere già in esecuzione su entrambi i siti, riducendo così l'RTO complessivo durante le operazioni di failover.

MetroCluster

Disaster recovery con MetroCluster

MetroCluster è una funzionalità di ONTAP in grado di proteggere i database Oracle con RPO=0 mirroring sincrono tra i siti, per poi scalare in verticale e supportare centinaia di database su un singolo sistema MetroCluster.

È anche semplice da usare. L'utilizzo di MetroCluster non aggiunge o modifica necessariamente i migliori percorsi per la gestione di applicazioni e database aziendali.

Le normali Best practice vengono comunque applicate e se le tue esigenze richiedono solo RPO=0:1 di data Protection, allora MetroCluster ne soddisfa l'esigenza. Tuttavia, la maggior parte dei clienti utilizza MetroCluster non solo per la protezione dei dati con RPO=0, ma anche per migliorare l'RTO in scenari di disastro, oltre a fornire un failover trasparente come parte delle attività di manutenzione del sito.

Architettura fisica

Per comprendere il funzionamento dei database Oracle in un ambiente MetroCluster è necessario spiegare la progettazione fisica di un sistema MetroCluster.

Questa documentazione sostituisce il report tecnico precedentemente pubblicato *TR-4592: Oracle su MetroCluster.*

MetroCluster è disponibile in 3 diverse configurazioni

- Coppie HA con connettività IP
- Coppie HA con connettività FC
- Controller singolo con connettività FC

Il termine connettività si riferisce alla connessione cluster utilizzata per la replica tra siti. Non si riferisce ai protocolli host. Tutti i protocolli lato host sono supportati come di consueto in una configurazione MetroCluster indipendentemente dal tipo di connessione utilizzata per la comunicazione tra cluster.

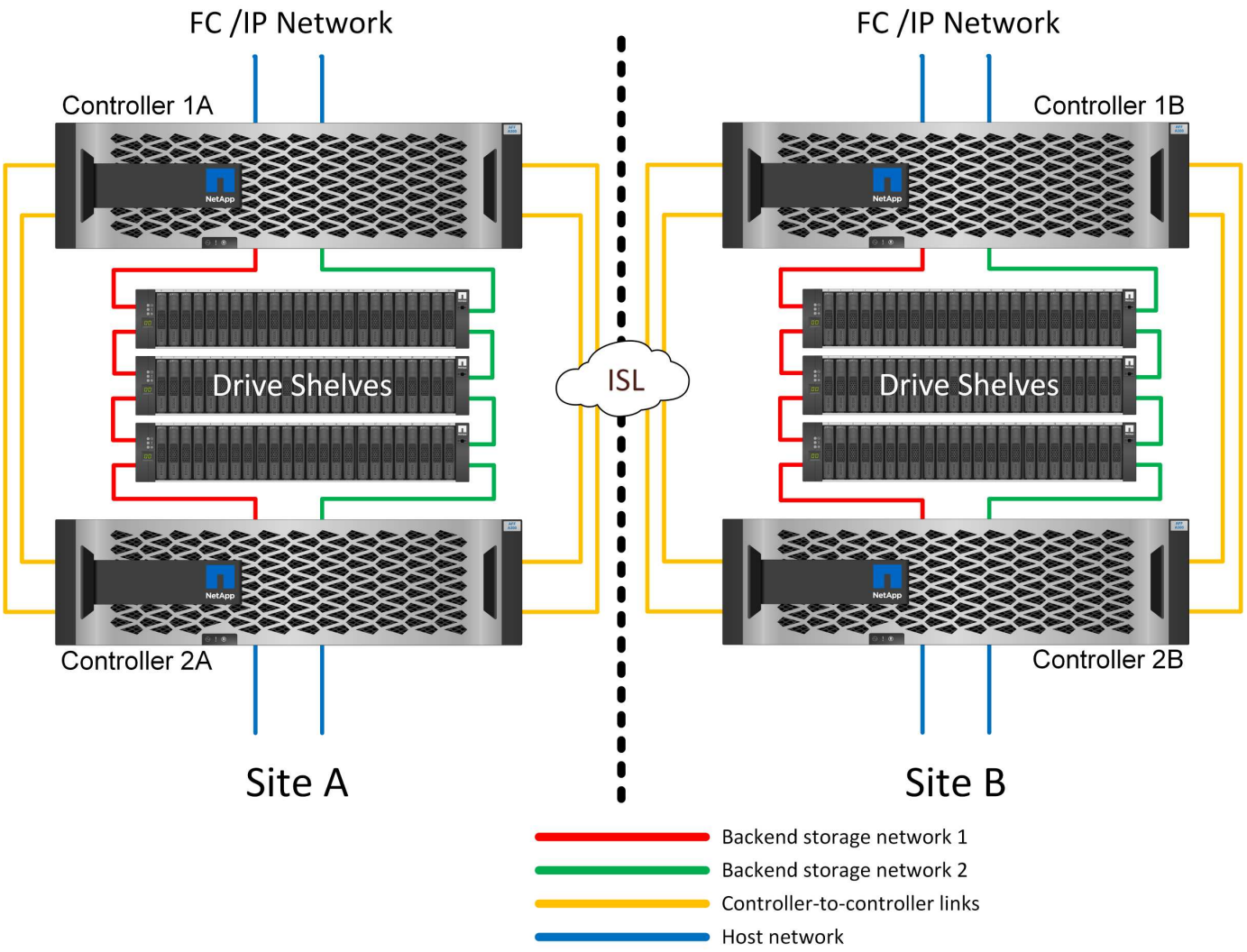
IP MetroCluster

La configurazione MetroCluster IP ha-Pair utilizza due o quattro nodi per sito. Questa opzione di configurazione aumenta la complessità e i costi rispetto all'opzione a due nodi, ma offre un vantaggio importante: La ridondanza intrasite. Un semplice errore del controller non richiede l'accesso ai dati nella WAN. L'accesso ai dati rimane locale attraverso il controller locale alternativo.

La maggior parte dei clienti sceglie la connettività IP perché i requisiti dell'infrastruttura sono più semplici. In passato, la connettività cross-site ad alta velocità era generalmente più semplice da fornire utilizzando gli switch FC e in fibra scura, ma oggi i circuiti IP ad alta velocità e a bassa latenza sono più prontamente disponibili.

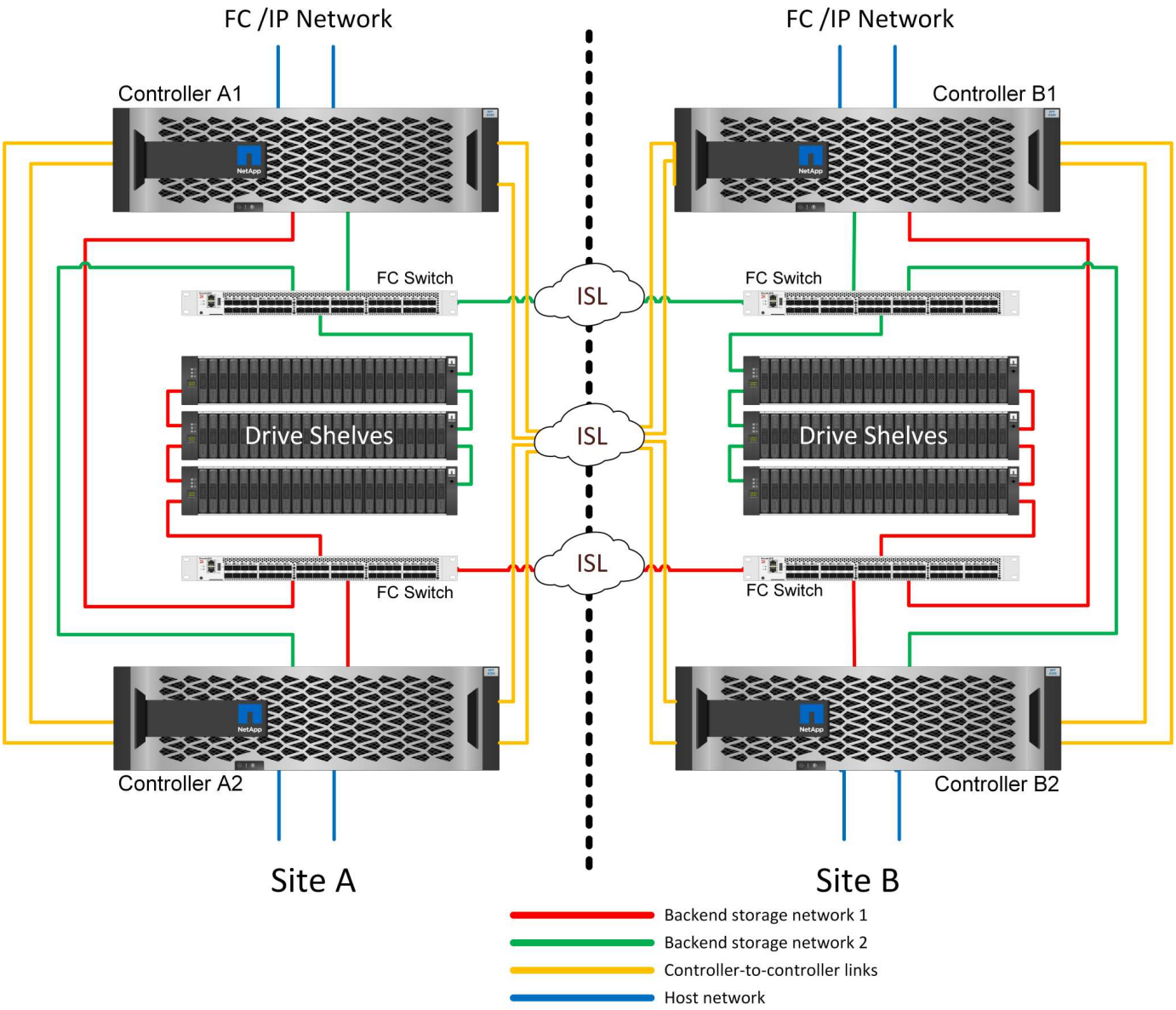
L'architettura è anche più semplice perché le uniche connessioni cross-site sono per i controller. Nei MetroClusters collegati a FC SAN, un controller scrive direttamente sulle unità del sito opposto e quindi richiede connessioni SAN, switch e bridge aggiuntivi. Al contrario, un controller in una configurazione IP scrive sulle unità opposte tramite il controller.

Per ulteriori informazioni, consultare la documentazione ufficiale di ONTAP e. "Architettura e progettazione della soluzione IP di MetroCluster".



MetroCluster HA-Pair FC SAN-Attached

La configurazione ha-Pair MetroCluster FC utilizza due o quattro nodi per sito. Questa opzione di configurazione aumenta la complessità e i costi rispetto all'opzione a due nodi, ma offre un vantaggio importante: La ridondanza intrasite. Un semplice errore del controller non richiede l'accesso ai dati nella WAN. L'accesso ai dati rimane locale attraverso il controller locale alternativo.

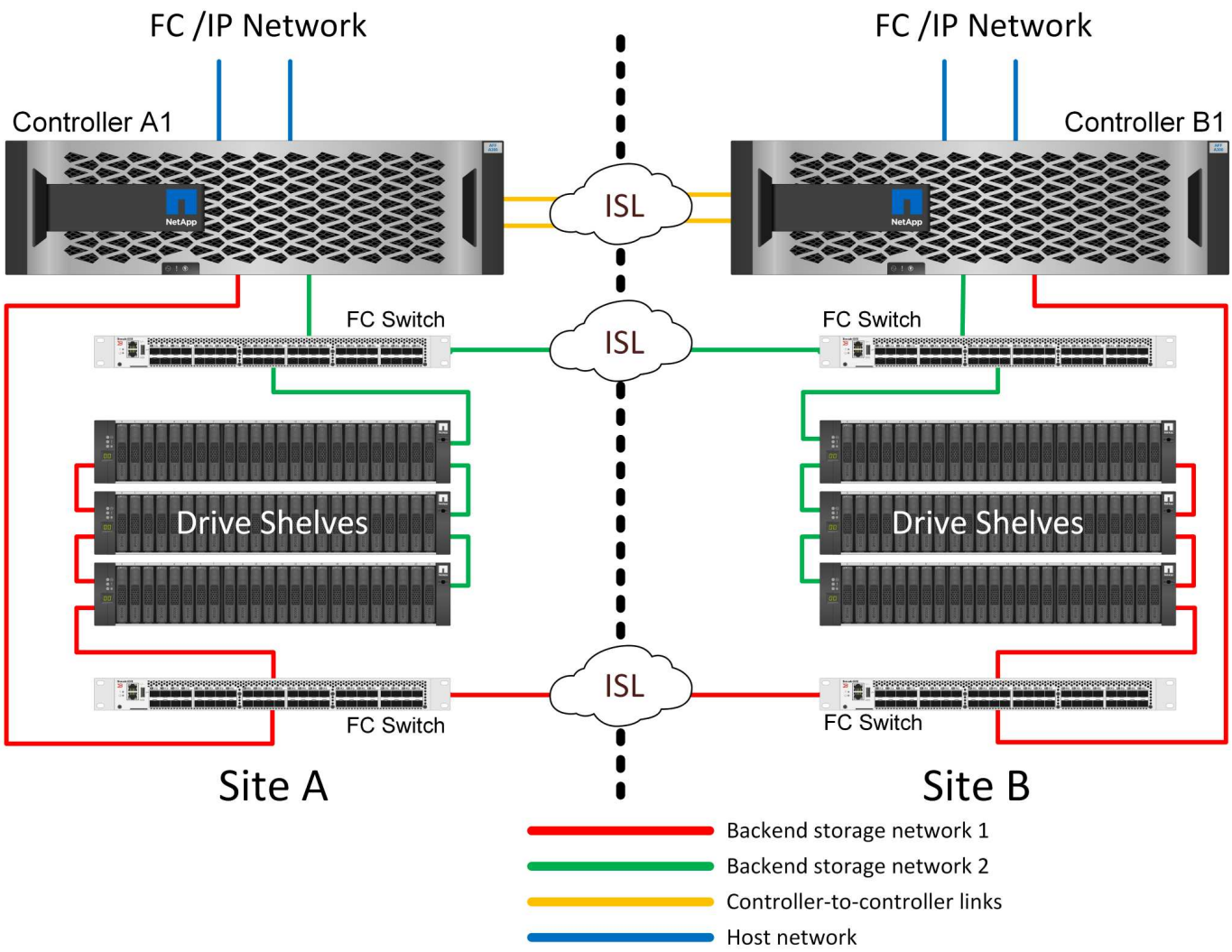


Alcune infrastrutture multisito non sono progettate per le operazioni Active-Active, ma vengono utilizzate maggiormente come sito primario e sito di disaster recovery. In questa situazione, è generalmente preferibile un'opzione ha-Pair MetroCluster per i seguenti motivi:

- Anche se un cluster MetroCluster a due nodi è un sistema ha, un guasto imprevisto di un controller o una manutenzione pianificata richiedono che i servizi dati vengano online sul sito opposto. Se la connettività di rete tra i siti non supporta la larghezza di banda richiesta, le prestazioni ne risentono. L'unica opzione sarebbe anche eseguire il failover dei vari sistemi operativi host e dei servizi associati al sito alternativo. Il cluster MetroCluster ha-Pair elimina questo problema grazie alla perdita di un controller che consente di eseguire un semplice failover all'interno dello stesso sito.
- Alcune topologie di rete non sono progettate per l'accesso tra siti, ma utilizzano sottoreti o SAN FC isolate. In questi casi, il cluster MetroCluster a due nodi non funziona più come sistema ha, perché il controller alternativo non può fornire dati ai server del sito opposto. L'opzione ha-Pair MetroCluster è necessaria per garantire ridondanza completa.
- Se un'infrastruttura a due siti viene vista come una singola infrastruttura ad alta disponibilità, la configurazione MetroCluster a due nodi è adatta. Tuttavia, se il sistema deve funzionare per un periodo di tempo prolungato dopo il guasto del sito, è preferibile una coppia ha perché continua a fornire ha all'interno di un singolo sito.

MetroCluster FC SAN-attached a due nodi

La configurazione MetroCluster a due nodi utilizza un solo nodo per sito. Questo design è più semplice rispetto all'opzione ha-Pair perché richiede meno componenti da configurare e gestire. Inoltre, ha ridotto le richieste di infrastruttura in termini di cablaggio e switch FC. Infine, riduce i costi.



L'evidente impatto di questa progettazione è che un guasto del controller su un singolo sito implica che i dati sono disponibili dal sito opposto. Questa restrizione non è necessariamente un problema. Molte aziende hanno operazioni di data center multisito con reti estese, ad alta velocità e a bassa latenza che funzionano essenzialmente come una singola infrastruttura. In questi casi, la configurazione preferita è la versione a due nodi di MetroCluster. Diversi service provider utilizzano attualmente sistemi a due nodi con scalabilità di petabyte.

Funzionalità di resilienza di MetroCluster

Non esistono single point of failure in una soluzione MetroCluster:

- Ogni controller dispone di due percorsi indipendenti verso gli shelf di dischi sul sito locale.
- Ogni controller dispone di due percorsi indipendenti verso gli shelf di dischi sul sito remoto.
- Ciascun controller dispone di due percorsi indipendenti verso i controller sul sito opposto.
- Nella configurazione ha-Pair, ogni controller ha due percorsi verso il partner locale.

Riassumendo, qualsiasi componente della configurazione può essere rimosso senza compromettere la capacità di MetroCluster di fornire dati. L'unica differenza in termini di resilienza tra le due opzioni è che la versione ha-Pair è ancora un sistema storage ha generale dopo un guasto del sito.

Architettura logica

Per comprendere il funzionamento dei database Oracle in un ambiente MetroCluster alsop è necessario spiegare alcune delle funzionalità logiche di un sistema MetroCluster.

Protezione da errori del sito: NVRAM e MetroCluster

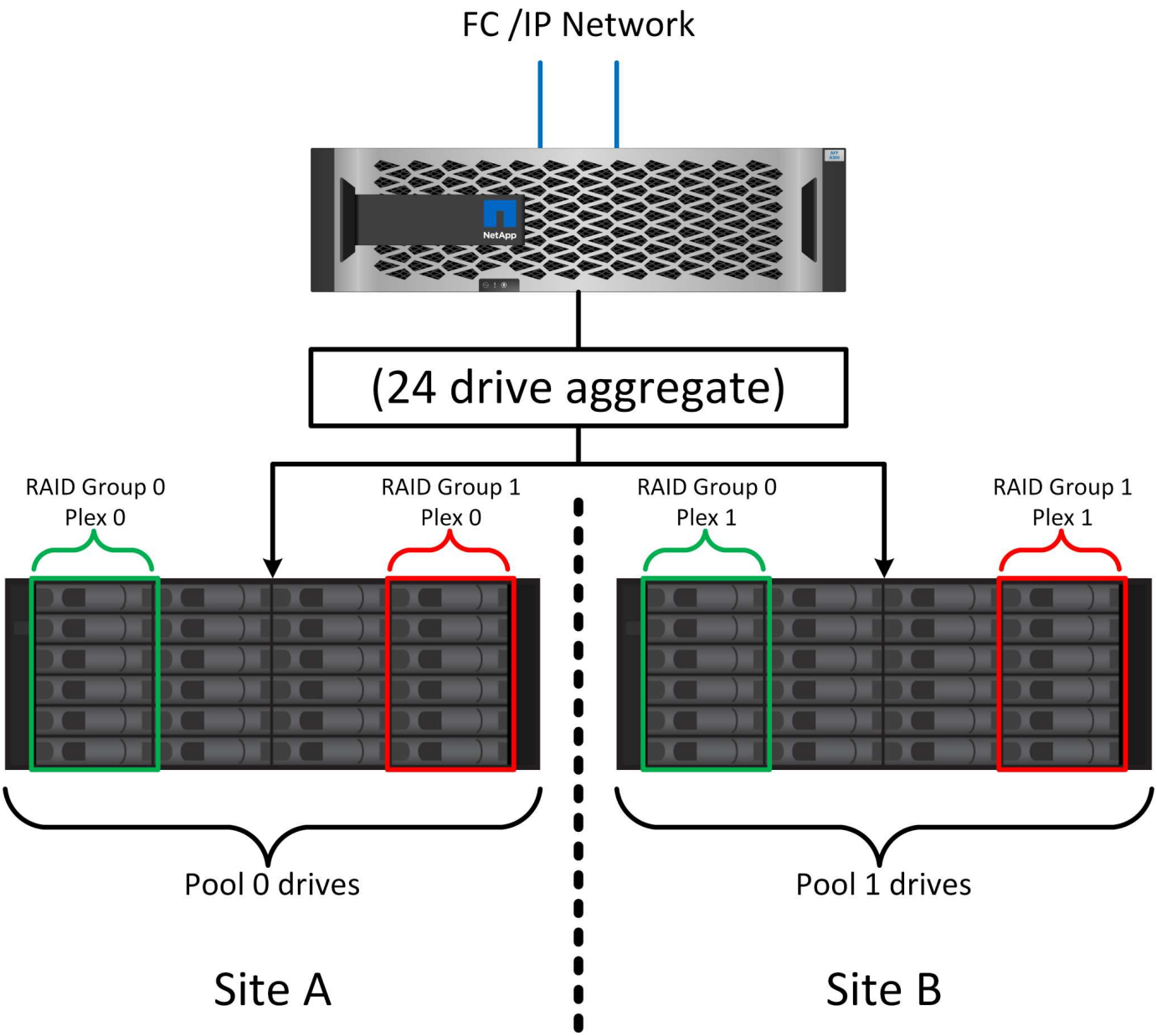
MetroCluster estende la protezione dei dati NVRAM nei seguenti modi:

- In una configurazione a due nodi, i dati NVRAM vengono replicati attraverso i collegamenti Inter-Switch (ISL) al partner remoto.
- In una configurazione ha-Pair, i dati NVRAM vengono replicati sia nel partner locale che in un partner remoto.
- Una scrittura non viene riconosciuta fino a quando non viene replicata a tutti i partner. Questa architettura protegge gli i/o in fase di trasferimento dai guasti del sito replicando i dati NVRAM a un partner remoto. Il processo non è coinvolto nella replica dei dati a livello di unità. Il controller proprietario degli aggregati si occupa della replica dei dati per iscritto a entrambi i plessi dell'aggregato, ma in caso di perdita del sito occorre comunque proteggere dalle perdite di i/o in fase di trasferimento. I dati NVRAM replicati sono utilizzati solo se un partner controller deve subentrare a un controller guasto.

Protezione dai guasti di shelf e siti: SyncMirror e plessi

SyncMirror è una tecnologia di mirroring che migliora, ma non sostituisce, RAID DP o RAID-TEC. Esegue il mirroring del contenuto di due gruppi RAID indipendenti. La configurazione logica è la seguente:

1. I dischi sono configurati in due pool in base alla posizione. Un pool è composto da tutti i dischi sul sito A, mentre il secondo è composto da tutti i dischi sul sito B.
2. Viene quindi creato un pool di storage comune, detto aggregato, in base a set di gruppi RAID con mirroring. Viene ottenuto lo stesso numero di unità per ciascun sito. Ad esempio, un aggregato SyncMirror da 20 dischi sarebbe composto da 10 dischi del sito A e 10 dischi del sito B.
3. Ogni set di unità su un dato sito viene configurato automaticamente come uno o più gruppi RAID DP o RAID-TEC completamente ridondanti, indipendentemente dall'utilizzo del mirroring. Questo utilizzo di RAID sottostante il mirroring garantisce la protezione dei dati anche dopo la perdita di un sito.



La figura precedente illustra una configurazione SyncMirror di esempio. È stato creato un aggregato di 24 dischi sul controller con 12 dischi da uno shelf allocato sul sito A e 12 dischi da uno shelf allocato sul sito B. I dischi sono stati raggruppati in due gruppi RAID con mirroring. Il gruppo RAID 0 include un plesso A 6 dischi sul sito A con mirroring su un plesso A 6 dischi sul sito B. Analogamente, il gruppo RAID 1 include un plesso A 6 dischi sul sito A con mirroring su un plesso A 6 dischi sul sito B.

Di norma, SyncMirror viene utilizzato per fornire il mirroring remoto con i sistemi MetroCluster, con una copia dei dati in ciascun sito. A volte, è stato utilizzato per fornire un livello di ridondanza extra in un unico sistema. In particolare, fornisce ridondanza a livello di shelf. Uno shelf di dischi contiene già doppi controller e alimentatori e nel complesso è poco più di una lamiera, ma in alcuni casi è consigliabile garantire una protezione extra. Ad esempio, un cliente NetApp ha implementato SyncMirror per una piattaforma mobile di analytics in tempo reale utilizzata durante i test nel settore automobilistico. Il sistema è stato separato in due rack fisici forniti con alimentatori indipendenti e sistemi UPS indipendenti.

Errore di ridondanza: NVFAIL

Come discusso in precedenza, una scrittura non viene riconosciuta fino a quando non è stata registrata nella

NVRAM locale e nella NVRAM su almeno un altro controller. Questo approccio garantisce che un guasto dell'hardware o un'interruzione di corrente non comporti la perdita dell'i/o in-flight Se si verifica un guasto nella NVRAM locale o nella connettività ad altri nodi, i dati non verranno più mirrorati.

Se la NVRAM locale riporta un errore, il nodo si arresta. Questo arresto determina il failover su un partner controller quando vengono utilizzate coppie ha. Con MetroCluster, il comportamento dipende dalla configurazione complessiva scelta, ma può portare al failover automatico della nota remota. In ogni caso, nessun dato viene perso perché il controller che subisce l'errore non ha confermato l'operazione di scrittura.

Un guasto di connettività site-to-site che blocca la replica NVRAM ai nodi remoti è una situazione più complicata. Le scritture non vengono più replicate sui nodi remoti, con la possibilità di perdita di dati in caso di errore catastrofico su un controller. Cosa più importante, il tentativo di failover su un nodo diverso in queste condizioni comporta una perdita di dati.

Il fattore di controllo è se la NVRAM è sincronizzata. Se la NVRAM è sincronizzata, il failover da nodo a nodo può procedere in tutta sicurezza senza rischio di perdita di dati. In una configurazione MetroCluster, se la NVRAM e i plessi degli aggregati sottostanti sono sincronizzati, è possibile procedere con lo switchover senza rischio di perdita di dati.

ONTAP non consente alcun failover o switchover quando i dati non sono sincronizzati, a meno che non sia forzato il failover o lo switchover. La forzatura di una modifica delle condizioni in questo modo riconosce che i dati potrebbero essere lasciati indietro nel controllore originale e che la perdita di dati è accettabile.

I database e altre applicazioni sono particolarmente vulnerabili al danneggiamento se un failover o uno switchover è forzato perché mantengono cache interne di dati su disco di dimensioni maggiori. In caso di failover o switchover forzato, le modifiche riconosciute in precedenza vengono eliminate del tutto. Il contenuto dell'array di storage torna indietro nel tempo e lo stato della cache non riflette più lo stato dei dati su disco.

Per evitare questa situazione, ONTAP consente di configurare i volumi per una protezione speciale contro i guasti della NVRAM. Quando attivato, questo meccanismo di protezione determina l'ingresso di un volume nello stato chiamato NVFAIL. Questo stato causa errori di i/o che causano un crash dell'applicazione. Questo blocco causa l'arresto delle applicazioni in modo che non utilizzino dati obsoleti. I dati non devono essere persi perché i dati delle transazioni devono essere presenti nei registri. Solitamente, gli amministratori dovranno arrestare completamente gli host prima di riportare manualmente LUN e volumi in linea. Sebbene queste fasi possano comportare un certo lavoro, questo approccio è il modo più sicuro per garantire l'integrità dei dati. Non tutti i dati richiedono questa protezione, motivo per cui il comportamento di NVFAIL può essere configurato in base al volume.

Coppie HA e MetroCluster

MetroCluster è disponibile in due configurazioni: Due nodi e coppia ha. La configurazione a due nodi si comporta come una coppia ha in relazione alla NVRAM. In caso di guasto improvviso, il nodo partner può riprodurre i dati della NVRAM per rendere i dischi coerenti e garantire che non vengano perse scritture riconosciute.

La configurazione ha-Pair replica la NVRAM anche sul nodo partner locale. Un semplice guasto al controller porta a un replay della NVRAM sul nodo partner, come nel caso di una coppia ha standalone, senza MetroCluster. In caso di improvvisa perdita completa del sito, il sito remoto dispone anche della NVRAM necessaria per rendere i dischi coerenti e iniziare a fornire i dati.

Un aspetto importante di MetroCluster è che i nodi remoti non hanno accesso ai dati partner in normali condizioni operative. Ogni sito funziona essenzialmente come un sistema indipendente che può assumere la personalità del sito opposto. Questo processo, noto come switchover, include uno switchover pianificato, in cui le operazioni del sito vengono migrate senza interruzioni nel sito opposto. Include anche le situazioni non pianificate in cui si perde un sito ed è necessario uno switchover manuale o automatico come parte del

disaster recovery.

Switchover e switchback

I termini switchover e switchback si riferiscono al processo di transizione dei volumi tra controller remoti in una configurazione MetroCluster. Questo processo si applica solo ai nodi remoti. Se viene utilizzato MetroCluster in una configurazione a quattro volumi, il failover di nodo locale utilizza il medesimo processo di takeover e giveback descritto in precedenza.

Switchover e switchback pianificati

Uno switchover o uno switchback pianificato è simile a un takeover o un giveback tra i nodi. Il processo prevede diverse fasi e potrebbe richiedere alcuni minuti, ma in realtà si tratta di una transizione graduale delle risorse di storage e di rete. Il momento in cui il trasferimento del controllo avviene molto più rapidamente del tempo richiesto per l'esecuzione del comando completo.

La differenza principale tra takeover/giveback e switchover/switchback influisce sulla connettività FC SAN. Grazie al takeover/giveback locale, un host subisce la perdita di tutti i percorsi FC nel nodo locale e si affida al proprio MPIO nativo per passare ai percorsi alternativi disponibili. Le porte non vengono ricollocate. Grazie a switchover e switchback, le porte di destinazione FC virtuali sui controller passano all'altro sito. Di fatto, smettono di esistere sulla SAN per un momento e ricompaiono su un controller alternativo.

Timeout SyncMirror

SyncMirror è una tecnologia di mirroring ONTAP che fornisce protezione dai guasti agli shelf. Quando gli shelf sono separati su una distanza, il risultato è una data Protection remota.

SyncMirror non fornisce mirroring sincrono universale. Il risultato è una maggiore disponibilità. Alcuni sistemi di archiviazione utilizzano un mirroring costante tutto o niente, talvolta chiamato modalità domino. Questa forma di mirroring è limitata nell'applicazione poiché tutte le attività di scrittura devono cessare se la connessione al sito remoto viene persa. Altrimenti, una scrittura esisterebbe in un sito ma non nell'altro. Generalmente, tali ambienti sono configurati per portare le LUN offline in caso di perdita della connettività sito-sito per più di un breve periodo (ad esempio 30 secondi).

Questo comportamento è desiderabile per un piccolo sottoinsieme di ambienti. Tuttavia, la maggior parte delle applicazioni richiede una soluzione che offra una replica sincrona garantita in normali condizioni operative, ma con la possibilità di sospendere la replica. Una perdita completa della connettività da sito a sito viene spesso considerata una situazione quasi disastrosa. Generalmente, tali ambienti vengono mantenuti online e forniscono dati fino al ripristino della connettività o alla decisione formale di arrestare l'ambiente per proteggere i dati. Un requisito per l'arresto automatico dell'applicazione solo a causa di un errore di replica remota è insolito.

SyncMirror supporta i requisiti di mirroring sincrono con la flessibilità di un timeout. Se la connettività al telecomando e/o al plex viene persa, inizia il conto alla rovescia un timer di 30 secondi. Quando il contatore raggiunge 0, l'elaborazione i/o in scrittura riprende a utilizzare i dati locali. La copia remota dei dati è utilizzabile, ma viene bloccata in tempo fino a quando non viene ripristinata la connettività. La risincronizzazione sfrutta le snapshot a livello di aggregato per riportare il sistema in modalità sincrona il più rapidamente possibile.

In particolare, in molti casi, questo tipo di replica universale in modalità domino a tutto o niente è meglio implementato a livello di applicazione. Ad esempio, Oracle DataGuard include la modalità di protezione massima, che garantisce la replica a lunga istanza in tutte le circostanze. Se il collegamento di replica non riesce per un periodo superiore a un timeout configurabile, i database vengono arrestati.

Switchover automatico senza intervento dell'utente con MetroCluster fabric-attached

Lo switchover automatico non assistito (ASOLO) è una funzione MetroCluster collegata al fabric che offre un tipo di ha cross-site. Come indicato in precedenza, MetroCluster è disponibile in due tipi: Un singolo controller su ciascun sito o una coppia ha su ciascun sito. Il vantaggio principale dell'opzione ha è che l'arresto pianificato o non pianificato del controller consente comunque a tutti gli i/o di essere locali. Il vantaggio dell'opzione a nodo singolo consiste nella riduzione di costi, complessità e infrastruttura.

Il valore primario di AUSO è migliorare le capacità ha dei sistemi MetroCluster fabric-attached. Ciascun sito esegue il monitoraggio dello stato di salute del sito opposto e, se non sono ancora presenti nodi che forniscono dati, AUDO esegue un rapido switchover. Questo approccio è particolarmente utile nelle configurazioni MetroCluster con un solo nodo per sito, perché consente di avvicinare la configurazione a una coppia ha in termini di disponibilità.

AUSO non è in grado di offrire un monitoraggio completo a livello di coppia ha. Una coppia ha può offrire una disponibilità estremamente elevata, perché include due cavi fisici ridondanti per la comunicazione diretta da nodo a nodo. Inoltre, entrambi i nodi di una coppia ha hanno accesso allo stesso set di dischi in loop ridondanti, offrendo un altro percorso a un nodo per monitorare la salute di un altro.

I cluster MetroCluster esistono tra i siti per i quali le comunicazioni nodo-nodo e l'accesso al disco si basano sulla connettività di rete site-to-site. La capacità di monitorare il battito cardiaco del resto del cluster è limitata. AUSO deve discriminare tra una situazione in cui l'altro sito è effettivamente inattivo piuttosto che non disponibile a causa di un problema di rete.

Di conseguenza, un controller in una coppia ha può richiedere un takeover se rileva un guasto del controller verificatosi per un motivo specifico, ad esempio un panico del sistema. Può anche richiedere un takeover in caso di perdita totale della connettività, talvolta nota come battito cardiaco perso.

Un sistema MetroCluster può eseguire uno switchover automatico in modo sicuro solo quando viene rilevato un guasto specifico nel sito originale. Inoltre, il controller che prende la proprietà del sistema di storage deve essere in grado di garantire che i dati su disco e NVRAM siano sincronizzati. Il controller non è in grado di garantire la sicurezza di uno switchover solo perché ha perso il contatto con il sito di origine, cosa che potrebbe essere ancora operativa. Per ulteriori opzioni per automatizzare uno switchover, vedere le informazioni sulla soluzione MetroCluster Tiebreaker (MCTB) nella sezione successiva.

Tiebreaker MetroCluster con MetroCluster fabric-attached

Il "Tiebreaker NetApp MetroCluster" software può essere eseguito su un terzo sito per monitorare lo stato dell'ambiente MetroCluster, inviare notifiche e, facoltativamente, imporre uno switchover in una situazione di emergenza. Una descrizione completa di Tiebreaker "Sito di supporto NetApp"è disponibile sul , ma lo scopo principale di MetroCluster Tiebreaker è quello di rilevare la perdita del sito. Inoltre, deve discriminare tra la perdita del sito e la perdita della connettività. Ad esempio, lo switchover non deve essere eseguito perché il tiebreaker non è riuscito a raggiungere il sito primario; questo spiega perché il tiebreaker monitora anche la capacità del sito remoto di contattare il sito primario.

Lo switchover automatico con AUSO è compatibile anche con l'MCTB. AUSO reagisce in modo molto rapido perché è progettato per rilevare eventi di errore specifici e quindi richiamare lo switchover solo quando i plex NVRAM e SyncMirror sono sincronizzati.

Al contrario, il Tiebreaker è localizzato a distanza e quindi deve attendere che un temporizzatore trascorra prima di dichiarare un sito morto. Il tiebreaker alla fine rileva il tipo di guasto del controller coperto da AUSO, ma in generale AUSO ha già avviato lo switchover e, eventualmente, ha completato lo switchover prima che il tiebreaker agisca. Il secondo comando switchover risultante proveniente dal tiebreaker verrebbe rifiutato.

Il software MCTB non verifica se NVRAM era e/o i plex sono sincronizzati quando si forza uno switchover. Lo switchover automatico, se configurato, deve essere disattivato durante le attività di manutenzione che causano una perdita di sincronizzazione dei plessi NVRAM o SyncMirror.

Inoltre, l'MCTB potrebbe non risolvere un disastro continuo che porta alla seguente sequenza di eventi:

1. La connettività tra i siti viene interrotta per più di 30 secondi.
2. Timeout della replica SyncMirror e proseguimento delle operazioni sul sito primario, lasciando inattiva la replica remota.
3. Il sito primario viene perso. Il risultato è la presenza di modifiche non replicate sul sito primario. Uno switchover potrebbe quindi essere indesiderato per una serie di motivi, tra cui:
    - I dati critici potrebbero essere presenti sul sito primario e quindi ripristinabili. Uno switchover che ha permesso all'applicazione di continuare a funzionare eliminava efficacemente i dati critici.
    - Un'applicazione sul sito rimasto che stava utilizzando le risorse di storage sul sito primario al momento della perdita del sito potrebbe avere memorizzato nella cache i dati. Uno switchover introdurrebbe una versione obsoleta dei dati che non corrisponde alla cache.
    - Un sistema operativo del sito rimasto che utilizzava le risorse di storage del sito primario al momento della perdita del sito potrebbe avere memorizzato i dati nella cache. Uno switchover introdurrebbe una versione obsoleta dei dati che non corrisponde alla cache. L'opzione più sicura è configurare tiebreaker in modo da inviare un avviso se rileva un guasto del sito e chiedere a una persona di decidere se forzare uno switchover. Potrebbe essere necessario arrestare le applicazioni e/o i sistemi operativi per cancellare i dati memorizzati nella cache. Inoltre, è possibile utilizzare le impostazioni NVFAIL per aggiungere ulteriore protezione e semplificare il processo di failover.

ONTAP Mediator con MetroCluster IP

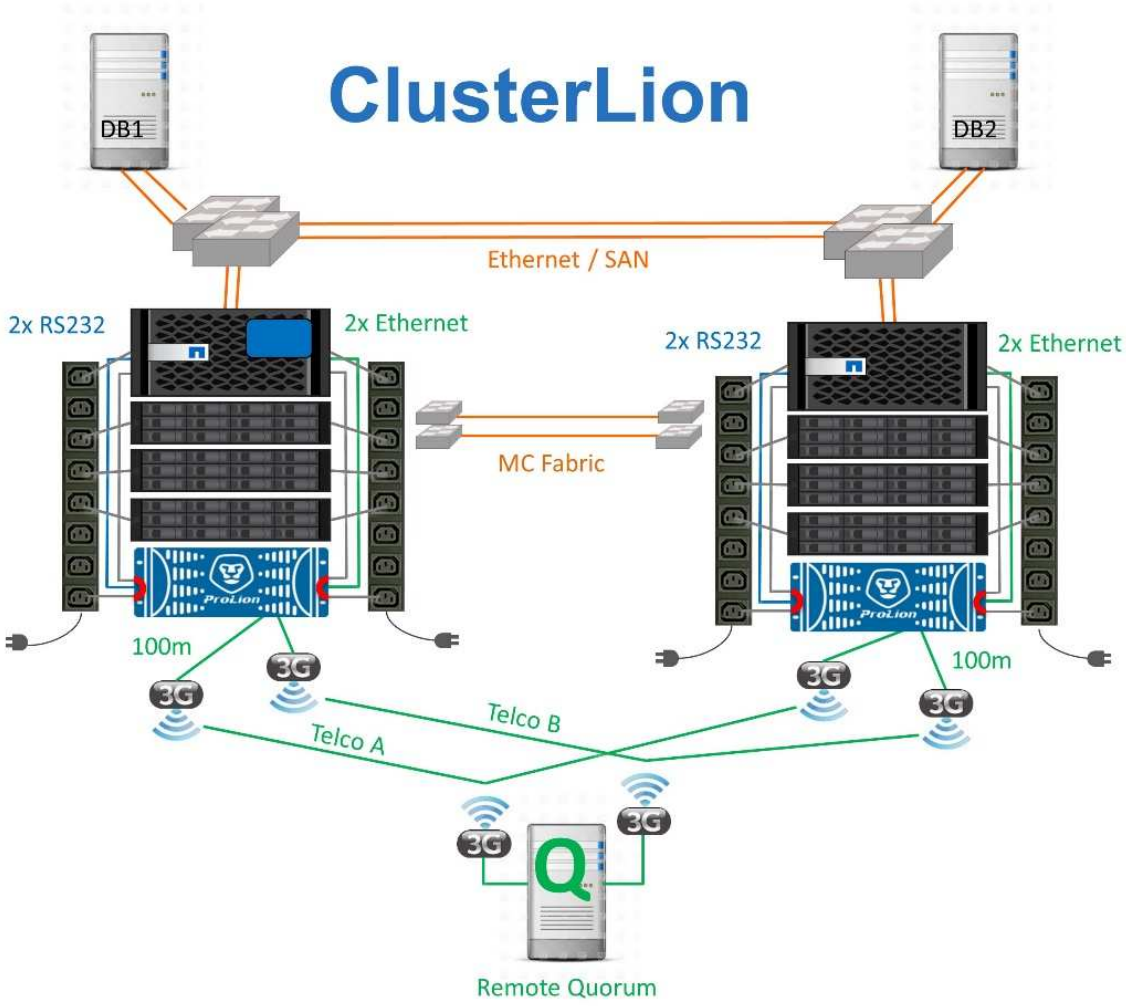
ONTAP Mediator viene utilizzato con MetroCluster IP e con alcune altre soluzioni ONTAP. Funziona come un servizio di tiebreaker tradizionale, proprio come il software MetroCluster Tiebreaker descritto in precedenza, ma include anche una funzione critica: Eseguire uno switchover automatizzato e non assistito.

Un MetroCluster fabric-attached ha accesso diretto ai dispositivi di storage del sito opposto. Ciò consente a un controller MetroCluster di monitorare lo stato degli altri controller leggendo i dati heartbeat dalle unità. In questo modo, un controller riconosce il guasto di un altro controller ed esegue uno switchover.

Al contrario, l'architettura IP di MetroCluster instrada tutti i/o esclusivamente attraverso la connessione controller-controller; non vi è accesso diretto ai dispositivi di storage sul sito remoto. Questo limita la possibilità per un controller di rilevare gli errori ed eseguire uno switchover. Pertanto, come dispositivo di tiebreaker occorre il ONTAP Mediator per rilevare la perdita di un sito ed eseguire automaticamente uno switchover.

Terzo sito virtuale con ClusterLion

ClusterLion è un'appliance di monitoraggio MetroCluster avanzata che funziona come un terzo sito virtuale. Questo approccio consente di implementare MetroCluster in maniera sicura in una configurazione a due siti con una funzionalità di switchover completamente automatizzata. Inoltre, ClusterLion può eseguire ulteriori operazioni di monitoraggio a livello di rete ed eseguire operazioni post-switchover. La documentazione completa è disponibile presso ProLion.



- Gli appliance ClusterLion monitorano lo stato dei controller con cavi Ethernet e seriali collegati direttamente.
- I due dispositivi sono collegati tra loro mediante connessioni wireless 3G ridondanti.
- L'alimentazione alla centralina ONTAP viene instradata attraverso i relè interni. In caso di guasto a un sito, ClusterLion, che contiene un sistema UPS interno, interrompe i collegamenti di alimentazione prima di richiamare uno switchover. Questo processo assicura che non si verifichi alcuna condizione split-brain.
- ClusterLion esegue uno switchover entro il timeout SyncMirror di 30 secondi o non lo esegue affatto.
- ClusterLion non esegue uno switchover a meno che gli stati della NVRAM e dei plex SyncMirror non siano sincronizzati.
- Poiché ClusterLion esegue uno switchover solo se MetroCluster è completamente sincronizzato, NVFAIL non è necessario. Questa configurazione consente ad ambienti che si estendono tra diversi siti, come un Oracle RAC esteso, di rimanere online anche durante uno switchover non pianificato.
- Il supporto include MetroCluster fabric-attached e MetroCluster IP

SyncMirror

La base della protezione dei dati di Oracle con un sistema MetroCluster è SyncMirror, una tecnologia di mirroring sincrono scale-out dalle performance massime.

Data Protection con SyncMirror

Al livello più semplice, la replica sincrona significa che qualsiasi modifica deve essere apportata a entrambi i lati dello storage con mirroring prima che venga riconosciuta. Ad esempio, se un database sta scrivendo un registro o un guest VMware viene aggiornato, non deve mai andare persa una scrittura. Come livello di protocollo, il sistema di storage non deve riconoscere la scrittura fino a quando non è stato assegnato a un supporto non volatile in entrambi i siti. Solo allora è sicuro procedere senza il rischio di perdita dei dati.

L'utilizzo di una tecnologia di replica sincrona è il primo passo nella progettazione e nella gestione di una soluzione di replica sincrona. La considerazione più importante è capire cosa potrebbe accadere durante i vari scenari di guasto pianificati e non pianificati. Non tutte le soluzioni di replica sincrona offrono le stesse funzionalità. Se hai bisogno di una soluzione che offra un recovery point objective (RPO) pari a zero, ovvero zero data loss, devi prendere in considerazione tutti gli scenari di guasto. In particolare, qual è il risultato previsto quando la replica è impossibile a causa della perdita di connettività tra i siti?

Disponibilità dei dati SyncMirror

La replica MetroCluster si basa sulla tecnologia NetApp SyncMirror, che è progettata per passare in modo efficiente dalla modalità sincrona alla modalità sincrona e viceversa. Questa funzionalità soddisfa i requisiti dei clienti che richiedono una replica sincrona, ma che hanno bisogno anche di un'alta disponibilità per i propri servizi dati. Ad esempio, se la connettività a un sito remoto viene interrotta, è generalmente preferibile che il sistema di archiviazione continui a funzionare in uno stato non replicato.

Molte soluzioni di replica sincrona sono in grado di funzionare solo in modalità sincrona. Questo tipo di replica "tutto o niente" viene talvolta chiamato modalità domino. Tali sistemi storage smettono di fornire i dati piuttosto che permettere che le copie locali e remote dei dati diventino non sincronizzate. Se la replica viene forzata, la risincronizzazione può richiedere molto tempo e lasciare un cliente esposto a una perdita di dati completa durante il tempo in cui il mirroring viene ristabilita.

Non solo SyncMirror può passare alla modalità sincrona senza problemi se il sito remoto non è raggiungibile, ma può anche risincronizzare rapidamente uno stato RPO = 0 al ripristino della connettività. La copia obsoleta dei dati nel sito remoto può anche essere preservata in uno stato utilizzabile durante la risincronizzazione, garantendo l'esistenza in ogni momento di copie locali e remote dei dati.

Quando è richiesta la modalità domino, NetApp offre SnapMirror Synchronous (SM-S). Esistono anche opzioni a livello di applicazione, come Oracle DataGuard o SQL Server Always on Availability Groups. Il mirroring del disco a livello del sistema operativo può essere opzionale. Per ulteriori informazioni e opzioni, consulta il tuo NetApp o il partner account team.

MetroCluster e NVFAIL

NVFAIL è una funzionalità generale di integrità dei dati di ONTAP progettata per massimizzare la protezione dell'integrità dei dati con i database.

> Questa sezione espande la spiegazione di ONTAP NVFAIL di base per affrontare argomenti specifici di MetroCluster.

Con MetroCluster, una scrittura non viene riconosciuta fino a quando non è stata registrata nella NVRAM locale e nella NVRAM su almeno un altro controller. Questo approccio garantisce che un guasto dell'hardware o un'interruzione di corrente non comporti la perdita dell'i/o in-flight Se si verifica un guasto nella NVRAM locale o nella connettività ad altri nodi, i dati non verranno più mirrorati.

Se la NVRAM locale riporta un errore, il nodo si arresta. Questo arresto determina il failover su un partner controller quando vengono utilizzate coppie ha. Con MetroCluster, il comportamento dipende dalla

configurazione complessiva scelta, ma può portare al failover automatico della nota remota. In ogni caso, nessun dato viene perso perché il controller che subisce l'errore non ha confermato l'operazione di scrittura.

Un guasto di connettività site-to-site che blocca la replica NVRAM ai nodi remoti è una situazione più complicata. Le scritture non vengono più replicate sui nodi remoti, con la possibilità di perdita di dati in caso di errore catastrofico su un controller. Cosa più importante, il tentativo di failover su un nodo diverso in queste condizioni comporta una perdita di dati.

Il fattore di controllo è se la NVRAM è sincronizzata. Se la NVRAM è sincronizzata, il failover da nodo a nodo può procedere in tutta sicurezza senza il rischio di perdita di dati. In una configurazione MetroCluster, se la NVRAM e i plessi degli aggregati sottostanti sono sincronizzati, è possibile effettuare lo switchover senza correre il rischio di perdita di dati.

ONTAP non consente alcun failover o switchover quando i dati non sono sincronizzati, a meno che non sia forzato il failover o lo switchover. La forzatura di una modifica delle condizioni in questo modo riconosce che i dati potrebbero essere lasciati indietro nel controllore originale e che la perdita di dati è accettabile.

I database sono particolarmente vulnerabili al danneggiamento se un failover o uno switchover è forzato, perché mantengono cache interne di dati su disco di dimensioni maggiori. In caso di failover o switchover forzato, le modifiche riconosciute in precedenza vengono eliminate del tutto. Il contenuto dell'array di storage torna indietro nel tempo e lo stato della cache del database non riflette più lo stato dei dati su disco.

Per proteggere le applicazioni da questa situazione, ONTAP consente di configurare i volumi per una protezione speciale contro gli errori della NVRAM. Quando attivato, questo meccanismo di protezione determina l'ingresso di un volume nello stato chiamato NVFAIL. Questo stato causa errori di i/o che causano l'arresto di un'applicazione in modo che non utilizzino dati obsoleti. I dati non devono essere persi perché eventuali scritture riconosciute sono ancora presenti nel sistema di storage e, nel caso dei database, tutti i dati delle transazioni con commit devono essere presenti nei registri.

Solitamente, gli amministratori dovranno arrestare completamente gli host prima di riportare manualmente LUN e volumi in linea. Sebbene queste fasi possano comportare un certo lavoro, questo approccio è il modo più sicuro per garantire l'integrità dei dati. Non tutti i dati richiedono questa protezione, motivo per cui il comportamento di NVFAIL può essere configurato in base al volume.

NVFAIL forzato manualmente

L'opzione più sicura per forzare uno switchover con un cluster di applicazioni (inclusi VMware, Oracle RAC e altri) distribuito tra i siti dipende da come specificato `-force-nvfail-all` alla riga di comando. Questa opzione è disponibile come misura di emergenza per assicurarsi che tutti i dati memorizzati nella cache vengano eliminati. Se un host utilizza risorse di storage situate originariamente nel sito colpito da disastro, riceve errori di i/o o un handle di file obsoleto (`ESTALE`). I database Oracle si arrestano in modo anomalo e i file system possono andare completamente offline o passare alla modalità di sola lettura.

Al termine dello switchover, il `in-nvfailed-state` Il flag deve essere cancellato e i LUN devono essere messi online. Al termine di questa attività, è possibile riavviare il database. È possibile automatizzare queste attività per ridurre l'RTO.

dr-force-nvfail

Come misura di sicurezza generale, impostare `dr-force-nvfail` contrassegnare tutti i volumi a cui è possibile accedere da un sito remoto durante le normali operazioni, ovvero si tratta di attività utilizzate prima del failover. Il risultato di questa impostazione è che i volumi remoti selezionati diventano non disponibili quando vengono immessi `in-nvfailed-state` durante uno switchover. Al termine dello switchover, il `in-nvfailed-state` Il flag deve essere cancellato e i LUN devono essere messi online. Al termine di queste attività, è possibile riavviare le applicazioni. È possibile automatizzare queste attività per ridurre l'RTO.

Il risultato è come usare l' `-force-nvfail-all` flag per commutatori manuali. Tuttavia, il numero di volumi interessati può essere limitato solo a quei volumi che devono essere protetti da applicazioni o sistemi operativi con cache obsolete.

Ci sono due requisiti critici per un ambiente che non utilizza `dr-force-nvfail` su volumi applicativi:

- Uno switchover forzato non deve avvenire più di 30 secondi dopo la perdita del sito primario.
- Lo switchover non deve essere eseguito durante le attività di manutenzione o in altre condizioni in cui i plex SyncMirror o la replica della NVRAM non sono sincronizzati. Il primo requisito può essere soddisfatto con il software tiebreaker configurato per eseguire uno switchover entro 30 secondi da un guasto del sito. Questo requisito non significa che lo switchover debba essere eseguito entro 30 secondi dal rilevamento di un guasto del sito. Ciò significa che non è più sicuro forzare uno switchover se sono trascorsi 30 secondi da quando un sito è stato confermato operativo.

Il secondo requisito può essere parzialmente soddisfatto disattivando tutte le funzionalità di switchover automatico quando la configurazione di MetroCluster non è sincronizzata. Un'opzione migliore è quella di disporre di una soluzione di tiebreaker in grado di monitorare lo stato di salute della replica NVRAM e dei plessi SyncMirror. Se il cluster non è completamente sincronizzato, il tiebreaker non deve attivare uno switchover.

Il software NetApp MCTB non è in grado di monitorare lo stato di sincronizzazione, pertanto deve essere disattivato quando MetroCluster non è sincronizzato per alcun motivo. ClusterLion include funzionalità di monitoraggio NVRAM e plex e può essere configurato in modo da non attivare lo switchover a meno che il sistema MetroCluster non sia confermato completamente sincronizzato.

Istanza singola di Oracle

Come indicato in precedenza, la presenza di un sistema MetroCluster non implica necessariamente l'aggiunta o la modifica delle Best practice per l'utilizzo di un database. La maggior parte dei database attualmente in esecuzione sui sistemi MetroCluster dei clienti è a singola istanza e segue le raccomandazioni contenute nella documentazione relativa a Oracle su ONTAP.

Failover con un sistema operativo preconfigurato

SyncMirror fornisce una copia sincrona dei dati nel sito di disaster recovery, ma per renderli disponibili sono necessari un sistema operativo e le applicazioni associate. L'automazione di base può migliorare notevolmente il tempo di failover dell'ambiente complessivo. I prodotti Clusterware come Veritas Cluster Server (VCS) vengono spesso utilizzati per creare un cluster in tutti i siti e in molti casi il processo di failover può essere guidato con semplici script.

In caso di perdita dei nodi primari, il clusterware (o gli script) viene configurato in modo da portare i database online nel sito alternativo. Un'opzione è creare server di standby preconfigurati per le risorse NFS o SAN che compongono il database. Se il sito primario non funziona, il clusterware o l'alternativa con script esegue una sequenza di azioni simile alle seguenti:

1. Forzare uno switchover su MetroCluster
2. Rilevamento di LUN FC (solo SAN)
3. Montaggio di file system e/o montaggio di gruppi di dischi ASM
4. Avvio del database

Il requisito principale di questo approccio è rappresentato da un sistema operativo in esecuzione sul sito remoto. Deve essere preconfigurato con i file binari di Oracle, il che significa anche che attività come l'applicazione delle patch Oracle devono essere eseguite sul sito primario e di standby. In alternativa, è possibile eseguire il mirroring dei file binari di Oracle nel sito remoto e montarli se viene dichiarato un disastro.

La procedura di attivazione effettiva è semplice. Comandi come il rilevamento delle LUN richiedono solo pochi comandi per ogni porta FC. Il montaggio del file system non è altro che un `mount` E sia i database che ASM possono essere avviati e arrestati dalla CLI con un unico comando. Se i volumi e i file system non vengono utilizzati nel sito di disaster recovery prima dello switchover, non è necessario impostare alcun requisito `dr-force- nvfail` sui volumi.

Failover con un sistema operativo virtualizzato

Il failover degli ambienti di database può essere esteso per includere il sistema operativo stesso. In teoria, questo failover può essere eseguito con le LUN di avvio, ma nella maggior parte dei casi con un sistema operativo virtualizzato. La procedura è simile ai seguenti passaggi:

1. Forzare uno switchover su MetroCluster
2. Montaggio dei datastore che ospitano le macchine virtuali del server di database
3. Avvio delle macchine virtuali
4. Avvio manuale dei database o configurazione delle macchine virtuali per avviare automaticamente i database, ad esempio, un cluster ESX può estendersi su diversi siti. In caso di disastro, dopo lo switchover, è possibile portare online le macchine virtuali nel sito di disaster recovery. Fino a quando i datastore che ospitano i database server virtualizzati non saranno in uso in occasione di un evento di emergenza, non sarà necessario impostare alcun valore `dr-force- nvfail` sui volumi associati.

Oracle Extended RAC

Molti clienti ottimizzano il proprio RTO estendendo un cluster Oracle RAC tra i vari siti, ottenendo una configurazione completamente Active-Active. La progettazione complessiva diventa più complicata perché deve includere la gestione del quorum di Oracle RAC. Inoltre, entrambi i siti accedono ai dati, il che significa che uno switchover forzato può portare all'utilizzo di una copia dei dati non aggiornata.

Sebbene una copia dei dati sia presente in entrambi i siti, solo il controller attualmente proprietario di un aggregato può fornire i dati. Pertanto, con i cluster RAC estesi, i nodi remoti devono eseguire l'i/o attraverso una connessione site-to-site. Il risultato è un'aggiunta di latenza i/o, ma generalmente questa latenza non rappresenta un problema. Anche la rete di interconnessione RAC deve essere estesa su più siti, il che significa che è comunque necessaria una rete ad alta velocità e a bassa latenza. Se la latenza aggiunta causa un problema, il cluster può essere azionato in maniera Active-passive. Quindi, le operazioni i/o-intensive devono essere indirizzate ai nodi RAC locali del controller proprietario degli aggregati. I nodi remoti eseguono quindi operazioni i/o più chiare o vengono utilizzati esclusivamente come server warm standby.

Se è richiesto un RAC esteso Active-Active, la sincronizzazione attiva di SnapMirror deve essere presa in considerazione al posto di MetroCluster. La replica SM-As consente di preferire una replica specifica dei dati. Pertanto, può essere integrato un cluster RAC esteso in cui tutte le letture avvengono localmente. Gli i/o in lettura non attraversano mai i siti, offrendo la minore latenza possibile. Tutte le attività di scrittura devono comunque transito sulla connessione tra siti, ma tale traffico è inevitabile con qualsiasi soluzione di mirroring sincrono.

Se con Oracle RAC vengono utilizzate LUN di avvio, compresi i dischi di avvio virtualizzati, `misscount` potrebbe essere necessario modificare il parametro. Per ulteriori informazioni sui parametri di timeout RAC, vedere "Oracle RAC con ONTAP".

Configurazione a due siti

Una configurazione RAC estesa a due siti può fornire servizi di database Active-Active che possono sopravvivere a molti scenari ma non a tutti.

File di voto RAC

La prima considerazione da prendere in considerazione per la distribuzione di RAC esteso su MetroCluster deve essere la gestione del quorum. Oracle RAC dispone di due meccanismi per gestire il quorum: Heartbeat del disco e heartbeat della rete. L'heartbeat del disco controlla l'accesso allo storage utilizzando i file di voto. Con una configurazione RAC a sito singolo, una singola risorsa di voto è sufficiente fintanto che il sistema storage sottostante offre funzionalità ha.

Nelle versioni precedenti di Oracle, i file di voto erano posizionati su dispositivi di archiviazione fisici, ma nelle versioni correnti di Oracle i file di voto sono memorizzati in gruppi di dischi ASM.

Oracle RAC è supportato con NFS. Durante il processo di installazione della griglia, viene creata una serie di processi ASM per presentare la posizione NFS utilizzata per i file della griglia come un gruppo di dischi ASM. Il processo è quasi trasparente per l'utente finale e non richiede alcuna gestione ASM continua al termine dell'installazione.

Il primo requisito di una configurazione a due siti è garantire che ogni sito possa sempre accedere a più della metà dei file di voto in modo da garantire un processo di disaster recovery senza interruzioni. Questa attività era semplice prima che i file di voto fossero memorizzati in gruppi di dischi ASM, ma oggi gli amministratori devono comprendere i principi di base della ridondanza ASM.

I gruppi di dischi ASM hanno tre opzioni di ridondanza `external`, `normal`, e. `high`. In altre parole, senza mirror, con mirroring e a 3 vie con mirroring. Un'opzione più recente chiamata `Flex` è anche disponibile, ma raramente utilizzato. Il livello di ridondanza e il posizionamento dei dispositivi ridondanti controllano ciò che accade negli scenari di errore. Ad esempio:

- Posizionamento dei file di votazione su un `diskgroup` con `external` la risorsa di ridondanza garantisce l'eliminazione di un sito se la connettività tra siti viene persa.
- Posizionamento dei file di votazione su un `diskgroup` con `normal` La ridondanza con un solo disco ASM per sito garantisce l'eliminazione dei nodi su entrambi i siti se la connettività tra i siti viene persa perché nessuno dei due siti dispone di un quorum di maggioranza.
- Posizionamento dei file di votazione su un `diskgroup` con `high` la ridondanza con due dischi su un sito e un singolo disco sull'altro sito consente operazioni active-active quando entrambi i siti sono operativi e reciprocamente raggiungibili. Tuttavia, se il sito a disco singolo è isolato dalla rete, il sito viene eliminato.

Heartbeat rete RAC

L'heartbeat della rete Oracle RAC monitora la raggiungibilità dei nodi in tutta l'interconnessione cluster. Per rimanere nel cluster, un nodo deve essere in grado di contattare più della metà degli altri nodi. In un'architettura a due siti, questo requisito crea le seguenti scelte per il numero di nodi RAC:

- Il posizionamento di un numero uguale di nodi per sito comporta l'espulsione in un sito nel caso in cui la connettività di rete venga persa.

- Il posizionamento di N nodi su un sito e N+1 nodi sul sito opposto garantisce che la perdita di connettività intersito determini nel sito con il maggior numero di nodi rimanenti nel quorum di rete e nel sito con meno nodi evicting.

Prima di Oracle 12cR2, non era fattibile controllare quale lato avrebbe subito un'eviction durante la perdita del sito. Quando ogni sito ha un numero uguale di nodi, l'evocazione è controllata dal nodo master, che in generale è il primo nodo RAC da avviare.

Oracle 12cR2 introduce la funzionalità di ponderazione dei nodi. Questa funzionalità consente agli amministratori di controllare in che modo Oracle risolve le condizioni split-brain. Ad esempio, il seguente comando imposta la preferenza per un nodo specifico in un RAC:

```
[root@host-a ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.
```

Dopo aver riavviato Oracle High-Availability Services, la configurazione si presenta come segue:

```
[root@host-a lib]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
```

Nodo `host-a` è ora designato come server critico. Se i due nodi RAC sono isolati, `host-a` sopravvive, e. `host-b` è sfrattato.

Per informazioni dettagliate, consultare il white paper Oracle "Panoramica tecnica su Oracle Clusterware 12c Release 2. "

Per le versioni di Oracle RAC precedenti a 12cR2, il nodo master può essere identificato controllando i registri CRS come segue:

```
[root@host-a ~]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
 [root@host-a ~]# grep -i 'master node' /grid/diag/crs/host-
a/crs/trace/crsd.trc
2017-05-04 04:46:12.261525 :   CRSSE:2130671360: {1:16377:2} Master Change
Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:01:24.979716 :   CRSSE:2031576832: {1:13237:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
2017-05-04 05:11:22.995707 :   CRSSE:2031576832: {1:13237:221} Master
Change Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:28:25.797860 :   CRSSE:3336529664: {1:8557:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
```

Questo registro indica che il nodo master è `2` e il nodo `host-a` Ha un ID di `1`. Questo significa che `host-a` non è il nodo master. L'identità del nodo master può essere confermata con il comando `olsnodes -n`.

```
[root@host-a ~]# /grid/bin/olsnodes -n
host-a  1
host-b  2
```

Il nodo con un ID di `2` è `host-b`, che è il nodo master. In una configurazione con un numero uguale di nodi su ogni sito, il sito con `host-b` è il sito che sopravvive se i due set perdono la connettività di rete per qualsiasi motivo.

È possibile che la voce di log che identifica il nodo master rimanga fuori dal sistema. In questa situazione, è possibile utilizzare i timestamp dei backup OCR (Oracle Cluster Registry).

```
[root@host-a ~]#  /grid/bin/ocrconfig -showbackup
host-b     2017/05/05 05:39:53     /grid/cdata/host-cluster/backup00.ocr
0
host-b     2017/05/05 01:39:53     /grid/cdata/host-cluster/backup01.ocr
0
host-b     2017/05/04 21:39:52     /grid/cdata/host-cluster/backup02.ocr
0
host-a     2017/05/04 02:05:36     /grid/cdata/host-cluster/day.ocr     0
host-a     2017/04/22 02:05:17     /grid/cdata/host-cluster/week.ocr     0
```

Questo esempio mostra che il nodo master è `host-b`. Indica anche una modifica nel nodo master da `host-a` a. `host-b` Da qualche parte tra il 2:05 e il 21:39 maggio 4. Questo metodo di identificazione del nodo master è sicuro da utilizzare solo se sono stati controllati anche i log CRS, poiché è possibile che il nodo master sia

cambiato dal precedente backup OCR. Se questa modifica si è verificata, dovrebbe essere visibile nei registri OCR.

La maggior parte dei clienti sceglie un singolo gruppo di dischi di voto che gestisce l'intero ambiente e un numero uguale di nodi RAC su ciascun sito. Il gruppo di dischi deve essere collocato nel sito che contiene il database. Il risultato è che la perdita di connettività provoca sfratto sul sito remoto. Il sito remoto non dispone più del quorum né avrebbe accesso ai file di database, ma il sito locale continua a funzionare normalmente. Quando la connettività viene ripristinata, l'istanza remota può essere riportata nuovamente in linea.

In caso di emergenza, è necessario uno switchover per portare online i file di database e il gruppo di dischi di voto sul sito rimasto. Se il disastro consente AD AUSO di attivare lo switchover, NVFAIL non viene attivato perché il cluster è sincronizzato e le risorse di storage vengono normalmente online. AUSO è un'operazione molto veloce e dovrebbe essere completata prima del `disktimeout` il periodo scade.

Poiché ci sono solo due siti, non è possibile utilizzare alcun tipo di software di rottura automatica esterna, il che significa che lo switchover forzato deve essere un'operazione manuale.

Configurazioni a tre siti

Un cluster RAC esteso è molto più semplice da progettare con tre siti. I due siti che ospitano ciascuna metà del sistema MetroCluster supportano anche i carichi di lavoro del database, mentre il terzo sito funge da tiebreaker sia per il database che per il sistema MetroCluster. La configurazione di Oracle Tiebreaker può essere semplice come collocare un membro del gruppo di dischi ASM utilizzato per il voto su un sito 3rd e può anche includere un'istanza operativa sul sito 3rd per garantire che vi sia un numero dispari di nodi nel cluster RAC.

> Per informazioni importanti sull'utilizzo di NFS in una configurazione RAC estesa, consultare la documentazione Oracle relativa al "gruppo di errori del quorum". In sintesi, potrebbe essere necessario modificare le opzioni di montaggio NFS per includere l'opzione soft per garantire che la perdita di connettività alle risorse quorum di hosting del sito 3rd non blocchi i server Oracle primari o i processi Oracle RAC.

Sincronizzazione attiva di SnapMirror

Panoramica

SnapMirror Active Sync ti consente di creare ambienti di database Oracle a disponibilità ultra elevata in cui le LUN sono disponibili da due cluster di storage diversi.

Con la sincronizzazione attiva di SnapMirror, non vi è alcuna copia "primaria" e "secondaria" dei dati. Ogni cluster può fornire i/o in lettura dalla propria copia locale dei dati e ciascun cluster replicherà una scrittura al partner. Il risultato è un comportamento io simmetrico.

Tra le altre opzioni, questo consente di eseguire Oracle RAC come cluster esteso con istanze operative su entrambi i siti. In alternativa, è possibile creare cluster di database attivi-passivi RPO=0 in cui è possibile spostare i database a singola istanza tra i siti durante un'interruzione del sito e questo processo può essere automatizzato tramite prodotti come Pacemaker o VMware ha. La base di queste opzioni è la replica sincrona gestita dalla sincronizzazione attiva di SnapMirror.

Replica sincrona

Durante il funzionamento normale, la sincronizzazione attiva di SnapMirror fornisce sempre RPO=0 replica sincrona, con un'eccezione. Se i dati non possono essere replicati, ONTAP rilascerà il requisito di replicare i dati e riprendere la distribuzione io su un sito mentre le LUN dell'altro sito vengono portate offline.

Hardware per lo storage

A differenza di altre soluzioni di disaster recovery per lo storage, SnapMirror Active Sync offre una flessibilità asimmetrica della piattaforma. Non è necessario che l'hardware di ciascun sito sia identico. Questa funzionalità consente di dimensionare correttamente l'hardware utilizzato per supportare la sincronizzazione attiva di SnapMirror. Il sistema di storage remoto può essere identico al sito primario se deve supportare un carico di lavoro di produzione completo, ma se un disastro determina una riduzione dell'i/o, rispetto a un sistema più piccolo nel sito remoto potrebbe risultare più conveniente.

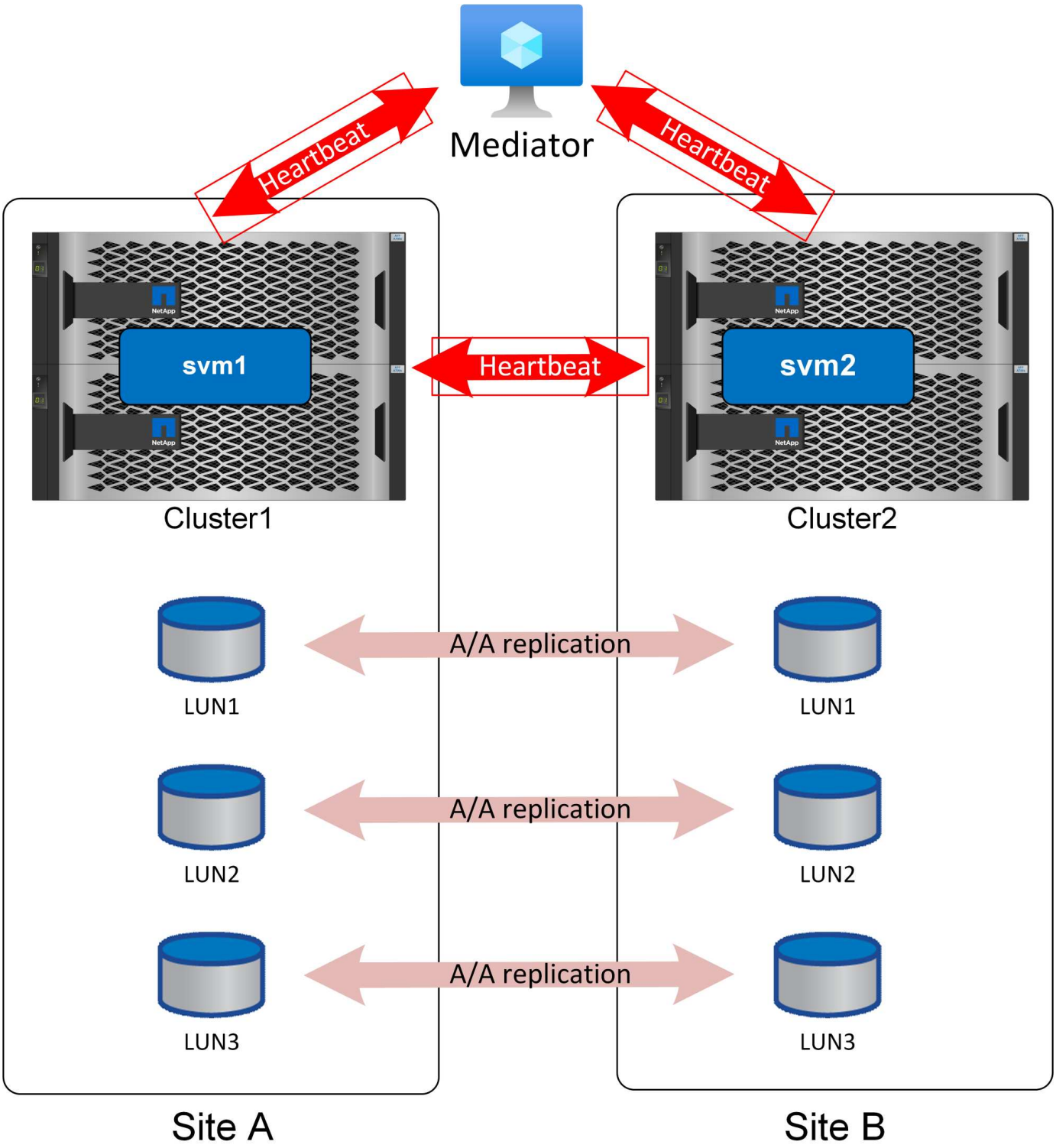
Mediatore ONTAP

ONTAP Mediator è un'applicazione software che viene scaricata dal supporto NetApp e che viene in genere distribuita su una piccola macchina virtuale. Il ONTAP Mediator non è un Tiebreaker quando viene utilizzato con la sincronizzazione attiva di SnapMirror. È un canale di comunicazione alternativo per i due cluster che partecipano alla replica sincrona attiva SnapMirror. Le operazioni automatizzate sono gestite da ONTAP in base alle risposte ricevute dal partner tramite connessioni dirette e tramite il mediatore.

Mediatore ONTAP

Il mediatore è necessario per automatizzare in modo sicuro il failover. Idealmente, sarebbe posizionato su un sito 3rd indipendente, ma può comunque funzionare per la maggior parte delle esigenze se colocated con uno dei cluster che partecipano alla replica.

Il mediatore non è realmente un tiebreaker, anche se quella è effettivamente la funzione che fornisce. Non esegue alcuna azione, ma fornisce un canale di comunicazione alternativo per la comunicazione tra cluster e cluster.



La sfida #1 con il failover automatizzato è il problema split-brain, e questo problema sorge se i due siti perdono la connettività tra loro. Che cosa dovrebbe accadere? Non si desidera che due siti diversi si designino come copie sopravvissute dei dati, ma in che modo un singolo sito può distinguere tra la perdita effettiva del sito opposto e l'impossibilità di comunicare con il sito opposto?

Qui entra il mediatore nell'immagine. Se si trova in un sito 3rd e ciascun sito dispone di una connessione di rete separata per tale sito, è disponibile un percorso aggiuntivo per ciascun sito per convalidare lo stato dell'altro. Esaminare nuovamente l'immagine sopra riportata e considerare i seguenti scenari.

- Cosa succede se il mediatore non riesce o è irraggiungibile da uno o entrambi i siti?
  - I due cluster possono ancora comunicare tra loro sullo stesso link utilizzato per i servizi di replica.
  - I dati sono ancora serviti con protezione RPO=0/7
- Cosa succede se il sito A non funziona?
  - Il sito B vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito B sostituirà i servizi dati, ma senza RPO = mirroring 0:1
- Cosa succede se il sito B non funziona?
  - Il sito A vedrà che entrambi i canali di comunicazione si interrompono.
  - Il sito A sostituirà i servizi dati, ma senza RPO = mirroring 0:1

Esiste un altro scenario da considerare: La perdita del collegamento di replica dei dati. In caso di perdita del link di replica tra i siti, RPO=0 Mirroring sarà ovviamente impossibile. Che cosa dovrebbe accadere allora?

Questo è controllato dallo stato del sito preferito. In una relazione SM-AS, uno dei siti è secondario all'altro. Questo non ha alcun effetto sulle normali operazioni e tutto l'accesso ai dati è simmetrico, ma se la replica viene interrotta, il legame dovrà essere interrotto per riprendere le operazioni. Ne risulta che il sito preferito continuerà le operazioni senza mirroring e il sito secondario interromperà l'elaborazione io fino al ripristino della comunicazione di replica.

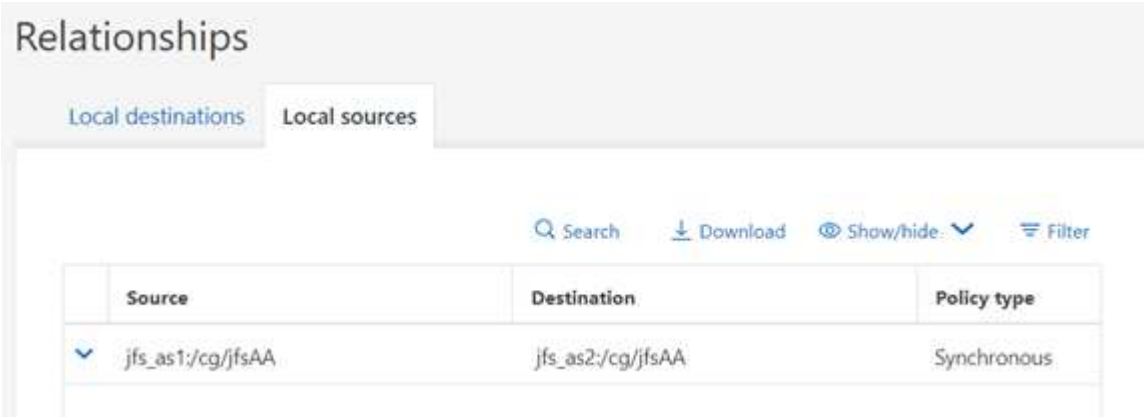
Sito preferito sincronizzazione attiva SnapMirror

Il comportamento di sincronizzazione attiva di SnapMirror è simmetrico, con una eccezione importante: Configurazione del sito preferita.

La sincronizzazione attiva di SnapMirror considererà un sito la "fonte" e l'altro la "destinazione". Ciò implica una relazione di replica unidirezionale, ma ciò non si applica al comportamento io. La replica è bidirezionale e simmetrica, mentre i tempi di risposta io sono identici su entrambi i lati del mirror.

La `source` designazione è controlla il sito preferito. In caso di perdita del link di replica, i percorsi delle LUN nella copia di origine continueranno a fornire i dati mentre i percorsi delle LUN nella copia di destinazione non saranno disponibili finché la replica non viene ristabilita e SnapMirror ritorna allo stato sincrono. I percorsi riprenderanno a fornire i dati.

La configurazione di origine/destinazione può essere visualizzata tramite SystemManager:



O all'interfaccia CLI:

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                     Destination Path: jfs_as2:/cg/jfsAA
                    Relationship Type: XDP
                Relationship Group Type: consistencygroup
                    SnapMirror Schedule: -
                 SnapMirror Policy Type: automated-failover-duplex
                      SnapMirror Policy: AutomatedFailOverDuplex
                            Tries Limit: -
                      Throttle (KB/sec): -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

La chiave è che source è l'SVM su cluster1. Come menzionato sopra, i termini "origine" e "destinazione" non descrivono il flusso di dati replicati. Entrambi i siti possono elaborare una scrittura e replicarla nel sito opposto. In effetti, entrambi i cluster sono origini e destinazioni. L'effetto della designazione di un cluster come origine controlla semplicemente quale cluster sopravvive come sistema di storage in lettura e scrittura in caso di perdita del link di replica.

Topologia di rete

Accesso uniforme

Una rete ad accesso uniforme significa che gli host sono in grado di accedere ai percorsi su entrambi i siti (o ai domini di errore all'interno dello stesso sito).

Una caratteristica importante di SM-AS è la capacità di configurare i sistemi storage per sapere dove si trovano gli host. Quando si mappano i LUN a un determinato host, è possibile indicare se sono prossimali o meno a un determinato sistema di archiviazione.

Impostazioni di prossimità

La prossimità si riferisce a una configurazione per cluster che indica che un determinato WWN host o ID iniziatore iSCSI appartiene a un host locale. Si tratta di un secondo passo opzionale per la configurazione dell'accesso LUN.

Il primo passo è la normale configurazione di igroup. Ogni LUN deve essere mappato a un igroup che contiene gli ID WWN/iSCSI degli host che devono accedere a quel LUN. Controlla quale host ha *accesso* a un LUN.

Il secondo passo, opzionale, consiste nel configurare la prossimità all'host. Questo non controlla l'accesso, controlla *priority*.

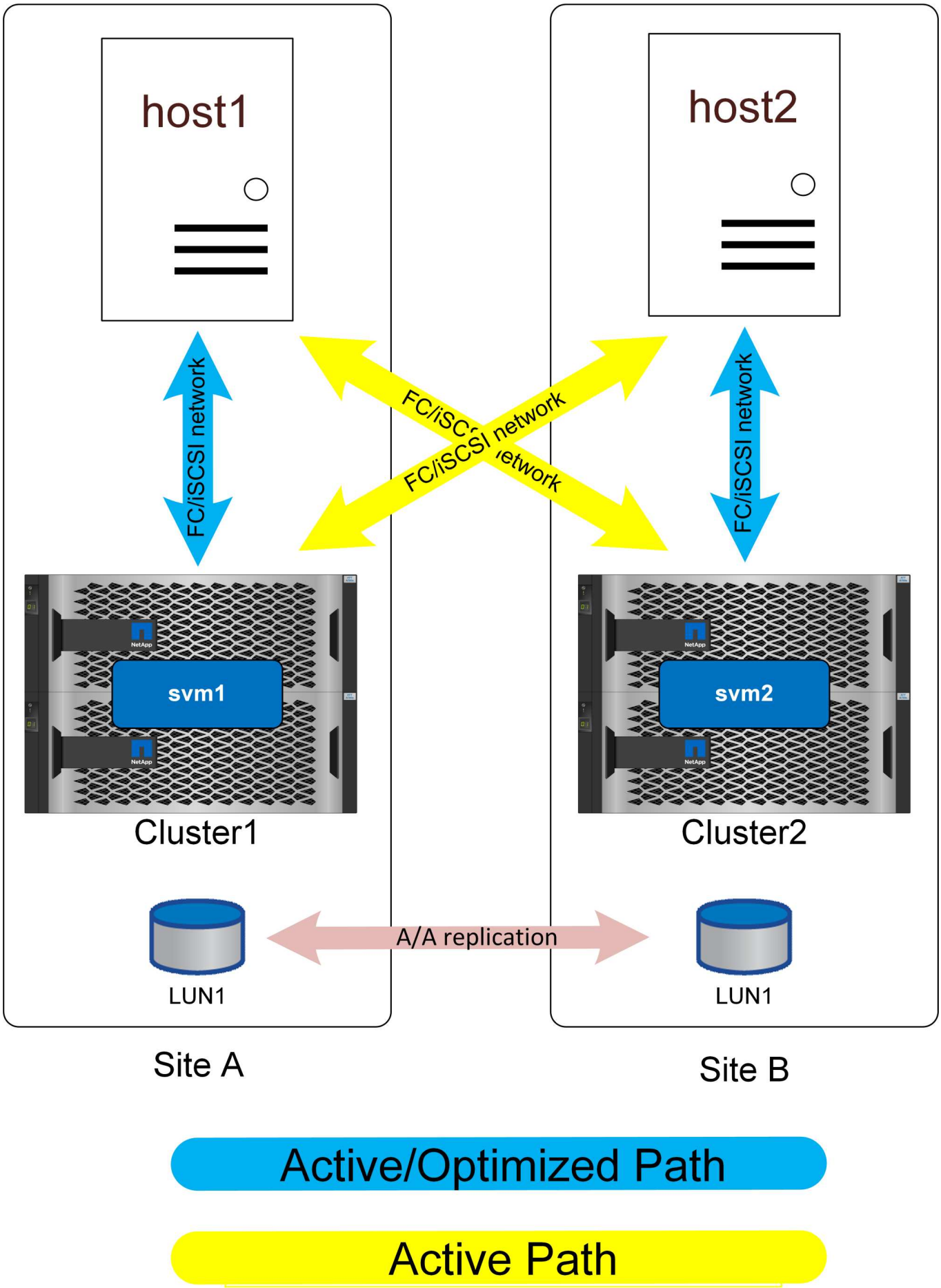
Ad esempio, un host del sito A potrebbe essere configurato in modo da accedere a una LUN protetta dalla sincronizzazione attiva di SnapMirror e, poiché la SAN è estesa tra i siti, i percorsi sono disponibili per tale LUN utilizzando lo storage sul sito A o lo storage sul sito B.

Senza impostazioni di prossimità, l'host utilizzerà entrambi i sistemi storage allo stesso modo perché entrambi i sistemi storage pubblicizzeranno i percorsi attivi/ottimizzati. Se la latenza SAN e/o la larghezza di banda tra i siti è limitata, questa operazione potrebbe non essere disattivabile e potrebbe essere necessario assicurarsi

che durante il normale funzionamento ogni host utilizzi preferenzialmente i percorsi verso il sistema di storage locale. Viene configurato aggiungendo l'ID WWN/iSCSI dell'host al cluster locale come host prossimale. Questa operazione può essere eseguita dalla CLI o da SystemManager.

AFF

Con un sistema AFF, i percorsi vengono visualizzati come mostrato di seguito quando è stata configurata la prossimità dell'host.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
svm2
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

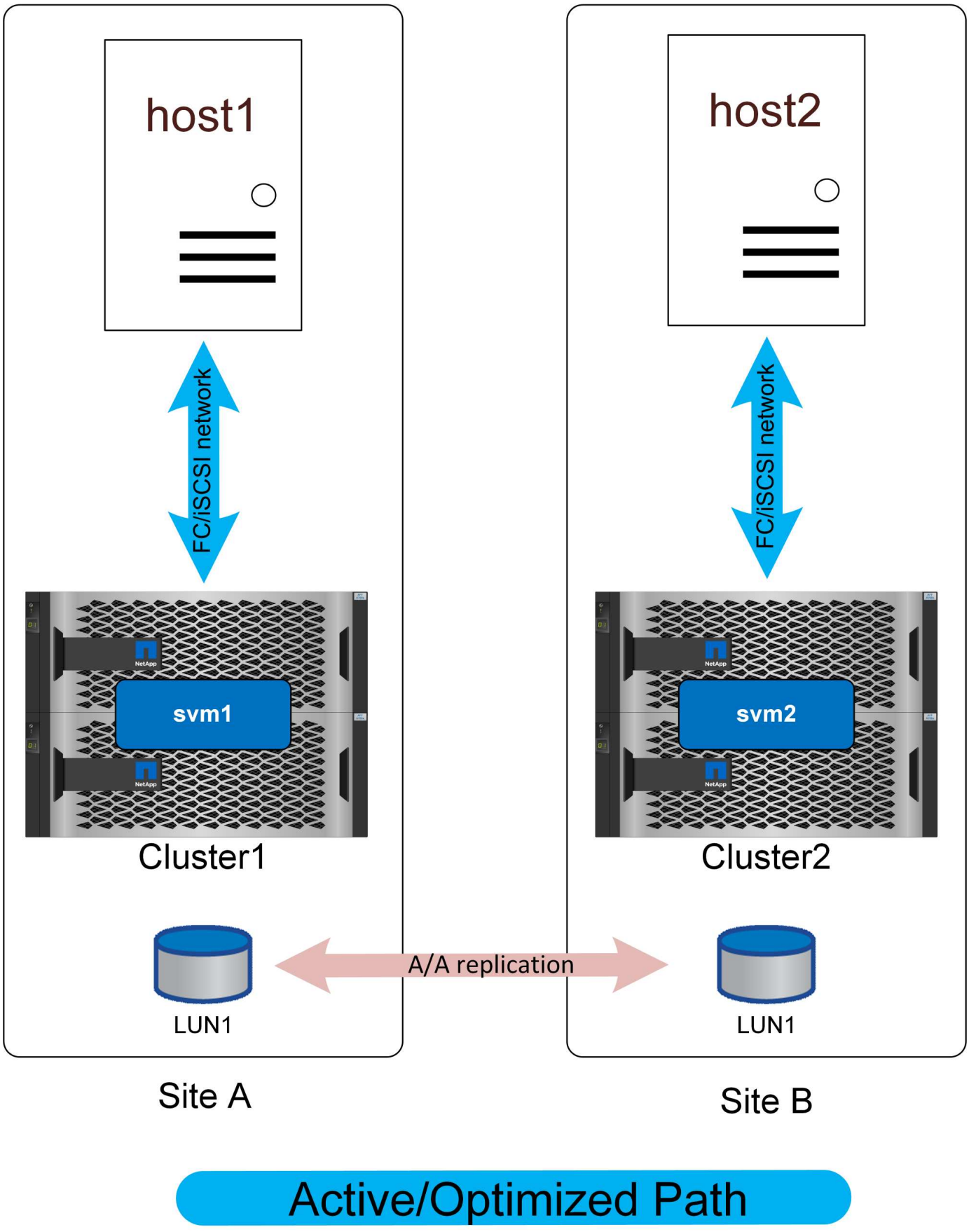
Durante le normali operazioni, tutti i io sono i/o locali. Le letture e le scritture sono gestite dallo storage array locale. Gli io in scrittura, naturalmente, dovranno anche essere replicati dal controller locale sul sistema remoto prima di essere riconosciuti, ma tutti gli io in lettura saranno serviti localmente e non subiranno una latenza aggiuntiva attraverso il collegamento SAN tra i siti.

L'unica volta in cui verranno utilizzati i percorsi non ottimizzati è quando tutti i percorsi attivi/ottimizzati vengono persi. Ad esempio, se l'intero array sul sito A perde energia, gli host sul sito A sarebbero comunque in grado di accedere ai percorsi dell'array sul sito B e di rimanere quindi operativi, anche se sperimenterebbero una latenza più elevata.

Esistono percorsi ridondanti attraverso il cluster locale non mostrati in questi diagrammi per semplicità. I sistemi di storage ONTAP sono ad alta disponibilità, pertanto un guasto a un controller non dovrebbe causare guasti nel sito. Ciò dovrebbe comportare solo una modifica dei percorsi locali utilizzati nel sito interessato.

ASA

I sistemi NetApp ASA offrono multipathing Active-Active su tutti i percorsi di un cluster. Questo vale anche per le configurazioni SM-AS.



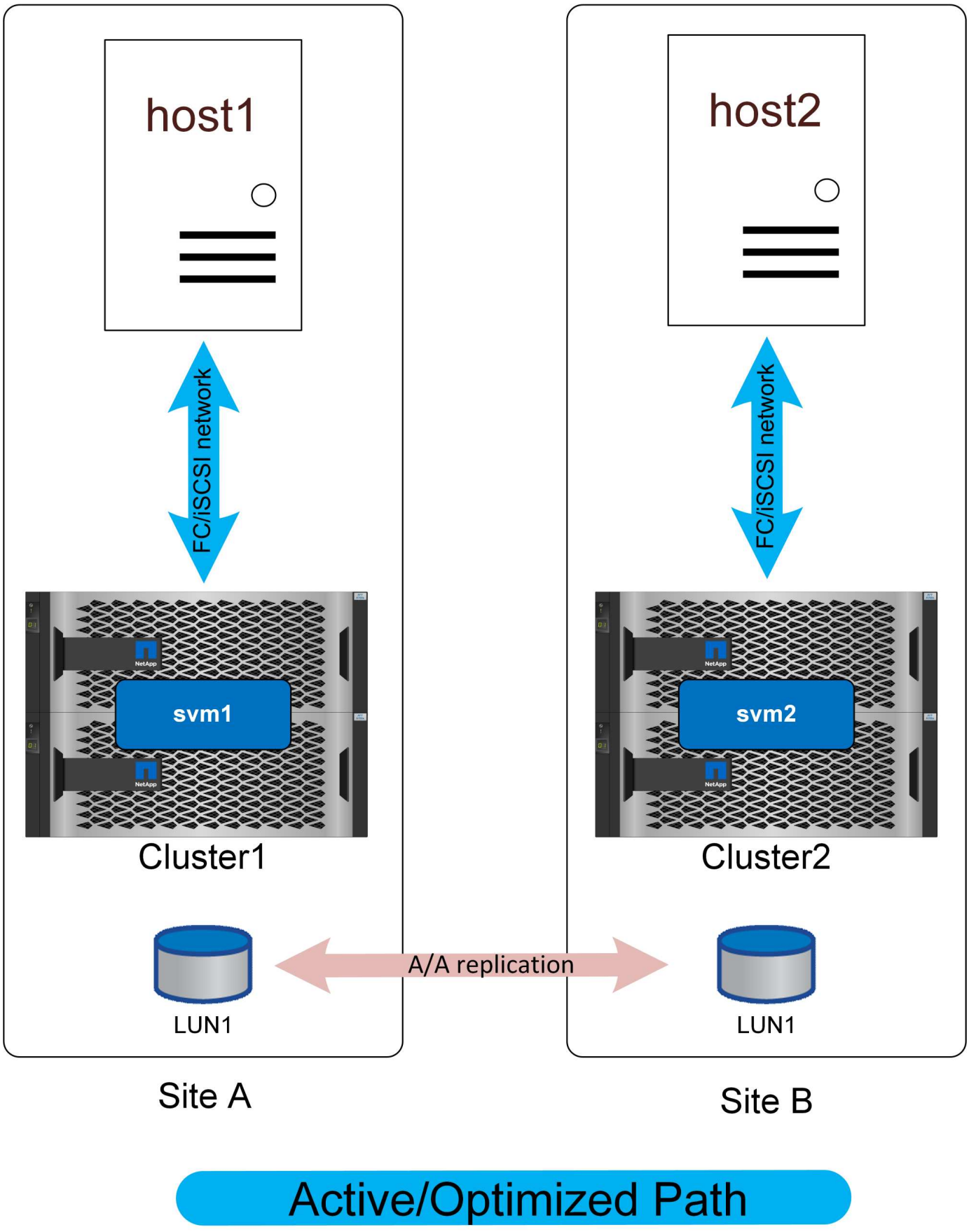
Una configurazione ASA con accesso non uniforme funzionerebbe in gran parte allo stesso modo di AFF. Con un accesso uniforme, io attraverserebbe la WAN. Ciò può essere o non può essere desiderabile.

Se i due siti fossero a una distanza di 100 metri con connettività in fibra non dovrebbe esserci una latenza aggiuntiva rilevabile che attraversa la WAN, ma se i siti fossero a lunga distanza gli uni dagli altri, le performance in lettura risulterebbero compromesse su entrambi i siti. Al contrario, con AFF questi percorsi WAN sarebbero utilizzati solo se non ci fossero percorsi locali disponibili e le performance quotidiane sarebbero migliori, perché tutti i io sarebbero io locali. ASA con rete di accesso non uniforme sarebbe un'opzione per ottenere i vantaggi in termini di costi e funzionalità di ASA senza incorrere in penalizzazioni per l'accesso alla latenza tra siti.

ASA con SM-as in una configurazione a bassa latenza offre due benefici interessanti. In primo luogo, essenzialmente **raddoppia** le prestazioni per ogni singolo host perché io può essere gestito dal doppio dei controller utilizzando il doppio dei percorsi. In secondo luogo, in un ambiente a sito singolo offre una disponibilità estrema, perché è possibile perdere un intero sistema storage senza interrompere l'accesso degli host.

Accesso non uniforme

Una rete di accesso non uniforme significa che ogni host ha solo accesso alle porte sul sistema di storage locale. LA SAN non è estesa a più siti (o presenta errori nei domini dello stesso sito).



Il vantaggio principale di questo approccio è la semplicità DELLE SAN, eliminando l'esigenza di stretching di una SAN via rete. Alcuni clienti non dispongono di connettività a latenza sufficientemente bassa tra i siti o non

dispongono dell'infrastruttura per il tunnel del traffico FC SAN su una rete intersito.

Lo svantaggio legato all'accesso non uniforme è che alcuni scenari di errore, inclusa la perdita del collegamento di replica, provocheranno la perdita dell'accesso allo storage da parte di alcuni host. In caso di interruzione della connettività dello storage locale, le applicazioni eseguite come istanze singole, come ad esempio i database non in cluster, eseguiti in maniera intrinseca solo su un singolo host in uno qualsiasi dei supporti di montaggio, si guasterebbero. I dati sarebbero comunque protetti, ma il server di database non avrebbe più accesso. Dovrebbe essere riavviato su un sito remoto, preferibilmente tramite un processo automatizzato. Ad esempio, VMware ha è in grado di rilevare una situazione di tutti i percorsi verso l'esterno su un server e di riavviare una macchina virtuale su un altro server in cui sono disponibili i percorsi.

Al contrario, un'applicazione in cluster come Oracle RAC può offrire un servizio disponibile contemporaneamente in due siti diversi. Perdere un sito non significa perdere il servizio dell'applicazione nel suo complesso. Le istanze sono ancora disponibili e in esecuzione nel sito sopravvissuto.

In molti casi, l'overhead della latenza aggiuntivo di un'applicazione che accede allo storage attraverso un collegamento da sito a sito sarebbe inaccettabile. Ciò significa che la migliore disponibilità di una rete uniforme è minima, poiché la perdita di storage su un sito comporterebbe comunque la necessità di arrestare i servizi sul sito in cui si è verificato l'errore.

> Esistono percorsi ridondanti attraverso il cluster locale non mostrati in questi diagrammi per semplicità. I sistemi di storage ONTAP sono ad alta disponibilità, pertanto un guasto a un controller non dovrebbe causare guasti nel sito. Ciò dovrebbe comportare solo una modifica dei percorsi locali utilizzati nel sito interessato.

Configurazioni Oracle

Panoramica

L'utilizzo della sincronizzazione attiva di SnapMirror non aggiunge o modifica necessariamente le procedure consigliate per il funzionamento di un database.

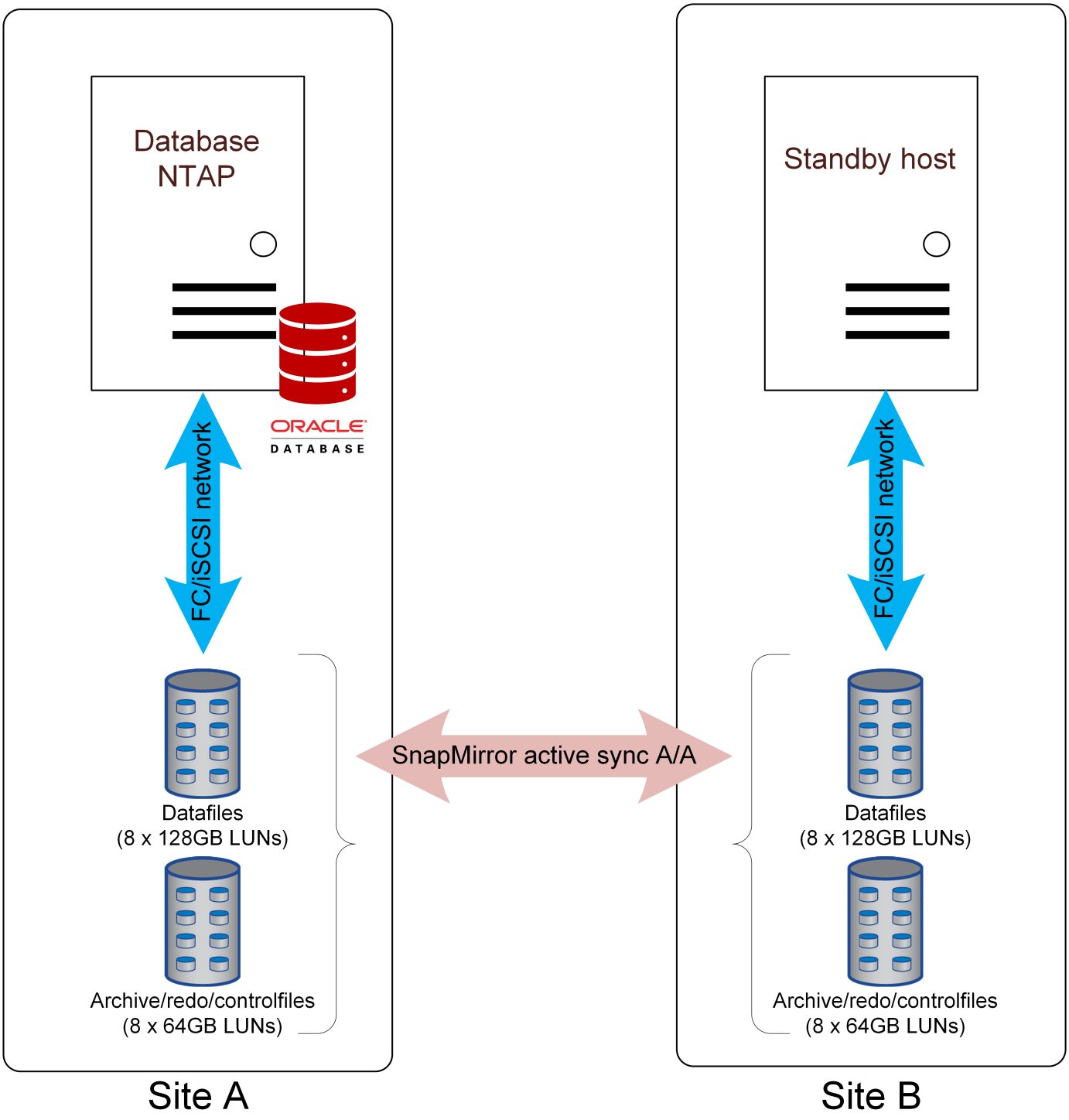
La migliore architettura dipende dai requisiti di business. Ad esempio, se l'obiettivo è ottenere una protezione RPO=0:1 contro la perdita di dati, ma l'RTO non è più così semplice, utilizzare i database Oracle Single Instance e replicare le LUN con SM, potrebbe essere sufficiente così come meno costoso da un approccio basato su licenze Oracle. Un guasto del sito remoto non interromperebbe le operazioni e, di conseguenza, la perdita del sito primario porterebbe all'utilizzo di LUN online e pronte per l'uso nel sito rimasto.

Se l'RTO fosse più rigoroso, l'automazione Active-passive di base tramite script o cluster come Pacemaker o Ansible migliorerebbe i tempi di failover. Ad esempio, VMware ha può essere configurato in modo da rilevare un guasto della VM nel sito primario e rendere attiva la VM nel sito remoto.

Infine, per un failover estremamente rapido, Oracle RAC può essere implementato su più siti. L'RTO sarebbe praticamente pari a zero perché il database sarebbe sempre online e disponibile su entrambi i siti.

Istanza singola Oracle

Gli esempi illustrati di seguito mostrano alcune delle numerose opzioni per la distribuzione dei database Oracle Single Instance con la replica di sincronizzazione attiva SnapMirror.



Failover con un sistema operativo preconfigurato

La sincronizzazione attiva SnapMirror fornisce una copia sincrona dei dati nel sito di disaster recovery, ma la loro disponibilità richiede un sistema operativo e le applicazioni associate. L'automazione di base può migliorare notevolmente il tempo di failover dell'ambiente complessivo. I prodotti Clusterware come Pacemaker vengono spesso utilizzati per creare un cluster in tutti i siti, e in molti casi il processo di failover può essere guidato con semplici script.

In caso di perdita dei nodi primari, il clusterware (o gli script) porterà i database online nel sito alternativo. Un'opzione è creare server di standby preconfigurati per le risorse SAN che compongono il database. Se il sito primario non funziona, il clusterware o l'alternativa con script esegue una sequenza di azioni simile alle

seguenti:

1. Rileva i guasti del sito primario
2. Rilevamento di LUN FC o iSCSI
3. Montaggio di file system e/o montaggio di gruppi di dischi ASM
4. Avvio del database

Il requisito principale di questo approccio è rappresentato da un sistema operativo in esecuzione sul sito remoto. Deve essere preconfigurato con i file binari di Oracle, il che significa anche che attività come l'applicazione delle patch Oracle devono essere eseguite sul sito primario e di standby. In alternativa, è possibile eseguire il mirroring dei file binari di Oracle nel sito remoto e montarli se viene dichiarato un disastro.

La procedura di attivazione effettiva è semplice. Comandi come il rilevamento delle LUN richiedono solo pochi comandi per ogni porta FC. Il montaggio del file system non è altro che un `mount` comando e sia i database che ASM possono essere avviati e arrestati dalla CLI con un unico comando.

Failover con un sistema operativo virtualizzato

Il failover degli ambienti di database può essere esteso per includere il sistema operativo stesso. In teoria, questo failover può essere eseguito con le LUN di avvio, ma nella maggior parte dei casi con un sistema operativo virtualizzato. La procedura è simile ai seguenti passaggi:

1. Rileva i guasti del sito primario
2. Montaggio dei datastore che ospitano le macchine virtuali del server di database
3. Avvio delle macchine virtuali
4. Avviare i database manualmente o configurare le macchine virtuali per avviare automaticamente i database.

Ad esempio, un cluster ESX può estendersi su diversi siti. In caso di disastro, dopo lo switchover, è possibile portare online le macchine virtuali nel sito di disaster recovery.

Protezione dai guasti dello storage

Il diagramma precedente mostra l'utilizzo di "accesso non uniforme", in cui la SAN non è estesa tra i siti. Questa configurazione potrebbe essere più semplice e, in alcuni casi, potrebbe essere l'unica opzione data le attuali funzionalità SAN, ma significa anche che un guasto del sistema di storage primario causerebbe un'interruzione del database fino a quando l'applicazione non è stata sottoposta a failover.

Per una maggiore resilienza, la soluzione potrebbe essere implementata con "accesso uniforme". Ciò consentirebbe alle applicazioni di continuare a funzionare utilizzando i percorsi pubblicizzati dal sito opposto.

Oracle Extended RAC

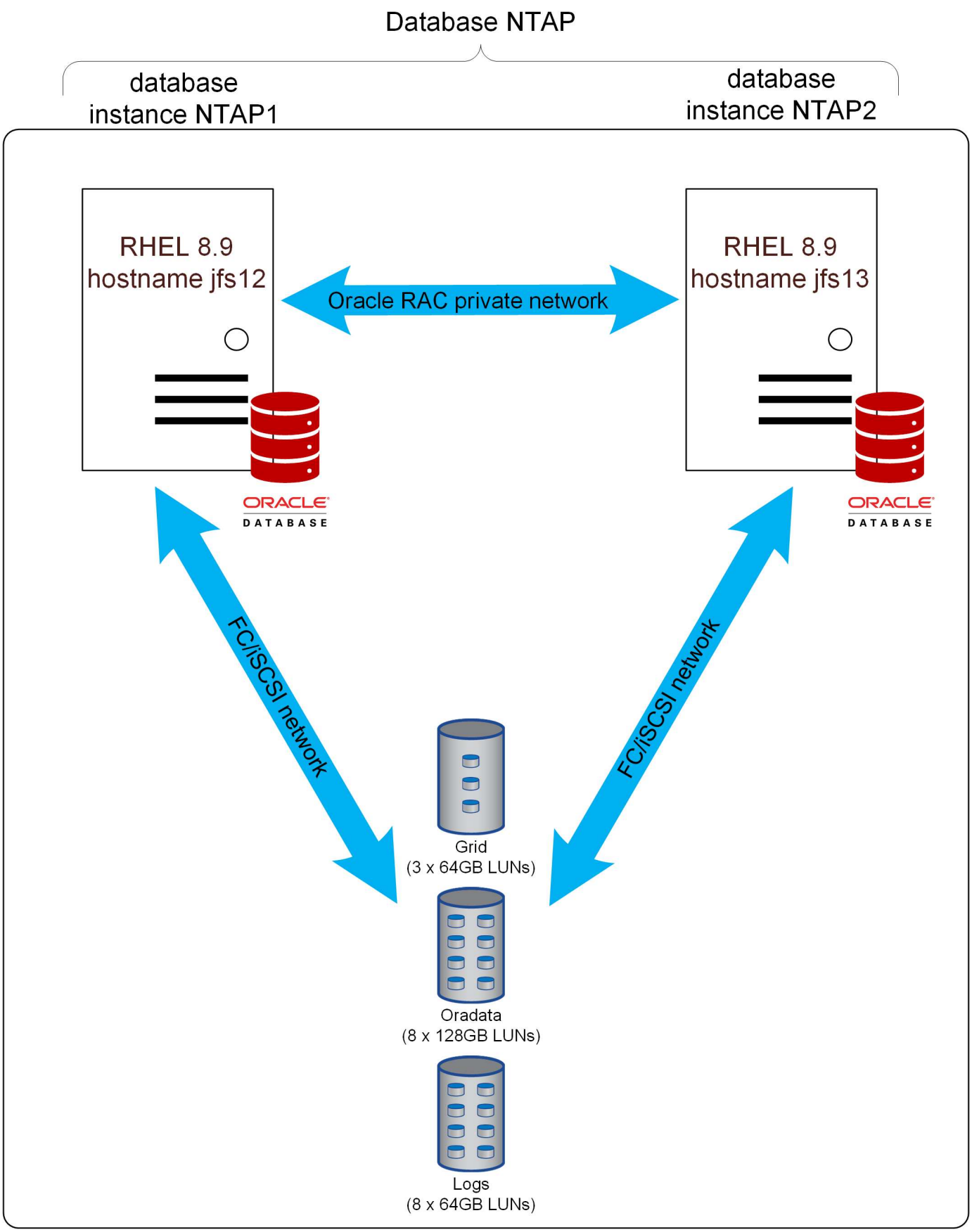
Molti clienti ottimizzano il proprio RTO estendendo un cluster Oracle RAC tra i vari siti, ottenendo una configurazione completamente Active-Active. La progettazione complessiva diventa più complicata perché deve includere la gestione del quorum di Oracle RAC.

Il cluster RAC esteso tradizionale si basava sul mirroring ASM per garantire la protezione dei dati. Questo approccio funziona, ma richiede anche molti passaggi di configurazione manuale e impone un overhead all'infrastruttura di rete. Al contrario, consentendo alla sincronizzazione attiva di SnapMirror di assumersi la

responsabilità della replica dei dati, la soluzione risulta notevolmente semplificata. Operazioni quali sincronizzazione, risincronizzazione dopo interruzioni, failover e gestione del quorum sono più semplici, inoltre la SAN non deve essere distribuita tra i siti, semplificando la progettazione e la gestione della SAN.

Replica

La chiave per comprendere le funzionalità RAC nella sincronizzazione attiva di SnapMirror è la visualizzazione dello storage come singolo set di LUN che si trovano sullo storage con mirroring. Ad esempio:



Nessuna copia primaria o copia speculare. A livello logico, esiste una sola copia di ogni LUN e tale LUN è disponibile nei percorsi SAN che si trovano su due sistemi di storage differenti. Dal punto di vista dell'host, non ci sono failover dello storage e modifiche del percorso. Diversi eventi di errore potrebbero causare la perdita di

determinati percorsi verso la LUN, mentre gli altri percorsi rimangono online. La sincronizzazione attiva di SnapMirror garantisce la disponibilità degli stessi dati su tutti i percorsi operativi.

Configurazione dello storage

In questa configurazione di esempio, i dischi ASM sono configurati come in qualsiasi configurazione RAC a sito singolo sullo storage Enterprise. Poiché il sistema di storage garantisce la protezione dei dati, viene utilizzata la ridondanza esterna ASM.

Accesso uniforme e non privo di informazioni

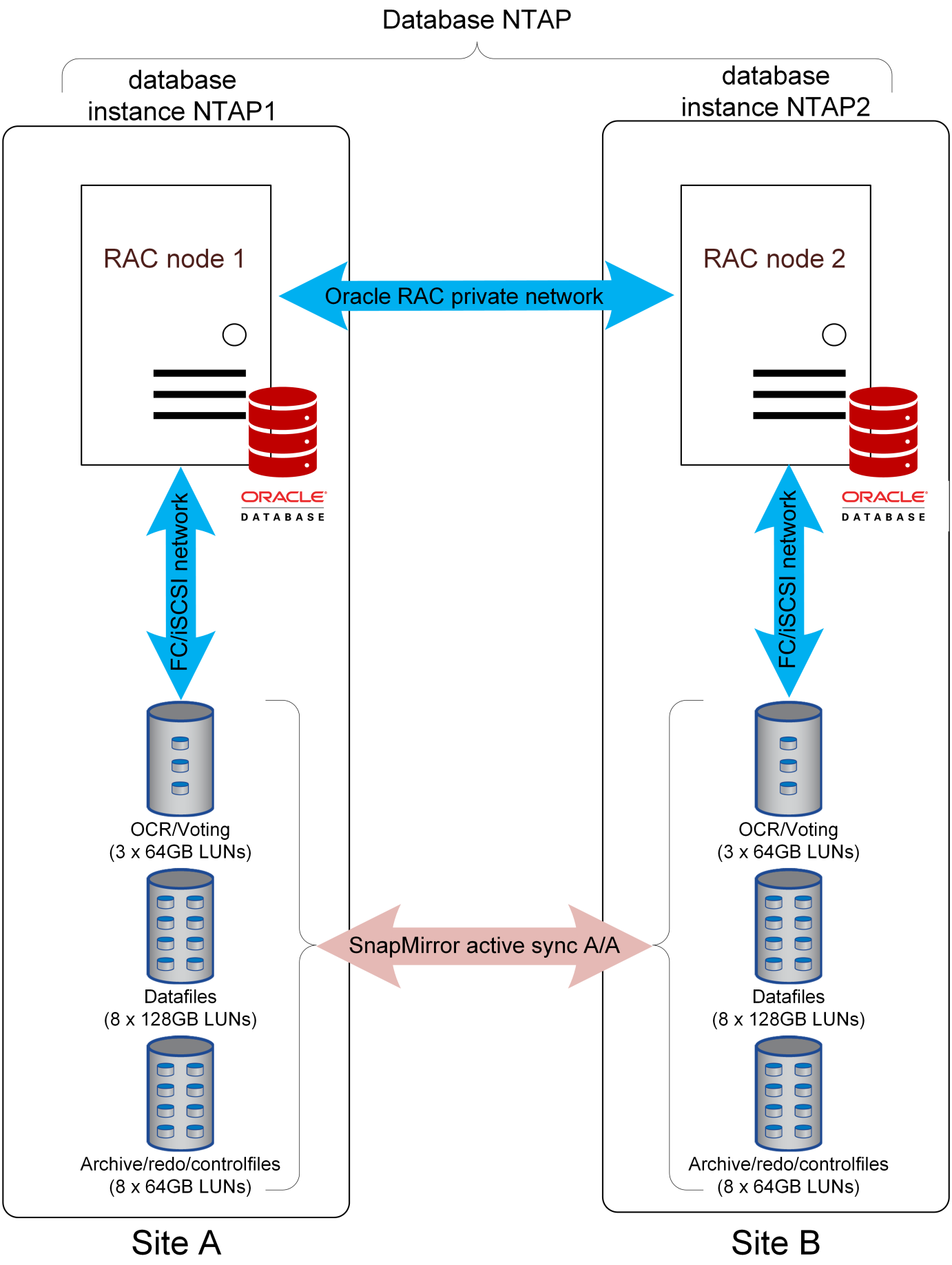
La considerazione più importante con Oracle RAC sulla sincronizzazione attiva di SnapMirror è se utilizzare un accesso uniforme o non uniforme.

Un accesso uniforme significa che ogni host può vedere i percorsi su entrambi i cluster. Accesso non uniforme significa che gli host possono vedere solo i percorsi verso il cluster locale.

Nessuna delle due opzioni è specificamente consigliata o scoraggiata. Alcuni clienti dispongono prontamente di fibra ottica spenta per la connessione ai siti, altri non dispongono di tale connettività oppure la loro infrastruttura SAN non supporta un ISL a lunga distanza.

Accesso non uniforme

L'accesso non uniforme è più semplice da configurare dal punto di vista della SAN.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
ORACLE DATABASE
ORACLE DATABASE
FC/iSCSI network
FC/iSCSI network
OCR/Voting (3 x 64GB LUNs)
OCR/Voting (3 x 64GB LUNs)
SnapMirror active sync A/A
Datafiles (8 x 128GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Site A
Site B

L’aspetto negativo principale "accesso non uniforme"dell’approccio è rappresentato dalla perdita della connettività ONTAP sito-sito o dalla perdita di un sistema di storage che causa la perdita delle istanze di database in un singolo sito. Questo ovviamente non è desiderabile, ma potrebbe essere un rischio accettabile in cambio di una configurazione SAN più semplice.

Accesso uniforme

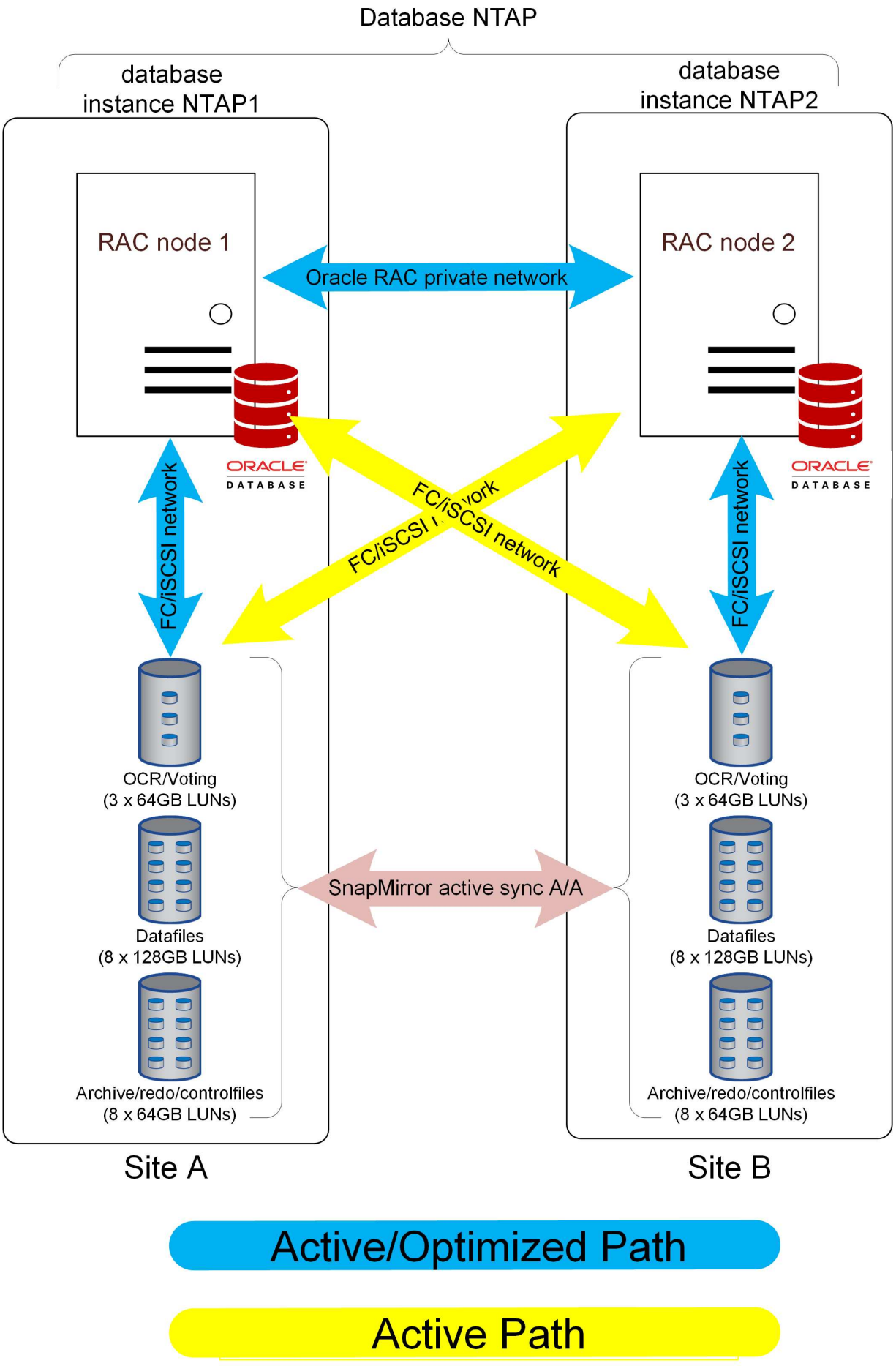
Un accesso uniforme richiede l’estensione della SAN tra i siti. Il vantaggio principale consiste nel fatto che la perdita di un sistema di storage non provocherà la perdita di un’istanza del database. Al contrario, si otterrebbe una modifica multipathing in cui i percorsi sono attualmente in uso.

Esistono diversi modi per configurare l’accesso non uniforme.

> Nei diagrammi di seguito sono presenti anche percorsi attivi ma non ottimizzati che sarebbero utilizzati durante semplici guasti del controller, ma tali percorsi non sono mostrati nell’interesse di semplificare i diagrammi.

AFF con impostazioni di prossimità

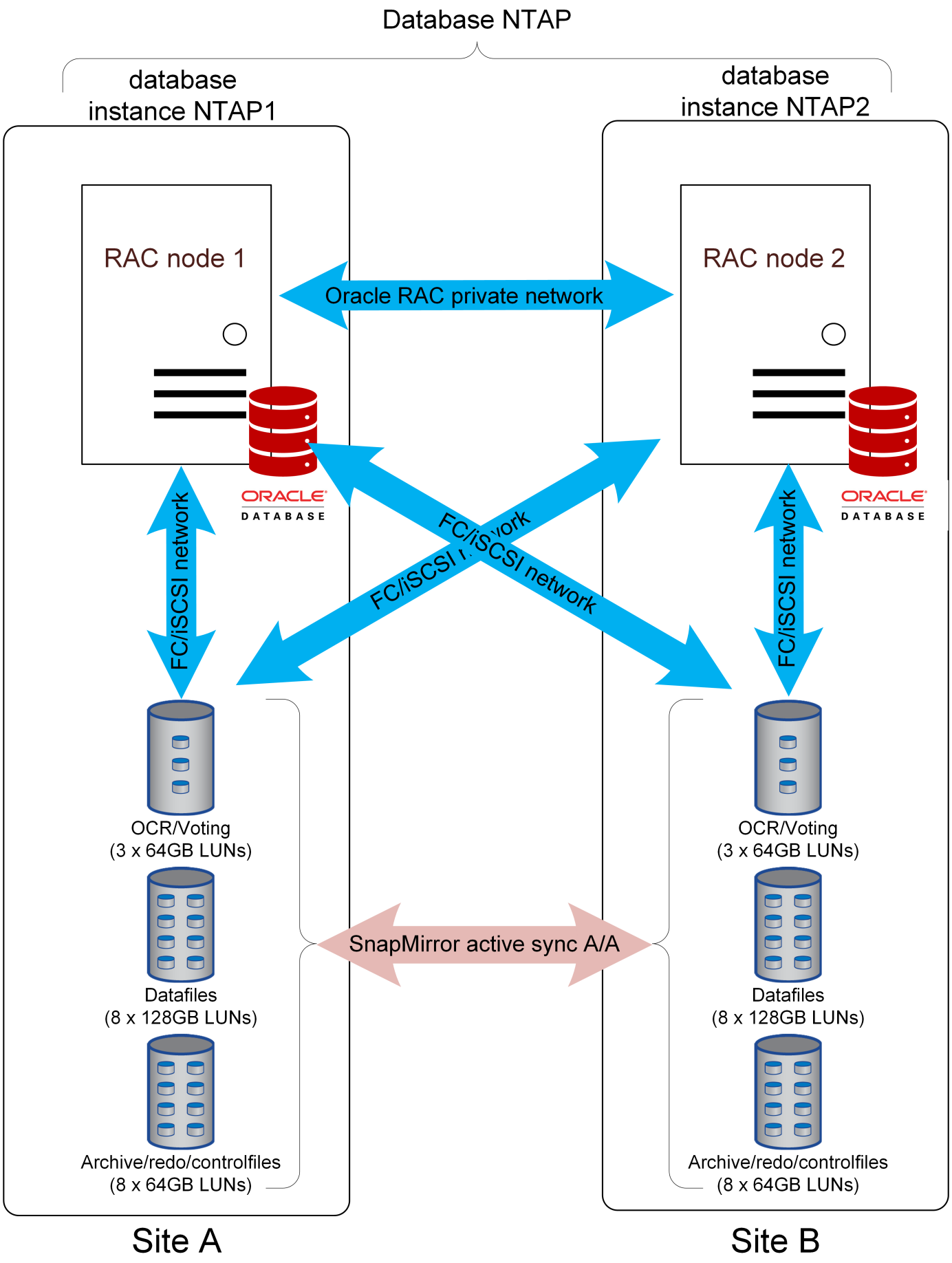
In presenza di una latenza significativa tra i siti, è possibile configurare i sistemi AFF con le impostazioni di prossimità dell’host. In questo modo, ciascun sistema storage può conoscere gli host locali e remoti e assegnare in maniera appropriata le priorità del percorso.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
ORACLE DATABASE
ORACLE DATABASE
OCR/Voting (3 x 64GB LUNs)
OCR/Voting (3 x 64GB LUNs)
SnapMirror active sync A/A
Datafiles (8 x 128GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Site A
Site B
Active/Optimized Path
Active Path

Durante il normale funzionamento, ogni istanza Oracle utilizzerebbe preferenzialmente i percorsi locali attivi/ottimizzati. Il risultato è che tutte le letture saranno gestite dalla copia locale dei blocchi. In questo modo, si ottiene la minore latenza possibile. L'io in scrittura viene analogamente inviato ai percorsi al controller locale. L'io deve ancora essere replicato prima di essere riconosciuto e quindi sarebbe comunque soggetto alla latenza aggiuntiva che attraversa la rete site-to-site, ma ciò non può essere evitato in una soluzione di replica sincrona.

ASA / AFF senza impostazioni di prossimità

In assenza di una latenza significativa tra i siti, è possibile configurare i sistemi AFF senza le impostazioni di prossimità dell'host oppure utilizzare ASA.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
ORACLE DATABASE
ORACLE DATABASE
OCR/Voting (3 x 64GB LUNs)
OCR/Voting (3 x 64GB LUNs)
SnapMirror active sync A/A
Datafiles (8 x 128GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Site A
Site B

Ciascun host sarà in grado di utilizzare tutti i percorsi operativi su entrambi i sistemi storage. Questo consente di migliorare potenzialmente le prestazioni consentendo a ciascun host di sfruttare il potenziale in termini di prestazioni di due cluster, non di uno solo.

Con ASA, non solo tutti i percorsi verso entrambi i cluster sono considerati attivi e ottimizzati, ma anche i percorsi sui partner controller sarebbero attivi. Ne risulterebbero percorsi SAN all-Active sull'intero cluster, in qualsiasi momento.

> I sistemi ASA possono anche essere utilizzati in una configurazione di accesso non uniforme. Poiché non esistono percorsi tra siti, non vi sarebbe alcun impatto sulle performance risultanti dall'io che attraversa l'ISL.

RAC tiebreaker

Sebbene il RAC esteso che utilizza la sincronizzazione attiva di SnapMirror sia un'architettura simmetrica rispetto all'io, esiste un'eccezione connessa alla gestione split-brain.

Cosa succede se il collegamento di replica viene perso e nessuno dei due siti ha un quorum? Che cosa dovrebbe accadere? Questa domanda è valida sia per Oracle RAC che per il comportamento di ONTAP. Se non è possibile replicare le modifiche tra i siti e si desidera riprendere le operazioni, uno dei siti dovrà sopravvivere e l'altro sito dovrà diventare non disponibile.

La "Mediatore ONTAP" soddisfa questo requisito al livello ONTAP. Esistono diverse opzioni per RAC Tiebreaking.

Gli Oracle Tiebreaker

Il metodo migliore per gestire i rischi di Oracle RAC split-brain consiste nell'utilizzare un numero dispari di nodi RAC, preferibilmente utilizzando un Tiebreaker a 3rd siti. Se un sito 3rd non è disponibile, l'istanza di Tiebreaker potrebbe essere posizionata in un sito dei due siti, designando effettivamente un sito di sopravvivenza preferito.

Oracle e css_Critical

Con un numero pari di nodi, il comportamento predefinito di Oracle RAC è che uno dei nodi nel cluster sarà considerato più importante degli altri nodi. Il sito con tale nodo con priorità più alta sopravviverà all'isolamento del sito, mentre i nodi sull'altro sito verranno eliminati. La priorità si basa su più fattori, ma è anche possibile controllare questo comportamento utilizzando l' `css_critical`impostazione.

Nell'"esempio"architettura, i nomi host per i nodi RAC sono jfs12 e jfs13. Di seguito sono riportate le impostazioni correnti di `css_critical`:

```
[root@jfs12 ~]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.

[root@jfs13 trace]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.
```

Se si desidera che il sito con jfs12 sia il sito preferito, impostare questo valore su sì su un nodo del sito A e riavviare i servizi.

```
[root@jfs12 ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.

[root@jfs12 ~]# /grid/bin/crsctl stop crs
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'jfs12'
CRS-2673: Attempting to stop 'ora.crsd' on 'jfs12'
CRS-2790: Starting shutdown of Cluster Ready Services-managed resources on
server 'jfs12'
CRS-2673: Attempting to stop 'ora.ntap.ntappdb1.pdb' on 'jfs12'
…
CRS-2673: Attempting to stop 'ora.gipcd' on 'jfs12'
CRS-2677: Stop of 'ora.gipcd' on 'jfs12' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'jfs12' has completed
CRS-4133: Oracle High Availability Services has been stopped.

[root@jfs12 ~]# /grid/bin/crsctl start crs
CRS-4123: Oracle High Availability Services has been started.
```

Scenari di errore

Panoramica

La pianificazione di un'architettura completa dell'applicazione SnapMirror Active Sync richiede la comprensione del modo in cui SM-AS risponderà in vari scenari di failover pianificati e non pianificati.

Per gli esempi seguenti, si supponga che il sito A sia configurato come sito preferito.

Interruzione della connettività di replica

Se la replica SM-AS viene interrotta, non è possibile completare la scrittura io perché sarebbe impossibile per un cluster replicare le modifiche al sito opposto.

Sito A (sito preferito)

Il risultato dell'errore del collegamento di replica sul sito preferito sarà una pausa di circa 15 secondi nell'elaborazione io in scrittura, poiché ONTAP ritenta le operazioni di scrittura replicate prima di determinare che il collegamento di replica è veramente irraggiungibile. Trascorsi i 15 secondi, il sistema del sito A riprende l'elaborazione io in lettura e scrittura. I percorsi SAN non vengono modificati e i LUN rimangono online.

Sito B

Poiché il sito B non è il sito preferito di sincronizzazione attiva SnapMirror, i relativi percorsi LUN non saranno più disponibili dopo circa 15 secondi.

Errore del sistema di storage

Il risultato di un errore del sistema di storage è quasi identico al risultato della perdita del collegamento di replica. Il sito sopravvissuto dovrebbe subire una pausa io di circa 15 secondi. Trascorso questo periodo di 15 secondi, io riprenderà sul sito come di consueto.

Perdita del mediatore

Il servizio di mediazione non controlla direttamente le operazioni di storage. Funziona come un percorso di controllo alternativo tra cluster. Esiste principalmente per automatizzare il failover senza il rischio di uno scenario split-brain. Durante l'utilizzo normale, ogni cluster replica le modifiche al partner e pertanto ogni cluster può verificare che il cluster partner sia online e fornisca i dati. Se il collegamento di replica non è riuscito, la replica viene interrotta.

Il motivo per cui è necessario un mediatore per il failover automatizzato sicuro è perché altrimenti sarebbe impossibile per un cluster di storage determinare se la perdita di comunicazione bidirezionale fosse il risultato di un'interruzione della rete o di un errore effettivo dello storage.

Il mediatore fornisce un percorso alternativo per ciascun cluster per verificare lo stato di salute del partner. Gli scenari sono i seguenti:

- Se un cluster può contattare direttamente il partner, i servizi di replica sono operativi. Non è richiesta alcuna azione.
- Se un sito preferito non può contattare direttamente il proprio partner o tramite il mediatore, presuppone che il partner sia effettivamente non disponibile oppure è stato isolato e ha portato i percorsi LUN offline. Il sito preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'io in lettura e in scrittura.
- Se un sito non preferito non può contattare direttamente il proprio partner, ma può contattarlo tramite il mediatore, prenderà i suoi percorsi offline e attenderà il ritorno della connessione di replica.
- Se un sito non preferito non può contattare direttamente il proprio partner o tramite un mediatore operativo, supporterà che il partner sia effettivamente non disponibile, oppure che sia stato isolato e che abbia portato i percorsi LUN offline. Il sito non preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'i/o in lettura e scrittura. Assumerà il ruolo dell'origine della replica e diventerà il nuovo sito preferito.

Se il mediatore non è completamente disponibile:

- In caso di guasto dei servizi di replica per qualsiasi motivo, incluso un guasto del sito o del sistema storage non preferito, il sito preferito rilascerà lo stato RPO=0 e riprenderà l'elaborazione i/o in lettura e scrittura. Il sito non preferito prenderà i suoi percorsi offline.
- Il guasto del sito preferito causerà un'interruzione poiché il sito non preferito non sarà in grado di verificare che il sito opposto sia effettivamente offline e quindi non sarebbe sicuro per il sito non preferito riprendere i servizi.

Ripristino dei servizi in corso

Dopo aver risolto un errore, come il ripristino della connettività da sito a sito o l'accensione di un sistema guasto, gli endpoint di sincronizzazione attivi di SnapMirror rilevano automaticamente la presenza di una relazione di replica difettosa e la riportano allo stato RPO=0. Una volta ristabilita la replica sincrona, i percorsi non riusciti torneranno in linea.

In molti casi, le applicazioni in cluster rilevano automaticamente la restituzione dei percorsi non riusciti e tali applicazioni tornano online. In altri casi, potrebbe essere necessaria una scansione SAN a livello di host

oppure potrebbe essere necessario riportare le applicazioni online manualmente. Dipende dall'applicazione e dal modo in cui è configurata, e in generale tali attività possono essere facilmente automatizzate. ONTAP si sta auto-riparando e non deve richiedere alcun intervento da parte dell'utente per riprendere le operazioni di storage RPO = 0 KB.

Failover manuale

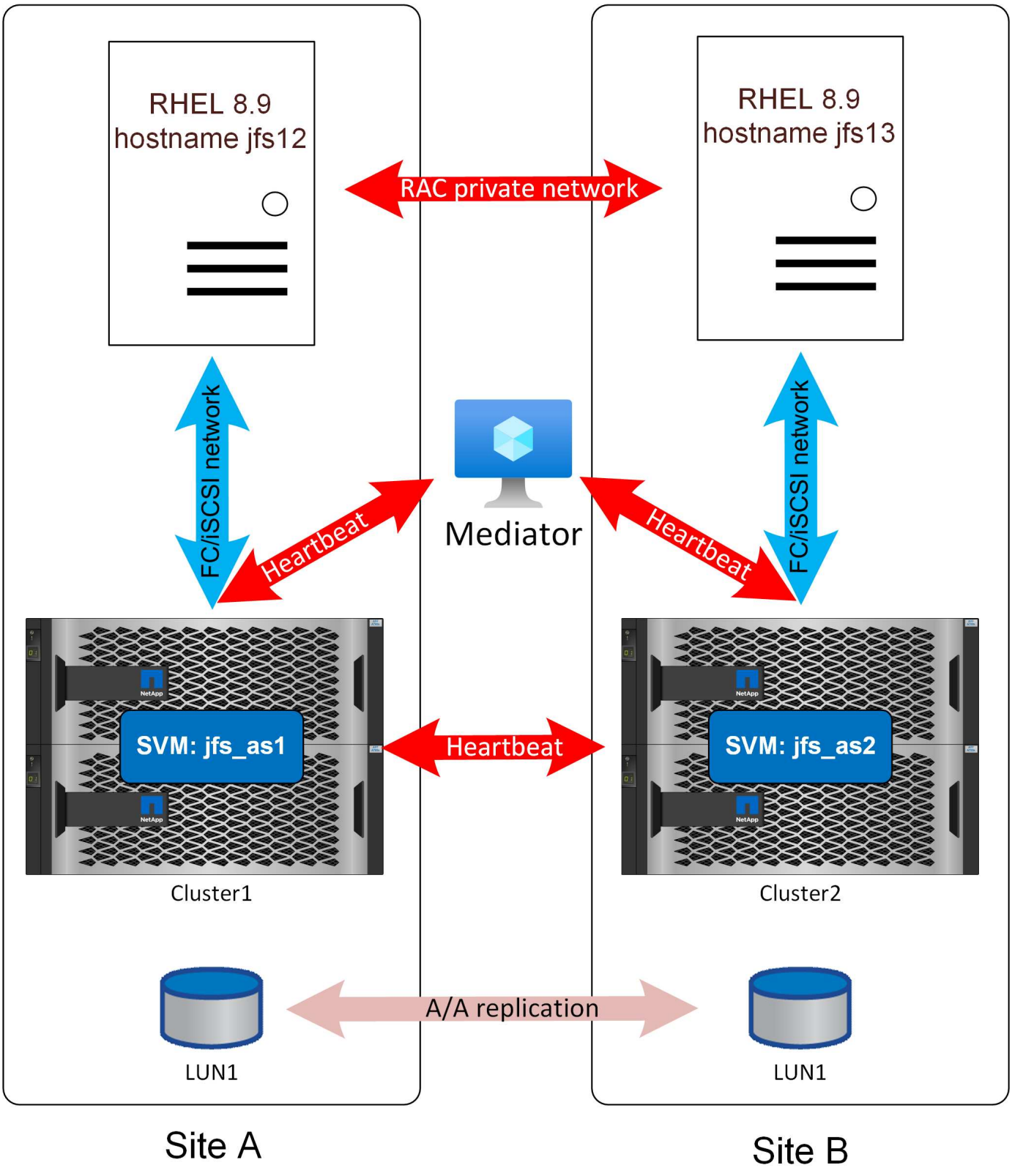
La modifica del sito preferito richiede un'operazione semplice. I/o si fermeranno per un secondo o due come autorità sugli switch del comportamento di replica tra i cluster, ma in caso contrario i/o non vengono influenzati.

Architettura di esempio

Gli esempi dettagliati di guasto illustrati in questa sezione si basano sull'architettura illustrata di seguito.

> Questa è solo una delle numerose opzioni per i database Oracle su SnapMirror Active Sync. Questo disegno è stato scelto perché illustra alcuni degli scenari più complicati.

In questa configurazione, si supponga che il sito A sia impostato su "sito preferito".



Errore di interconnessione RAC

La perdita del collegamento di replica di Oracle RAC produrrà un risultato simile alla perdita della connettività SnapMirror, tuttavia i timeout saranno più brevi per impostazione predefinita. Con le impostazioni predefinite, un nodo Oracle RAC attenderà 200 secondi

dopo la perdita della connettività di storage prima dell'eliminazione, ma attenderà solo 30 secondi dopo la perdita dell'heartbeat della rete RAC.

I messaggi CRS sono simili a quelli illustrati di seguito. È possibile visualizzare il timeout di 30 secondi. Poiché css_Critical è stato impostato su jfs12, situato sul sito A, questo sarà il sito per sopravvivere e jfs13 sul sito B sarà espulso.

```
2024-09-12 10:56:44.047 [ONMD(3528)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 6.980 seconds
2024-09-12 10:56:48.048 [ONMD(3528)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.980 seconds
2024-09-12 10:56:51.031 [ONMD(3528)]CRS-1607: Node jfs13 is being evicted
in cluster incarnation 621599354; details at (:CSSNM00007:) in
/gridbase/diag/crs/jfs12/crs/trace/onmd.trc.
2024-09-12 10:56:52.390 [CRSD(6668)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:33194;', interface list of remote node 'jfs13' is
'192.168.30.2:33621;'.
2024-09-12 10:56:55.683 [ONMD(3528)]CRS-1601: CSSD Reconfiguration
complete. Active nodes are jfs12 .
2024-09-12 10:56:55.722 [CRSD(6668)]CRS-5504: Node down event reported for
node 'jfs13'.
2024-09-12 10:56:57.222 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'Generic'.
2024-09-12 10:56:57.224 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'ora.NTAP'.
```

Errore di comunicazione SnapMirror

Se il collegamento di replica sincrona attiva di SnapMirror, l'io in scrittura non può essere completato perché sarebbe impossibile per un cluster replicare le modifiche al sito opposto.

Sito A

Il risultato sul sito A di un errore del collegamento di replica sarà una pausa di circa 15 secondi nell'elaborazione io in scrittura, poiché ONTAP tenta di replicare le scritture prima di determinare che il collegamento di replica è veramente inutilizzabile. Trascorsi i 15 secondi, il cluster ONTAP sul sito A riprende l'elaborazione i/o in lettura e scrittura. I percorsi SAN non vengono modificati e i LUN rimangono online.

Sito B

Poiché il sito B non è il sito preferito di sincronizzazione attiva SnapMirror, i relativi percorsi LUN non saranno più disponibili dopo circa 15 secondi.

Il collegamento di replica è stato tagliato alla data e ora 15:19:44. Il primo avviso da Oracle RAC arriva 100 secondi dopo l'avvicinarsi del timeout di 200 secondi (controllato dal disktimeout del parametro di Oracle RAC).

```
2024-09-10 15:21:24.702 [ONMD(2792)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99340 milliseconds.
2024-09-10 15:22:14.706 [ONMD(2792)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49330 milliseconds.
2024-09-10 15:22:44.708 [ONMD(2792)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19330 milliseconds.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.716 [ONMD(2792)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.731 [OCSSD(2794)]CRS-1652: Starting clean up of CRSD
resources.
```

Una volta raggiunto il timeout del disco di voting di 200 secondi, il nodo Oracle RAC verrà rimosso dal cluster e riavviato.

Errore interconnettività di rete totale

Se il collegamento di replica tra i siti viene perso completamente, la sincronizzazione attiva di SnapMirror e la connettività Oracle RAC verranno interrotte.

Il rilevamento split-brain di Oracle RAC dipende dall'heartbeat dello storage di Oracle RAC. Se la perdita della connettività da sito a sito determina la perdita simultanea dei servizi di replica dello storage e heartbeat della rete RAC, i siti RAC non saranno in grado di comunicare tra siti tramite RAC Interconnect o i dischi di voto RAC. Il risultato in un insieme di nodi pari-numerati può essere l'esclusione di entrambi i siti nelle impostazioni di default. Il comportamento esatto dipende dalla sequenza degli eventi e dalla tempistica dei sondaggi della rete RAC e degli heartbeat del disco.

È possibile risolvere il rischio di un'interruzione dei servizi dei siti 2 in due modi. In primo luogo, è possibile utilizzare una "tiebreaker" configurazione.

Se non è disponibile un sito 3rd, questo rischio può essere risolto regolando il parametro misscount sul cluster RAC. Per impostazione predefinita, il timeout heartbeat della rete RAC è di 30 secondi. Normalmente viene utilizzato dal RAC per identificare i nodi RAC guasti e rimuoverli dal cluster. Ha anche una connessione al disco di voto heartbeat.

Se, ad esempio, il condotto che trasporta il traffico tra i siti per Oracle RAC e i servizi di replica dello storage viene tagliato da un retroescavatore, inizia il conto alla rovescia per gli errori di 30 secondi. Se il nodo del sito preferito RAC non riesce a ristabilire il contatto con il sito opposto entro 30 secondi e non può utilizzare i dischi di voto per confermare che il sito opposto non sia attivo entro la stessa finestra di 30 secondi, anche i nodi del sito preferito verranno eliminati. Il risultato è un'interruzione completa del database.

A seconda di quando si verifica il polling dell'errore scount, 30 secondi potrebbero non essere sufficienti per il timeout della sincronizzazione attiva SnapMirror e consentire allo storage sul sito preferito di riprendere i servizi prima della scadenza della finestra di 30 secondi. Questa finestra di 30 secondi può essere aumentata.

```
[root@jfs12 ~]# /grid/bin/crsctl set css misscount 100
CRS-4684: Successful set of parameter misscount to 100 for Cluster
Synchronization Services.
```

Questo valore consente al sistema di storage sul sito preferito di riprendere le operazioni prima della scadenza del timeout per il conteggio degli errori. Il risultato sarà quindi l'eliminazione solo dei nodi nel sito in cui sono stati rimossi i percorsi LUN. Esempio seguente:

```
2024-09-12 09:50:59.352 [ONMD(681360)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 49.570 seconds
2024-09-12 09:51:10.082 [CRSD(682669)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:46039;', interface list of remote node 'jfs13' is
'192.168.30.2:42037;'.
2024-09-12 09:51:24.356 [ONMD(681360)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 24.560 seconds
2024-09-12 09:51:39.359 [ONMD(681360)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 9.560 seconds
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8011: reboot advisory message
from host: jfs13, component: cssagent, with time stamp: L-2024-09-12-
09:51:47.451
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8013: reboot advisory message
text: oracssdagent is about to reboot this node due to unknown reason as
it did not receive local heartbeats for 10470 ms amount of time
2024-09-12 09:51:48.925 [ONMD(681360)]CRS-1632: Node jfs13 is being
removed from the cluster in cluster incarnation 621596607
```

Oracle Support sconsiglia vivamente di modificare i parametri misscount o disktimeout per risolvere i problemi di configurazione. La modifica di questi parametri può tuttavia essere garantita e inevitabile in molti casi, incluso l'avvio SAN, la virtualizzazione e le configurazioni di replica dello storage. Se, ad esempio, si sono verificati problemi di stabilità con una rete SAN o IP che hanno provocato estrazioni RAC, è necessario risolvere il problema sottostante e non caricare i valori di misscount o disktimeout. La modifica dei timeout per correggere gli errori di configurazione è mascherare un problema, non risolvere un problema. La modifica di

questi parametri per configurare correttamente un ambiente RAC in base agli aspetti di progettazione dell'infrastruttura sottostante è diversa ed è coerente con le istruzioni di supporto di Oracle. Con l'avvio SAN, è comune regolare misscount fino a 200 per far corrispondere disktimeout. Per ulteriori informazioni, vedere"questo link".

Guasto del sito

Il risultato di un errore del sistema di storage o della sede è quasi identico al risultato della perdita del collegamento di replica. Il sito sopravvissuto dovrebbe subire una pausa io di circa 15 secondi sulle scritture. Trascorso questo periodo di 15 secondi, io riprenderà sul sito come di consueto.

Se il problema riguarda solo il sistema storage, il nodo Oracle RAC del sito guasto perderà i servizi di storage e inserirà lo stesso conto alla rovescia di 200 secondi per il disktime, prima dell'eliminazione e del successivo riavvio.

```
2024-09-11 13:44:38.613 [ONMD(3629)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99750 milliseconds.
2024-09-11 13:44:51.202 [ORAAGENT(5437)]CRS-5011: Check of resource "NTAP"
failed: details at "(:CLSN00007:)" in
"/gridbase/diag/crs/jfs13/crs/trace/crsd_oraagent_oracle.trc"
2024-09-11 13:44:51.798 [ORAAGENT(75914)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 75914
2024-09-11 13:45:28.626 [ONMD(3629)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49730 milliseconds.
2024-09-11 13:45:33.339 [ORAAGENT(76328)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 76328
2024-09-11 13:45:58.629 [ONMD(3629)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19730 milliseconds.
2024-09-11 13:46:18.630 [ONMD(3629)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-11 13:46:18.631 [ONMD(3629)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.638 [ONMD(3629)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.651 [OCSSD(3631)]CRS-1652: Starting clean up of CRSD
resources.
```

Lo stato del percorso SAN sul nodo RAC che ha perso i servizi storage è simile al seguente:

```
oradata7 (3600a0980383041334a3f55676c697347) dm-20 NETAPP,LUN C-Mode
size=128G features='3 queue_if_no_path pg_init_retries 50' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
| `- 34:0:0:18 sdam 66:96  failed faulty running
`-+- policy='service-time 0' prio=0 status=enabled
  `- 33:0:0:18 sdaj 66:48  failed faulty running
```

L'host linux ha rilevato la perdita dei percorsi molto più velocemente di 200 secondi, ma dal punto di vista del database le connessioni client all'host sul sito guasto saranno ancora bloccate per 200 secondi sotto le impostazioni predefinite di Oracle RAC. Le operazioni complete del database riprenderanno solo dopo che l'eviction è stata completata.

Nel frattempo, il nodo RAC di Oracle sul sito opposto registrerà la perdita dell'altro nodo RAC. In caso contrario, continua a funzionare come di consueto.

```
2024-09-11 13:46:34.152 [ONMD(3547)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 14.020 seconds
2024-09-11 13:46:41.154 [ONMD(3547)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 7.010 seconds
2024-09-11 13:46:46.155 [ONMD(3547)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.010 seconds
2024-09-11 13:46:46.470 [OHASD(1705)]CRS-8011: reboot advisory message
from host: jfs13, component: cssmonit, with time stamp: L-2024-09-11-
13:46:46.404
2024-09-11 13:46:46.471 [OHASD(1705)]CRS-8013: reboot advisory message
text: At this point node has lost voting file majority access and
oracssdmonitor is rebooting the node due to unknown reason as it did not
receive local hearbeats for 28180 ms amount of time
2024-09-11 13:46:48.173 [ONMD(3547)]CRS-1632: Node jfs13 is being removed
from the cluster in cluster incarnation 621516934
```

Errore mediatore

Il servizio di mediazione non controlla direttamente le operazioni di storage. Funziona come un percorso di controllo alternativo tra cluster. Esiste principalmente per automatizzare il failover senza il rischio di uno scenario split-brain.

Durante l'utilizzo normale, ogni cluster replica le modifiche al partner e pertanto ogni cluster può verificare che il cluster partner sia online e fornisca i dati. Se il collegamento di replica non è riuscito, la replica viene interrotta.

Il motivo per cui è necessario un mediatore per operazioni automatizzate sicure è che altrimenti sarebbe

impossibile per i cluster di storage determinare se la perdita di comunicazione bidirezionale fosse il risultato di un'interruzione della rete o di un errore effettivo dello storage.

Il mediatore fornisce un percorso alternativo per ciascun cluster per verificare lo stato di salute del partner. Gli scenari sono i seguenti:

- Se un cluster può contattare direttamente il partner, i servizi di replica sono operativi. Non è richiesta alcuna azione.
- Se un sito preferito non può contattare direttamente il proprio partner o tramite il mediatore, presuppone che il partner sia effettivamente non disponibile oppure è stato isolato e ha portato i percorsi LUN offline. Il sito preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'io in lettura e in scrittura.
- Se un sito non preferito non può contattare direttamente il proprio partner, ma può contattarlo tramite il mediatore, prenderà i suoi percorsi offline e attenderà il ritorno della connessione di replica.
- Se un sito non preferito non può contattare direttamente il proprio partner o tramite un mediatore operativo, supporterà che il partner sia effettivamente non disponibile, oppure che sia stato isolato e che abbia portato i percorsi LUN offline. Il sito non preferito procede quindi al rilascio dello stato RPO=0 e continua l'elaborazione dell'i/o in lettura e scrittura. Assumerà il ruolo dell'origine della replica e diventerà il nuovo sito preferito.

Se il mediatore non è completamente disponibile:

- In caso di guasto dei servizi di replica per qualsiasi motivo, il sito preferito rilascerà lo stato RPO=0 e riprenderà l'elaborazione i/o in lettura e scrittura. Il sito non preferito prenderà i suoi percorsi offline.
- Il guasto del sito preferito causerà un'interruzione poiché il sito non preferito non sarà in grado di verificare che il sito opposto sia effettivamente offline e quindi non sarebbe sicuro per il sito non preferito riprendere i servizi.

Ripristino del servizio

SnapMirror è a riparazione automatica. La sincronizzazione attiva di SnapMirror rileva automaticamente la presenza di una relazione di replica difettosa e la riporta allo stato RPO=0. Una volta ristabilita la replica sincrona, i percorsi torneranno online.

In molti casi, le applicazioni in cluster rilevano automaticamente la restituzione dei percorsi non riusciti e tali applicazioni tornano online. In altri casi, potrebbe essere necessaria una scansione SAN a livello di host oppure potrebbe essere necessario riportare le applicazioni online manualmente.

Dipende dall'applicazione e dal modo in cui è configurata, e in generale tali attività possono essere facilmente automatizzate. La sincronizzazione attiva di SnapMirror stessa risolve automaticamente il problema e non richiede alcun intervento da parte dell'utente per ripristinare l'RPO = 0 operazioni di storage una volta ripristinata l'alimentazione e la connettività.

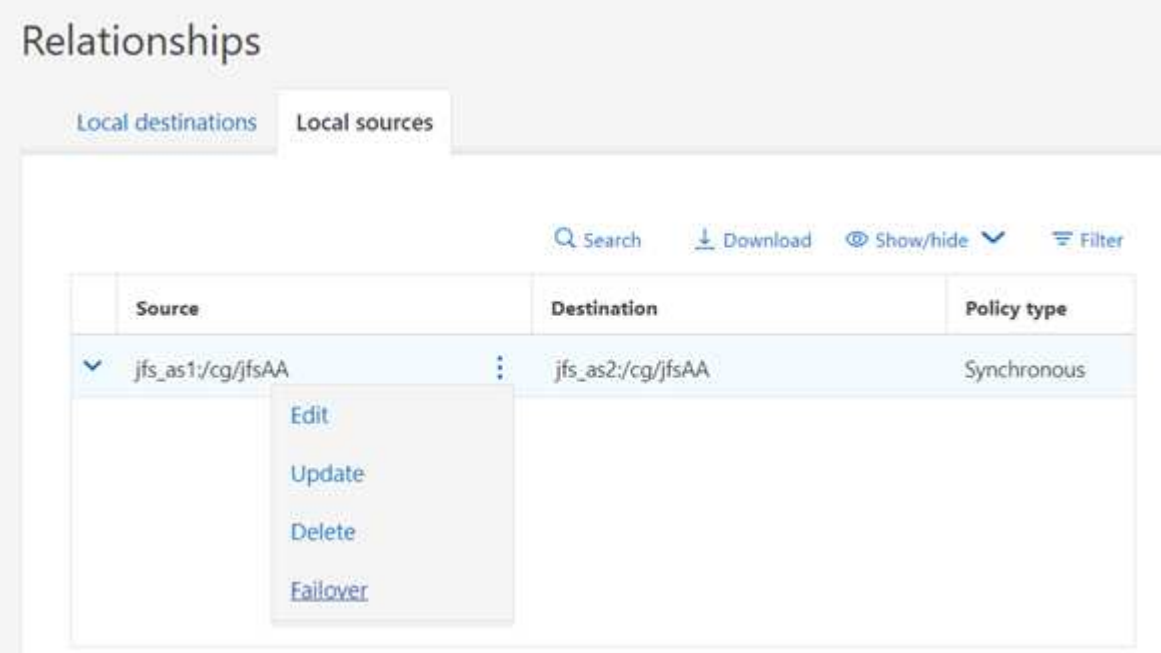
Failover manuale

Il termine "failover" non si riferisce alla direzione della replica con la sincronizzazione attiva di SnapMirror perché si tratta di una tecnologia di replica bidirezionale. "Failover" si riferisce invece al sistema di storage preferito in caso di guasto.

Ad esempio, è possibile eseguire un failover per modificare il sito preferito prima di chiudere un sito per la manutenzione o prima di eseguire un test di DR.

La modifica del sito preferito richiede un'operazione semplice. I/o si fermeranno per un secondo o due come autorità sugli switch del comportamento di replica tra i cluster, ma in caso contrario i/o non vengono influenzati.

Esempio di GUI:



Esempio di modifica tramite l'interfaccia CLI:

```
Cluster2::> snapmirror failover start -destination-path jfs_as2:/cg/jfsAA
[Job 9575] Job is queued: SnapMirror failover for destination
"jfs_as2:/cg/jfsAA                        ".

Cluster2::> snapmirror failover show

Source    Destination                                            Error
Path      Path        Type     Status    start-time end-time    Reason
--------- ----------  -------- --------- ---------- ---------- ----------
jfs_as1:/cg/jfsAA
          jfs_as2:/cg/jfsAA
                      planned  completed 9/11/2024  9/11/2024
                                         09:29:22   09:29:32

The new destination path can be verified as follows:


Cluster1::> snapmirror show -destination-path jfs_as1:/cg/jfsAA

                           Source Path: jfs_as2:/cg/jfsAA
                        Destination Path: jfs_as1:/cg/jfsAA
                       Relationship Type: XDP
                 Relationship Group Type: consistencygroup
                  SnapMirror Policy Type: automated-failover-duplex
                       SnapMirror Policy: AutomatedFailOverDuplex
                             Tries Limit: -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

Migrazione dei database Oracle

Panoramica

L'utilizzo delle funzionalità di una nuova piattaforma di storage impone un requisito inevitabile e prevede il posizionamento dei dati nel nuovo sistema di storage. ONTAP semplifica il processo di migrazione, inclusi aggiornamenti e migrazioni da ONTAP a ONTAP, importazioni di LUN esterne e procedure per l'utilizzo diretto del sistema operativo host o del software di database Oracle.

Questa documentazione sostituisce il report tecnico precedentemente pubblicato *TR-4534: Migrazione dei database Oracle in sistemi di storage NetApp*

Nel caso di un nuovo progetto di database, questo non rappresenta un problema, poiché gli ambienti di database e applicazioni sono stati costruiti in sede. La migrazione, tuttavia, pone sfide speciali in relazione all'interruzione del business, al tempo necessario per il completamento della migrazione, alle competenze

necessarie e alla minimizzazione del rischio.

Script

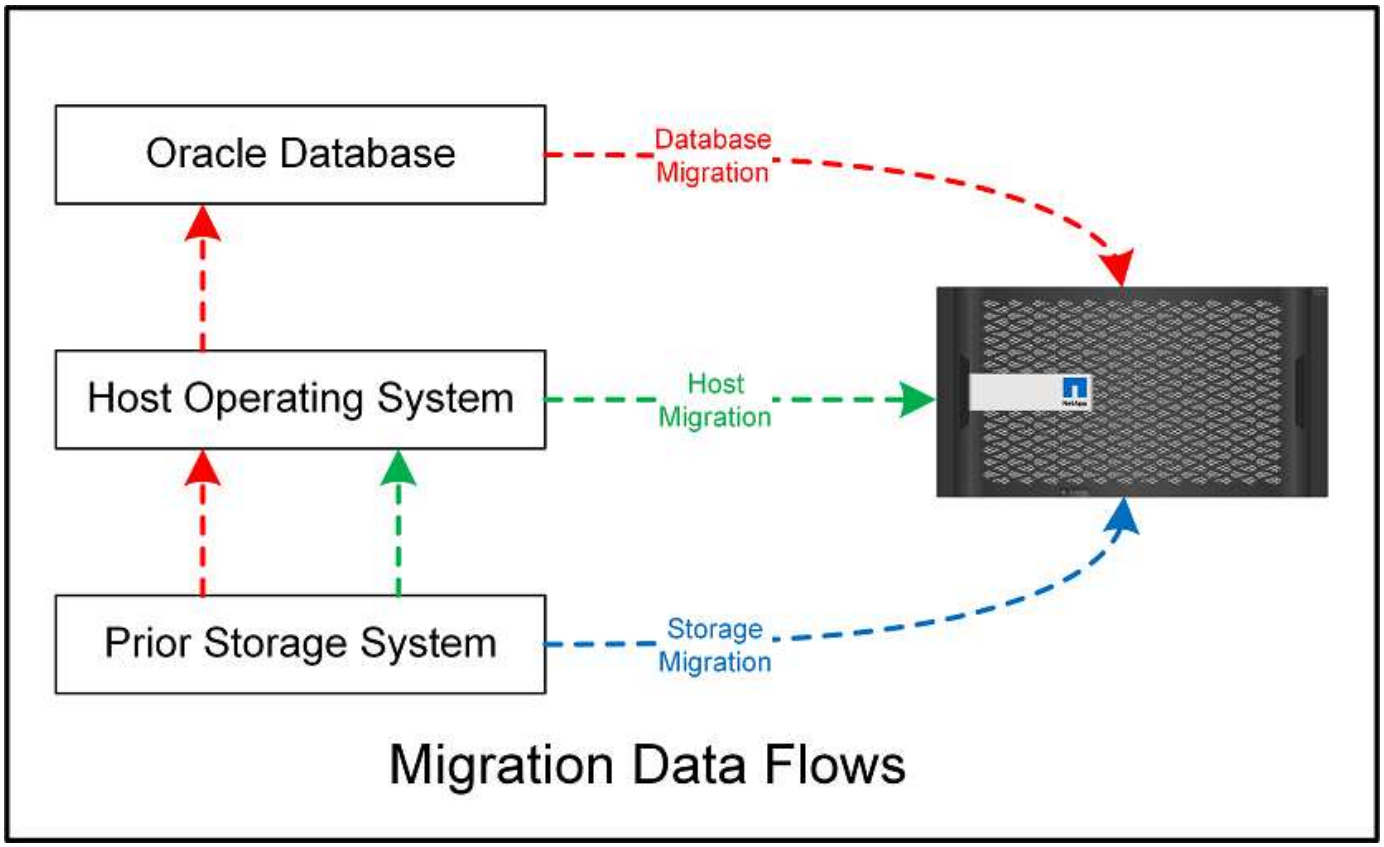
La presente documentazione contiene script di esempio. Questi script forniscono metodi di esempio per automatizzare vari aspetti della migrazione per ridurre la possibilità di errori da parte degli utenti. Gli script possono ridurre le richieste generali del personale IT responsabile della migrazione e accelerare il processo complessivo. Questi script sono ricavati da progetti di migrazione effettivi eseguiti dai servizi di assistenza professionale NetApp e dai partner NetApp. Nella presente documentazione sono riportati alcuni esempi del loro utilizzo.

Pianificazione della migrazione

La migrazione dei dati Oracle può avvenire a uno di tre livelli: Database, host o storage array.

Le differenze risiedono in quale componente della soluzione globale è responsabile dello spostamento dei dati: Il database, il sistema operativo host o il sistema di archiviazione.

La figura riportata di seguito mostra un esempio dei livelli di migrazione e del flusso di dati. In caso di migrazione a livello di database, i dati vengono spostati dal sistema di storage originale ai livelli di host e database nel nuovo ambiente. La migrazione a livello di host è simile, ma i dati non passano attraverso il livello di applicazione e vengono invece scritti nella nuova posizione utilizzando i processi degli host. Infine, con la migrazione a livello di storage, un array come un sistema NetApp FAS si occupa dello spostamento dei dati.



Una migrazione a livello di database si riferisce generalmente all'utilizzo di Oracle log shipping attraverso un database di standby per completare una migrazione a livello di Oracle. Le migrazioni a livello di host vengono eseguite utilizzando le funzionalità native della configurazione del sistema operativo host. Questa configurazione include le operazioni di copia dei file utilizzando comandi quali cp, tar e Oracle Recovery

Manager (RMAN) o un gestore del volume logico (LVM) per spostare i byte sottostanti di un file system. Oracle Automatic Storage Management (ASM) è classificato come capacità a livello di host perché viene eseguito al di sotto del livello dell'applicazione di database. ASM sostituisce il normale volume manager logico su un host. Infine, i dati possono essere migrati a livello di storage array, il che significa sotto il livello del sistema operativo.

Considerazioni sulla pianificazione

La scelta migliore per la migrazione dipende da una combinazione di fattori, inclusa la dimensione dell'ambiente da migrare, la necessità di evitare il downtime e lo sforzo complessivo necessario per eseguire la migrazione. Ovviamente, i database di grandi dimensioni richiedono più tempo e lavoro per la migrazione, ma la complessità di una migrazione di questo tipo è minima. I database di piccole dimensioni possono essere migrati rapidamente, ma se ne devono migrare migliaia, la portata dello sforzo può creare complicazioni. Infine, più grande è il database, più probabile è che l'IT sia business-critical, generando la necessità di ridurre al minimo i downtime mantenendo un percorso di back-out.

Alcune considerazioni per la pianificazione di una strategia di migrazione sono discusse qui.

Dimensioni dei dati

Le dimensioni dei database da migrare influiscono ovviamente sulla pianificazione della migrazione, sebbene le dimensioni non influiscano necessariamente sul tempo di cutover. Quando è necessario migrare una grande quantità di dati, l'aspetto più importante è la larghezza di banda. Le operazioni di copia vengono in genere eseguite con un efficiente i/o sequenziale Come stima conservativa, si presuppone un utilizzo del 50% della larghezza di banda della rete disponibile per le operazioni di copia. Ad esempio, una porta FC da 8GB GB può trasferire in teoria circa 800Mbps GB. Ipotizzando un utilizzo del 50%, è possibile copiare un database a una velocità di circa 400Mbps KB. Pertanto, un database 10TB può essere copiato in circa sette ore a questa velocità.

La migrazione su distanze più lunghe in genere richiede un approccio più creativo, ad esempio il processo di distribuzione dei log illustrato nella "Spostamento file dati online". Le reti IP a lunga distanza raramente dispongono di larghezza di banda in qualsiasi punto vicino alle velocità LAN o SAN. In un caso, NetApp ha assistito alla migrazione a lunga distanza di un database 220TB con tassi di generazione di log di archiviazione molto elevati. L'approccio scelto per il trasferimento dei dati era la spedizione giornaliera dei nastri, perché questo metodo offriva la massima larghezza di banda possibile.

Numero di database

In molti casi, il problema dello spostamento di una grande quantità di dati non è la dimensione dei dati, ma piuttosto la complessità della configurazione che supporta il database. Semplicemente sapere che 50TB database devono essere migrati non è sufficiente. Può essere un singolo database mission-critical 50TB, una raccolta di 4 database legacy 000 o un mix di dati di produzione e non. In alcuni casi, gran parte dei dati è costituita da cloni di un database di origine. Non è necessario migrare questi cloni perché possono essere facilmente ricreati, specialmente quando la nuova architettura è progettata per sfruttare i volumi FlexClone di NetApp.

Per la pianificazione della migrazione, è necessario comprendere il numero dei database interessati e la priorità da assegnare. Con l'aumento del numero di database, l'opzione di migrazione preferita tende a essere più bassa e più bassa nello stack. Ad esempio, la copia di un singolo database può essere eseguita facilmente con RMAN e con una breve interruzione del servizio. Si tratta di una replica a livello di host.

Se ci sono 50 database, potrebbe essere più facile evitare di impostare una nuova struttura del file system per ricevere una copia RMAN e spostare invece i dati sul posto. Questo processo può essere eseguito sfruttando la migrazione LVM basata su host per spostare i dati dalle vecchie LUN ai nuovi LUN. In tal modo, la responsabilità viene trasferita dal team di amministratori del database (DBA) al team del sistema operativo e,

di conseguenza, i dati vengono migrati in modo trasparente rispetto al database. La configurazione del file system rimane invariata.

Infine, se occorre migrare 500 database su 200 server, è possibile utilizzare opzioni basate sullo storage come la funzionalità FLI (ONTAP Foreign LUN Import) per eseguire una migrazione diretta delle LUN.

Requisiti di riarchitettura

In genere, per sfruttare le funzionalità del nuovo storage array è necessario modificare il layout dei file del database; tuttavia, non sempre questo avviene. Ad esempio, le funzionalità degli array all-flash EF-Series sono rivolte principalmente alle performance e all'affidabilità della SAN. Nella maggior parte dei casi, i database possono essere migrati su un array EF-Series senza particolari considerazioni sul layout dei dati. Gli unici requisiti sono IOPS elevati, bassa latenza e solida affidabilità. Sebbene esistano Best practice correlate a fattori quali la configurazione RAID o Dynamic Disk Pools, i progetti EF-Series raramente richiedono modifiche significative all'architettura dello storage generale per sfruttare tali funzionalità.

Al contrario, la migrazione a ONTAP richiede in genere una maggiore considerazione del layout del database per garantire che la configurazione finale fornisca il massimo valore. In sé, ONTAP offre molte funzionalità per un ambiente di database, anche senza interventi specifici sull'architettura. Soprattutto, offre la possibilità di migrare senza interruzioni al nuovo hardware quando l'hardware attuale termina la sua vita utile. In generale, la migrazione a ONTAP è l'ultima migrazione che è necessario eseguire. L'hardware successivo viene aggiornato e i dati vengono migrati senza interruzioni sui nuovi supporti.

Con una certa pianificazione, ancora più benefici sono disponibili. Le considerazioni più importanti riguardano l'uso delle istantanee. Le snapshot sono la base per l'esecuzione di backup, ripristini e operazioni di cloning quasi istantanei. Come esempio della potenza delle istantanee, il più grande utilizzo noto è con un singolo database di 996TB in esecuzione su circa 250 LUN su 6 controller. È possibile eseguire il backup di questo database in 2 minuti, ripristinarlo in 2 minuti e clonarlo in 15 minuti. Tra gli altri benefici, è inclusa la capacità di spostare i dati nel cluster in risposta alle variazioni del carico di lavoro e all'applicazione di controlli di qualità del servizio (QoS) per offrire performance buone e coerenti in un ambiente multi-database.

Tecnologie come controlli della QoS, trasferimento dei dati, snapshot e cloning funzionano praticamente in ogni configurazione. Tuttavia, un certo pensiero è generalmente richiesto per elevare i benefici. In alcuni casi, i layout dello storage del database possono richiedere modifiche di progettazione per massimizzare l'investimento nel nuovo storage array. Tali modifiche di progettazione possono influire sulla strategia di migrazione perché le migrazioni basate su host o su storage replicano il layout dei dati originale. Per completare la migrazione e offrire un layout dei dati ottimizzato per ONTAP potrebbero essere necessari ulteriori passaggi. Le procedure illustrate nella "Panoramica delle procedure di migrazione Oracle" in seguito, dimostrare alcuni metodi non solo per migrare un database, ma anche per eseguirne la migrazione nel layout finale ottimale con il minimo sforzo.

Tempo di cutover

Occorre determinare il disservizio massimo consentito del servizio durante il cutover. È un errore comune presumere che l'intero processo di migrazione causi interruzioni. È possibile eseguire numerose attività prima dell'inizio di qualsiasi interruzione del servizio e completare la migrazione senza interruzioni o black-out attraverso diverse opzioni. Anche quando è inevitabile un'interruzione, è comunque necessario definire il fuori servizio massimo consentito poiché la durata del tempo di cutover varia da procedura a procedura.

Ad esempio, la copia di un database 10TB richiede in genere circa sette ore. Se le esigenze aziendali rendono possibile un'interruzione di sette ore, la copia dei file è un'opzione semplice e sicura per la migrazione. Se cinque ore sono inaccettabili, un semplice log-processo di spedizione (vedere "Log shipping di Oracle") può essere impostato con il minimo sforzo per ridurre il tempo di cutover a circa 15 minuti. Durante questo periodo, un amministratore di database può completare il processo. Se 15 minuti sono inaccettabili, è possibile automatizzare il processo di cutover finale tramite script per ridurre il tempo di cutover a pochi minuti. È

sempre possibile accelerare una migrazione, anche se ciò comporta costi di tempo e lavoro. Gli obiettivi del tempo di cutover devono basarsi su ciò che è accettabile per l'azienda.

Percorso di ritorno

Nessuna migrazione è completamente priva di rischi. Anche se la tecnologia funziona perfettamente, c'è sempre la possibilità di errori da parte dell'utente. Il rischio associato a un percorso di migrazione scelto deve essere preso in considerazione insieme alle conseguenze di una migrazione non riuscita. Ad esempio, la capacità di migrazione trasparente dello storage online di Oracle ASM è una delle sue caratteristiche principali e questo metodo è uno dei più affidabili. Tuttavia, i dati vengono copiati irreversibilmente con questo metodo. Nel caso altamente improbabile in cui si verifichi un problema con ASM, non esiste un facile percorso di back-out. L'unica opzione è ripristinare l'ambiente originale o utilizzare ASM per riportare la migrazione ai LUN originali. Il rischio può essere minimizzato, ma non eliminato, eseguendo un backup di tipo snapshot sul sistema di storage originale, supponendo che il sistema sia in grado di eseguire tale operazione.

Prova

Alcune procedure di migrazione devono essere verificate completamente prima dell'esecuzione. La necessità di migrazione e verifica del processo di cutover è una richiesta comune con i database mission-critical per i quali la migrazione deve avere successo e il downtime deve essere ridotto al minimo. Inoltre, i test di accettazione da parte dell'utente sono spesso inclusi come parte del lavoro di post-migrazione e il sistema complessivo può essere riportato in produzione solo dopo il completamento di questi test.

In caso di necessità di prove, diverse funzionalità di ONTAP possono rendere il processo molto più semplice. In particolare, le istantanee possono ripristinare un ambiente di test e creare rapidamente più copie di un ambiente di database efficienti in termini di spazio.

Procedure

Panoramica

Sono disponibili molte procedure per il database di migrazione Oracle. La giusta dipende dalle vostre esigenze aziendali.

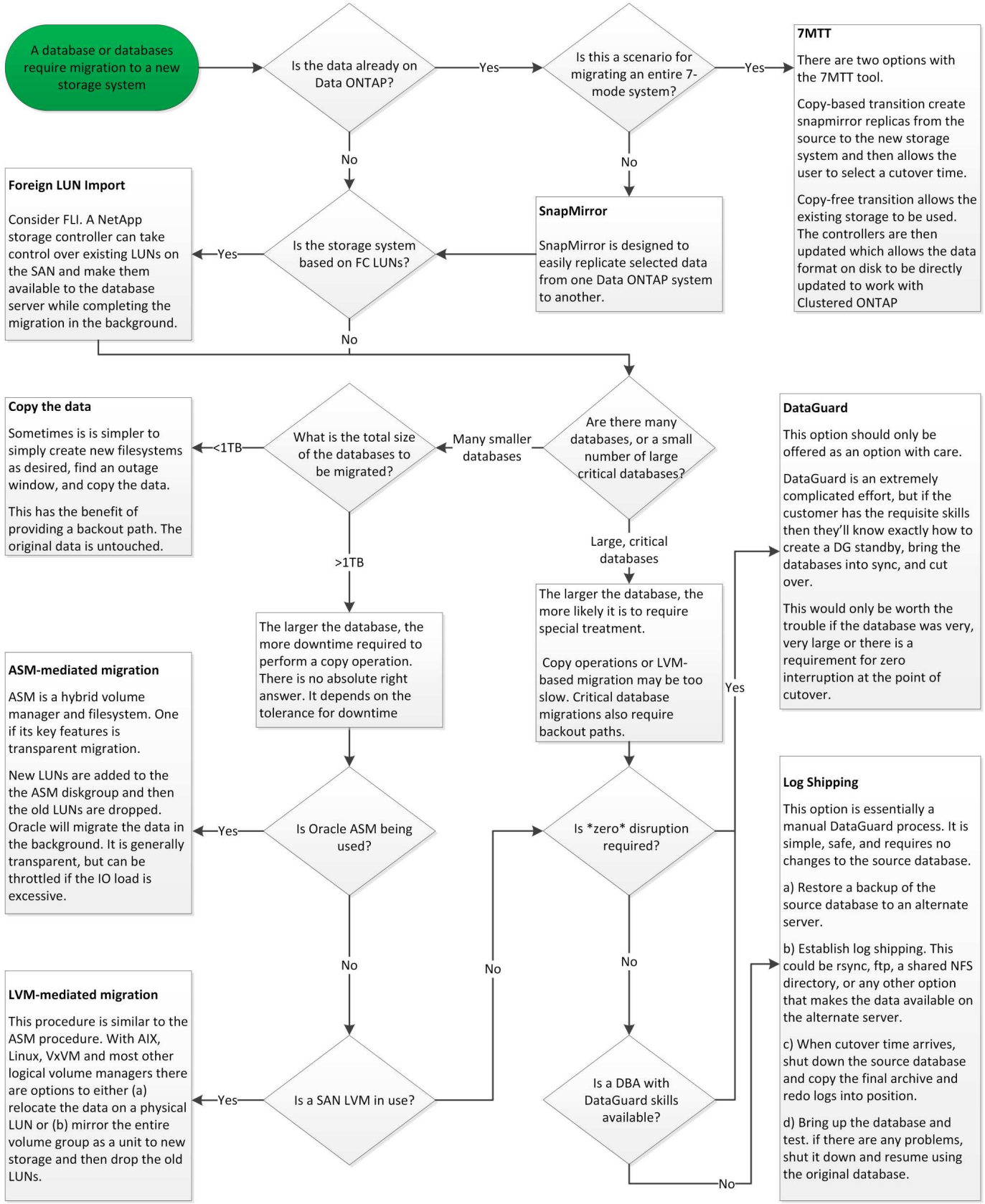
In molti casi, gli amministratori di sistema e i DBA dispongono dei propri metodi preferiti per trasferire i dati dei volumi fisici, eseguire il mirroring e il demirroring o utilizzare Oracle RMAN per copiare i dati.

Queste procedure vengono fornite principalmente come guida per il personale IT meno esperto di alcune delle opzioni disponibili. Inoltre, vengono illustrate le attività, i requisiti di tempo e le richieste di competenze per ogni approccio alla migrazione. Ciò consente ad altre parti, come NetApp e i servizi professionali dei partner o i responsabili dell'IT, di apprezzare più pienamente i requisiti di ogni procedura.

Non esiste un'unica Best practice per la creazione di una strategia di migrazione. La creazione di un piano richiede prima di tutto la comprensione delle opzioni di disponibilità e quindi la selezione del metodo più adatto alle esigenze dell'azienda. La figura seguente illustra le considerazioni di base e le conclusioni tipiche dei clienti, ma non è applicabile a tutte le situazioni.

Ad esempio, un passaggio solleva il problema della dimensione totale del database. Il passaggio successivo dipende dal fatto che il database sia maggiore o minore di 1TB. I passaggi consigliati sono solo questi: Consigli basati su pratiche tipiche del cliente. La maggior parte dei clienti non utilizzerebbe DataGuard per copiare un database di piccole dimensioni, ma alcuni potrebbero farlo. La maggior parte dei clienti non tenterebbe di copiare un database 50TB per il tempo necessario, ma alcuni potrebbero avere una finestra di manutenzione sufficientemente grande da consentire tale operazione.

Il diagramma di flusso riportato di seguito mostra i tipi di considerazioni sul percorso di migrazione più adatto. È possibile fare clic con il pulsante destro del mouse sull'immagine e aprirla in una nuova scheda per migliorare la leggibilità.



.

Spostamento file dati online

Oracle 12cR1 e versioni successive includono la possibilità di spostare un file dati mentre il database rimane online. Inoltre funziona tra diversi tipi di filesystem. Ad esempio, è possibile spostare un file dati da un filesystem xfs ad ASM. Questo metodo non viene generalmente utilizzato su larga scala a causa del numero di operazioni singole di spostamento del file di dati che sarebbero necessarie, ma è un'opzione che vale la pena considerare con database più piccoli con meno file di dati.

Inoltre, il semplice spostamento di un file dati è un'ottima opzione per la migrazione di parti di database esistenti. Ad esempio, è possibile ricollocare i file di dati meno attivi in uno storage più conveniente, ad esempio un volume FabricPool che consente di memorizzare blocchi inattivi in Object Store.

Migrazione a livello di database

La migrazione a livello di database significa consentire il trasferimento dei dati. In particolare, ciò significa spedizione dei log. Tecnologie come RMAN e ASM sono prodotti Oracle, ma, ai fini della migrazione, operano a livello di host in cui copiano i file e gestiscono i volumi.

Spedizione dei log

La base per la migrazione a livello di database è il log di archivio di Oracle, che contiene un registro delle modifiche apportate al database. Nella maggior parte dei casi, un registro di archiviazione fa parte di una strategia di backup e ripristino. Il processo di ripristino inizia con il ripristino di un database e quindi la riproduzione di uno o più log di archivio per portare il database allo stato desiderato. Questa stessa tecnologia di base può essere utilizzata per eseguire una migrazione con interruzioni delle operazioni minime o nulle. Cosa ancora più importante, questa tecnologia consente la migrazione senza intaccare il database originale, preservando un percorso di back-out.

Il processo di migrazione inizia con il ripristino di un backup del database su un server secondario. È possibile farlo in vari modi, ma la maggior parte dei clienti utilizza la normale applicazione di backup per ripristinare i file di dati. Una volta ripristinati i file di dati, gli utenti stabiliscono un metodo per la distribuzione dei log. L'obiettivo è creare un feed costante di log di archivio generati dal database primario e riprodurli sul database ripristinato per mantenerli entrambi vicini allo stesso stato. Quando arriva il tempo di cutover, il database di origine viene completamente arrestato e i log di archivio finali e, in alcuni casi, i log di redo vengono copiati e riprodotti. È fondamentale che i log di ripristino vengano presi in considerazione anche perché potrebbero contenere alcune delle transazioni finali impegnate.

Una volta trasferiti e riprodotti questi log, entrambi i database sono coerenti l'uno con l'altro. A questo punto, la maggior parte dei clienti esegue alcuni test di base. In caso di errori durante il processo di migrazione, la riproduzione del registro dovrebbe segnalare errori e errori. È comunque consigliabile eseguire alcuni test rapidi basati su query note o su attività guidate dalle applicazioni per verificare che la configurazione sia ottimale. È inoltre pratica comune creare una tabella di test finale prima di chiudere il database originale per verificare se è presente nel database migrato. Questa operazione garantisce che non siano stati commessi errori durante la sincronizzazione finale del registro.

Una semplice migrazione log-shipping può essere configurata fuori banda rispetto al database originale, il che lo rende particolarmente utile per i database mission-critical. Non sono richieste modifiche alla configurazione per il database di origine e il ripristino e la configurazione iniziale dell'ambiente di migrazione non hanno alcun effetto sulle operazioni di produzione. Una volta configurato, il log shipping pone alcune richieste di i/o sui server di produzione. Tuttavia, il log shipping è costituito da semplici letture sequenziali dei registri di archivio, che hanno scarse probabilità di influire sulle prestazioni del database di produzione.

La distribuzione dei log si è dimostrata particolarmente utile per progetti di migrazione a lunga distanza e ad alta velocità di cambiamento. In un'istanza, è stata eseguita la migrazione di un singolo database 220TB in una nuova posizione a circa 500 km di distanza. La velocità di modifica era estremamente elevata e le

restrizioni di sicurezza impedivano l'utilizzo di una connessione di rete. La spedizione dei log è stata eseguita utilizzando nastro e corriere. Una copia del database di origine è stata inizialmente ripristinata utilizzando le procedure descritte di seguito. Quindi, i registri sono stati spediti settimanalmente tramite corriere fino al momento del cutover, al momento della consegna del set finale di nastri e dell'applicazione dei registri al database di replica.

Oracle DataGuard

In alcuni casi, è garantito un ambiente DataGuard completo. Non è corretto utilizzare il termine DataGuard per fare riferimento a qualsiasi configurazione del database di standby o di distribuzione dei log. Oracle DataGuard è un framework completo per la gestione della replica dei database, ma non è una tecnologia di replica. Il vantaggio principale di un ambiente DataGuard completo in uno sforzo di migrazione è lo switchover trasparente da un database all'altro. DataGuard consente inoltre uno switchover trasparente nel database originale in caso di problemi, ad esempio problemi di prestazioni o connettività di rete nel nuovo ambiente. Un ambiente DataGuard completamente configurato richiede la configurazione non solo del livello del database ma anche delle applicazioni in modo che le applicazioni siano in grado di rilevare una modifica nella posizione del database primario. In generale, non è necessario utilizzare DataGuard per completare una migrazione, ma alcuni clienti hanno una vasta esperienza DataGuard in-house e già si affidano a essa per le attività di migrazione.

Riarchitettura

Come discusso in precedenza, per sfruttare le funzionalità avanzate degli storage array è talvolta necessario modificare il layout del database. Inoltre, una modifica nel protocollo di storage, come il passaggio da ASM a un file system NFS, altera necessariamente il layout del file system.

Uno dei principali vantaggi dei metodi di distribuzione dei log, incluso DataGuard, è che la destinazione di replica non deve corrispondere all'origine. Non vi sono problemi con l'utilizzo di un approccio di log-shipping per migrare da ASM a un normale file system o viceversa. Il layout preciso dei file di dati può essere modificato a destinazione per ottimizzare l'uso della tecnologia Pluggable Database (PDB) o per impostare i controlli QoS in modo selettivo su determinati file. In altre parole, un processo di migrazione basato sul log shipping consente di ottimizzare il layout dello storage del database in modo semplice e sicuro.

Risorse dei server

Un limite alla migrazione a livello di database è la necessità di un secondo server. Questo secondo server può essere utilizzato in due modi:

1. È possibile utilizzare il secondo server come nuova casa permanente per il database.
2. È possibile utilizzare il secondo server come server di staging temporaneo. Una volta completata e testata la migrazione dei dati nel nuovo storage array, i file system LUN o NFS vengono disconnessi dal server di staging e riconnessi al server originale.

La prima opzione è la più semplice, ma l'utilizzo potrebbe non essere possibile in ambienti molto grandi che richiedono server molto potenti. La seconda opzione richiede ulteriore lavoro per riportare i file system nella posizione originale. Si tratta di una semplice operazione in cui NFS viene utilizzato come protocollo storage, poiché i file system possono essere smontati dal server di staging e rimontati sul server originale.

I file system basati su blocchi richiedono lavoro extra per l'aggiornamento dello zoning FC o degli iSCSI initiator. Con la maggior parte dei gestori di volumi logici (incluso ASM), i LUN vengono automaticamente rilevati e portati online una volta resi disponibili sul server originale. Tuttavia, alcune implementazioni di file system e LVM potrebbero richiedere più lavoro per esportare e importare i dati. La procedura precisa può variare, ma in genere è facile stabilire una procedura semplice e ripetibile per completare la migrazione e ripristinare i dati sul server originale.

Sebbene sia possibile impostare la distribuzione dei log e replicare un database all'interno di un singolo ambiente server, la nuova istanza deve avere un SID di processo diverso per riprodurre i log. È possibile visualizzare temporaneamente il database con un diverso gruppo di ID di processo con un SID diverso e modificarlo in un secondo momento. Tuttavia, questo può portare a numerose e complicate attività di gestione ed espone l'ambiente di database al rischio di errori dell'utente.

Migrazione a livello di host

Migrare i dati a livello di host significa utilizzare il sistema operativo host e le utility associate per completare la migrazione. Questo processo include qualsiasi utility che copia i dati, inclusi Oracle RMAN e Oracle ASM.

Copia dei dati

Il valore di un'operazione di copia semplice non deve essere sottovalutato. Le moderne infrastrutture di rete sono in grado di spostare i dati a velocità misurate in gigabyte al secondo, mentre le operazioni di copia dei file si basano su un efficiente i/o di lettura e scrittura sequenziale L'interruzione è inevitabile con un'operazione di copia dell'host rispetto alla spedizione dei log, ma la migrazione non riguarda solo lo spostamento dei dati. In genere sono incluse le modifiche alla rete, il tempo di riavvio del database e i test post-migrazione.

Il tempo effettivo richiesto per copiare i dati potrebbe non essere significativo. Inoltre, l'operazione di copia preserva un percorso di back-out garantito perché i dati originali non vengono intatti. In caso di problemi durante il processo di migrazione, è possibile riattivare i file system originali con i dati originali.

Riformulazione

Replatforming si riferisce a una modifica del tipo di CPU. Quando un database viene migrato da una piattaforma Solaris, AIX o HP-UX tradizionale a x86 Linux, i dati devono essere riformattati a causa delle modifiche apportate all'architettura della CPU. Le CPU SPARC, IA64 e POWER sono note come grandi processori endian, mentre le architetture x86 e x86_64 sono note come Little endian. Di conseguenza, alcuni dati all'interno dei file di dati Oracle vengono ordinati in modo diverso a seconda del processore in uso.

Tradizionalmente, i clienti utilizzano DataPump per replicare i dati su più piattaforme. DataPump è un'utilità che crea un tipo speciale di esportazione dei dati logici che può essere importata più rapidamente nel database di destinazione. Poiché crea una copia logica dei dati, DataPump lascia alle spalle le dipendenze dell'endianness del processore. Anche se alcuni clienti usano DataPump per il replatform, con Oracle 11g è ora disponibile un'opzione più rapida: Tablespace trasportabili su più piattaforme. Questo avanzamento consente di convertire un tablespace in un diverso formato endian. Si tratta di una trasformazione fisica che offre prestazioni migliori rispetto a un'esportazione DataPump, che deve convertire i byte fisici in dati logici e quindi riconvertirli in byte fisici.

Una discussione completa su DataPump e tablespace trasportabili va oltre la documentazione relativa al NetApp dell'ambito, ma NetApp offre alcuni consigli basati sulla nostra esperienza nell'assistenza ai clienti durante la migrazione a un nuovo log di storage array con una nuova architettura della CPU:

- Se si utilizza DataPump, il tempo necessario per completare la migrazione deve essere misurato in un ambiente di test. A volte i clienti vengono sorpresi del tempo necessario per completare la migrazione. Questo downtime aggiuntivo e inatteso può causare interruzioni delle attività.
- Molti clienti credono erroneamente che gli spazi di tabella trasportabili su più piattaforme non richiedano la conversione dei dati. Quando si utilizza una CPU con un endian diverso, viene utilizzato un RMAN `convert` l'operazione deve essere eseguita sui file di dati in anticipo. Non si tratta di un'operazione istantanea. In alcuni casi, il processo di conversione può essere accelerato avendo più thread che operano su file di dati diversi, ma il processo di conversione non può essere evitato.

Migrazione guidata dal volume logico

Le LVM funzionano prendendo un gruppo di uno o più LUN e suddividendoli in piccole unità generalmente denominate estensioni. Il pool di estensioni viene quindi utilizzato come origine per creare volumi logici essenzialmente virtualizzati. Questo livello di virtualizzazione offre valore in vari modi:

- I volumi logici possono utilizzare estensioni tratte da più LUN. Quando un file system viene creato su un volume logico, può utilizzare le funzionalità con le performance complete di tutte le LUN. Inoltre, promuove il caricamento uniforme di tutte le LUN nel gruppo di volumi, offrendo performance più prevedibili.
- I volumi logici possono essere ridimensionati aggiungendo e, in alcuni casi, rimuovendo le estensioni. Il ridimensionamento di un file system su un volume logico avviene in genere senza interruzione delle attività.
- È possibile migrare i volumi logici senza interruzioni spostando le estensioni sottostanti.

La migrazione tramite LVM funziona in due modi: Spostare un'estensione o specchiare/demirrorizzare un'estensione. La migrazione LVM utilizza l'efficiente i/o sequenziale a blocchi di grandi dimensioni e solo raramente crea problemi di performance. In tal caso, sono solitamente disponibili opzioni per la riduzione della velocità di i/O. In tal modo, si aumenta il tempo necessario per completare la migrazione, riducendo al contempo il carico di i/o sui sistemi host e di storage.

Specchiatura e demirrorazione

Alcuni gestori di volumi, come AIX LVM, consentono all'utente di specificare il numero di copie per ogni estensione e di controllare quali periferiche ospitano ciascuna copia. La migrazione viene eseguita prelevando un volume logico esistente, eseguendo il mirroring delle estensioni sottostanti nei nuovi volumi, attendendo la sincronizzazione delle copie e rilasciando la copia precedente. Se si desidera un percorso di back-out, è possibile creare un'istantanea dei dati originali prima del punto in cui viene rilasciata la copia speculare. In alternativa, è possibile arrestare brevemente il server per mascherare i LUN originali prima di eliminare forzatamente le copie mirror contenute. In tal modo, si preserva una copia recuperabile dei dati nella loro posizione originale.

Estensione della migrazione

Quasi tutti i gestori di volumi consentono la migrazione delle estensioni e talvolta esistono diverse opzioni. Ad esempio, alcuni responsabili di volume consentono a un amministratore di spostare le singole estensioni per un volume logico specifico dal vecchio al nuovo storage. I gestori di volume come Linux LVM2 offrono `pvmove` Che riposiziona tutti gli extent sul dispositivo LUN specificato in un nuovo LUN. Una volta evacuata, la vecchia LUN può essere rimossa.

Il rischio principale per le operazioni è la rimozione delle LUN vecchie e non utilizzate dalla configurazione. È necessario prestare la massima attenzione quando si modifica la suddivisione in zone FC e si rimuovono i dispositivi LUN obsoleti.

Gestione automatica dello storage Oracle

Oracle ASM è un volume manager e un file system logici combinati. A un livello elevato, Oracle ASM prende una raccolta di LUN, le suddivide in piccole unità di allocazione e le presenta come un singolo volume noto come gruppo di dischi ASM. ASM include inoltre la possibilità di eseguire il mirroring del gruppo di dischi impostando il livello di ridondanza. Un volume può essere senza mirror (ridondanza esterna), con mirroring (ridondanza normale) o con mirroring a tre vie (ridondanza elevata). Prestare attenzione durante la configurazione del livello di ridondanza perché non può essere modificato dopo la creazione.

ASM fornisce anche funzionalità di file system. Sebbene il file system non sia visibile direttamente dall'host, il

database Oracle può creare, spostare ed eliminare file e directory in un gruppo di dischi ASM. Inoltre, è possibile navigare nella struttura utilizzando l'utilità asmcmd.

Come per altre implementazioni LVM, Oracle ASM ottimizza le performance di i/o mediante lo striping e il bilanciamento del carico dell'i/o di ciascun file su tutti i LUN disponibili. In secondo luogo, è possibile riposizionare le estensioni sottostanti per consentire sia il ridimensionamento del gruppo di dischi ASM sia la migrazione. Oracle ASM automatizza il processo mediante l'operazione di ribilanciamento. Le nuove LUN vengono aggiunte a un gruppo di dischi ASM e le vecchie LUN vengono eliminate, innescando il trasferimento dell'estensione e la successiva caduta della LUN evacuata dal gruppo di dischi. Questo processo è uno dei metodi di migrazione più comprovati e l'affidabilità di ASM nel fornire una migrazione trasparente è probabilmente la sua caratteristica più importante.

Poiché il livello di mirroring di Oracle ASM è fisso, non può essere utilizzato con il metodo di migrazione mirror e demirroring.

Migrazione a livello di storage

Migrazione a livello di storage: Migrazione al di sotto del livello dell'applicazione e del sistema operativo. In passato, questo a volte significava l'utilizzo di dispositivi specializzati che copiano i LUN a livello di rete, ma queste funzionalità ora si trovano in modo nativo in ONTAP.

SnapMirror

La migrazione di database da un sistema NetApp all'altro viene eseguita quasi universalmente con il software di replica dei dati NetApp SnapMirror. Il processo prevede la configurazione di una relazione di mirroring per i volumi da migrare, in modo che possano essere sincronizzati e quindi in attesa della finestra di cutover. Quando arriva, il database di origine viene arrestato, viene eseguito un aggiornamento finale del mirror e il mirror viene interrotto. I volumi di replica sono quindi pronti per l'uso, montando una directory del file system NFS contenuta oppure rilevando i LUN contenuti e avviando il database.

Il riposizionamento dei volumi in un singolo cluster ONTAP non viene preso in considerazione dalla migrazione, ma piuttosto da una routine `volume move` operazione. SnapMirror viene utilizzato come motore di replica dei dati all'interno del cluster. Questo processo è completamente automatizzato. Non esistono ulteriori passaggi da eseguire per la migrazione quando gli attributi del volume, come la mappatura delle LUN o le autorizzazioni di esportazione NFS, vengono spostati con il volume stesso. Il trasferimento non comporta interruzioni per le operazioni dell'host. In alcuni casi, l'accesso alla rete deve essere aggiornato per garantire che l'accesso ai dati appena ricollocati sia nel modo più efficiente possibile, ma anche queste attività non comportano interruzione delle attività.

Importazione di LUN esterne (FLI)

FLI è una funzione che consente a un sistema Data ONTAP con versione 8,3 o superiore di migrare una LUN esistente da un altro storage array. La procedura è semplice: Il sistema ONTAP viene sottoposto a zoning sull'array di storage esistente come se fosse un qualsiasi altro host SAN. Data ONTAP può quindi controllare le LUN legacy desiderate ed eseguire la migrazione dei dati sottostanti. Inoltre, il processo di importazione utilizza le impostazioni di efficienza del nuovo volume durante la migrazione dei dati, vale a dire che i dati possono essere compressi e deduplicati inline durante il processo di migrazione.

La prima implementazione di FLI in Data ONTAP 8,3 consentiva solo la migrazione offline. Si trattava di un trasferimento molto veloce, ma i dati LUN continuavano a non essere disponibili fino al completamento della migrazione. La migrazione online è stata introdotta in Data ONTAP 8,3.1. Questo tipo di migrazione consente di ridurre al minimo le interruzioni, consentendo a ONTAP di fornire dati LUN durante il processo di trasferimento. Si verifica una breve interruzione mentre l'host viene sottoposto a zoning per l'utilizzo dei LUN tramite ONTAP. Tuttavia, non appena tali modifiche vengono apportate, i dati sono ancora una volta accessibili

e rimangono accessibili per l'intero processo di migrazione.

L'i/o in lettura viene fornito con un proxy tramite ONTAP fino al completamento dell'operazione di copia, mentre l'i/o in scrittura viene scritta in modo sincrono su LUN esterna e ONTAP. Le due copie LUN vengono mantenute sincronizzate in questo modo fino a quando l'amministratore non esegue un cutover completo che rilascia la LUN esterna e non replica più le scritture.

FLI è progettato per funzionare con FC, ma se si desidera passare a iSCSI, la LUN migrata può essere facilmente rimappata come una LUN iSCSI al termine della migrazione.

Tra le caratteristiche di FLI vi è il rilevamento e la regolazione automatici dell'allineamento. In questo contesto, il termine allineamento si riferisce a una partizione su un dispositivo LUN. Per ottenere prestazioni ottimali è necessario allineare l'i/o ai blocchi da 4K KB. Se una partizione viene posizionata su un offset che non è multiplo di 4K, le prestazioni ne risentono.

Esiste un secondo aspetto dell'allineamento che non può essere corretto regolando un offset di partizione, ovvero la dimensione del blocco del file system. Ad esempio, un file system ZFS generalmente utilizza per impostazione predefinita una dimensione di blocco interna di 512 byte. Altri clienti che utilizzano AIX hanno occasionalmente creato file system JFS2 con dimensioni blocco di 512 o 1, 024 byte. Anche se il file system potrebbe essere allineato a un limite di 4K, i file creati all'interno di tale file system non lo sono e le prestazioni ne risentono.

FLI non deve essere usato in queste circostanze. Anche se i dati sono accessibili dopo la migrazione, il risultato sono file system con gravi limitazioni delle prestazioni. In linea di principio, qualsiasi file system che supporti un carico di lavoro di sovrascrittura casuale su ONTAP dovrebbe utilizzare una dimensione del blocco di 4K KB. Ciò è applicabile principalmente a workload come file di dati di database e implementazioni di VDI. La dimensione del blocco può essere identificata utilizzando i comandi del sistema operativo host pertinente.

Ad esempio, su AIX, la dimensione del blocco può essere visualizzata con `lsfs -q`. Con Linux, `xfs_info` e. `tune2fs` può essere utilizzato per `xfs` e. `ext3/ext4`, rispettivamente. Con `zfs`, il comando è `zdb -C`.

Il parametro che controlla la dimensione del blocco è `ashift` e generalmente il valore predefinito è 9, che significa 2^9, o 512 byte. Per prestazioni ottimali, la `ashift` Il valore deve essere 12 (2^12=4K). Questo valore viene impostato al momento della creazione di zpool e non può essere modificato, il che significa che i data zpool con un `ashift` oltre a 12 deve essere eseguita la migrazione copiando i dati in uno zpool appena creato.

Oracle ASM non ha dimensioni dei blocchi fondamentali. L'unico requisito è che la partizione su cui è stato creato il disco ASM sia allineata correttamente.

7-Mode Transition Tool

7-Mode Transition Tool (7MTT) è un'utility di automazione utilizzata per migrare configurazioni 7- Mode di grandi dimensioni a ONTAP. La maggior parte dei clienti che gestiscono i database trovano altri metodi più semplici, in parte perché eseguono di solito la migrazione dei database piuttosto che trasferire l'intero footprint dello storage. Inoltre, i database sono spesso solo una parte di un ambiente di storage più ampio. Pertanto, spesso i database vengono migrati singolarmente, quindi l'ambiente rimanente può essere spostato con 7MTT.

Alcuni clienti con sistemi di storage dedicati a ambienti di database complicati hanno un numero limitato ma significativo di essi. Questi ambienti potrebbero contenere molti volumi, snapshot e numerosi dettagli di configurazione, come autorizzazioni di esportazione, gruppi iniziatori LUN, autorizzazioni utente e configurazione del protocollo Lightweight Directory Access Protocol. In questi casi, le capacità di automazione di 7MTT possono semplificare una migrazione.

7MTT può funzionare in una delle due modalità seguenti:

- **Copy- Based Transition (CBT).** 7MTT con CBT imposta i volumi SnapMirror da un sistema 7- Mode esistente nel nuovo ambiente. Una volta sincronizzati i dati, 7MTT orchestra il processo di cutover.
- **Copy- Free Transition (CFT).** 7MTT con CFT si basa sulla conversione in-place degli shelf di dischi 7-Mode esistenti. I dati non vengono copiati e gli shelf di dischi esistenti possono essere riutilizzati. La configurazione esistente di data Protection ed efficienza dello storage viene preservata.

La differenza principale tra queste due opzioni consiste nel fatto che la transizione senza copie è un approccio a big-bang, in cui tutti gli shelf di dischi collegati alla coppia ha 7- Mode originale devono essere ricollocati nel nuovo ambiente. Non esiste alcuna opzione per spostare un sottoinsieme di shelf. L'approccio basato sulla copia consente lo spostamento dei volumi selezionati. Esiste anche potenzialmente una finestra di cutover più lunga con transizione priva di copie a causa del legame necessario per la riselezione degli shelf di dischi e la conversione dei metadati. In base all'esperienza sul campo, NetApp consiglia di lasciare trascorrere 1 ora per il riposizionamento e il ripristino degli shelf di dischi e tra 15 minuti e 2 ore per la conversione dei metadati.

Migrazione di file dati

È possibile spostare singoli file di dati Oracle con un singolo comando.

Ad esempio, il comando seguente sposta il file dati IOPST.dbf dal filesystem `/oradata2` al filesystem `/oradata3`.

```
SQL> alter database move datafile  '/oradata2/NTAP/IOPS002.dbf' to
'/oradata3/NTAP/IOPS002.dbf';
Database altered.
```

Lo spostamento di un file dati con questo metodo può essere lento, ma in genere non dovrebbe produrre i/o sufficienti da interferire con i carichi di lavoro del database quotidiani. Al contrario, la migrazione tramite il ribilanciamento di ASM può essere eseguita molto più rapidamente, ma con il rischio di rallentare il database globale durante lo spostamento dei dati.

È possibile misurare facilmente il tempo necessario per spostare i file di dati creando un file di dati di test e spostandolo. Il tempo trascorso per l'operazione viene registrato nei dati di v$session:

```
SQL> set linesize 300;
SQL> select elapsed_seconds||':'||message from v$session_longops;
ELAPSED_SECONDS||':'||MESSAGE
------------------------------------------------------------------------
---------------
351:Online data file move: data file 8: 22548578304 out of 22548578304
bytes done
SQL> select bytes / 1024 / 1024 /1024 as GB from dba_data_files where
FILE_ID = 8;
        GB
----------
        21
```

In questo esempio, il file spostato era datafile 8, della dimensione di 21GB GB e della durata di 6 minuti per la migrazione. Il tempo necessario dipende ovviamente dalle funzionalità del sistema di storage, della rete di

storage e dall'attività complessiva del database che si verifica al momento della migrazione.

Spedizione dei log

L'obiettivo di una migrazione utilizzando la distribuzione dei log è creare una copia dei file di dati originali in una nuova posizione e quindi stabilire un metodo per la distribuzione delle modifiche nel nuovo ambiente.

Una volta stabiliti, è possibile automatizzare la spedizione e la riproduzione dei log per mantenere il database di replica ampiamente sincronizzato con l'origine. Ad esempio, un job cron può essere programmato per (a) copiare i log più recenti nella nuova posizione e (b) riprodurli ogni 15 minuti. In questo modo si riduce al minimo l'interruzione al momento del cutover, in quanto è necessario riprodurre non più di 15 minuti dei registri di archivio.

La procedura illustrata di seguito è essenzialmente un'operazione di clonazione del database. La logica illustrata è simile al motore all'interno di NetApp SnapManager per Oracle (SMO) e al plug-in NetApp SnapCenter per Oracle. Alcuni clienti utilizzano la procedura indicata negli script o nei workflow Wfa per le operazioni di cloning personalizzate. Sebbene questa procedura sia più manuale che non utilizzi SMO o SnapCenter, viene comunque rapidamente script e le API di gestione dei dati all'interno di ONTAP semplificano ulteriormente il processo.

Log shipping - dal file system al file system

In questo esempio viene illustrata la migrazione di un database denominato WAFFLE da un normale file system a un altro normale file system situato su un server diverso. Illustra anche l'utilizzo di SnapMirror per eseguire una copia rapida dei file di dati, ma questa non è parte integrante della procedura generale.

Creare il backup del database

Il primo passo consiste nel creare un backup del database. In particolare, questa procedura richiede una serie di file di dati che possono essere utilizzati per la riproduzione del log di archivio.

Ambiente

In questo esempio, il database di origine si trova su un sistema ONTAP. Il metodo più semplice per creare un backup di un database consiste nell'utilizzare uno snapshot. Il database viene messo in modalità di backup a caldo per alcuni secondi mentre un `snapshot create` l'operazione viene eseguita sul volume che ospita i file di dati.

```
SQL> alter database begin backup;
Database altered.
```

```
Cluster01::*> snapshot create -vserver vserver1 -volume jfsc1_oradata
hotbackup
Cluster01::*>
```

```
SQL> alter database end backup;
Database altered.
```

Il risultato è un'istantanea sul disco chiamata `hotbackup` che contiene un'immagine dei file di dati in modalità di backup a caldo. Se combinati con i log di archivio appropriati per rendere i file di dati coerenti, i dati di questa snapshot possono essere utilizzati come base di un ripristino o di un clone. In questo caso, viene replicato sul nuovo server.

Ripristino in un nuovo ambiente

Ora il backup deve essere ripristinato nel nuovo ambiente. Questa operazione può essere eseguita in vari modi, tra cui Oracle RMAN, ripristino da un'applicazione di backup come NetBackup o semplice operazione di copia dei file di dati inseriti in modalità hot backup.

In questo esempio, SnapMirror viene utilizzato per replicare l'hot backup dello snapshot in una nuova posizione.

1. Creare un nuovo volume per ricevere i dati dello snapshot. Inizializzare il mirroring da `jfsc1_oradata` a. `vol_oradata`.

```
Cluster01::*> volume create -vserver vserver1 -volume vol_oradata
-aggregate data_01 -size 20g -state online -type DP -snapshot-policy
none -policy jfsc3
[Job 833] Job succeeded: Successful
```

```
Cluster01::*> snapmirror initialize -source-path vserver1:jfsc1_oradata
-destination-path vserver1:vol_oradata
Operation is queued: snapmirror initialize of destination
"vserver1:vol_oradata".
Cluster01::*> volume mount -vserver vserver1 -volume vol_oradata
-junction-path /vol_oradata
Cluster01::*>
```

2. Una volta impostato lo stato da SnapMirror, a indicare che la sincronizzazione è completa, aggiornare il mirror in base allo snapshot desiderato,

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields state
source-path             destination-path        state
----------------------- ----------------------- ------------
vserver1:jfsc1_oradata vserver1:vol_oradata SnapMirrored
```

```
Cluster01::*> snapmirror update -destination-path vserver1:vol_oradata
-source-snapshot hotbackup
Operation is queued: snapmirror update of destination
"vserver1:vol_oradata".
```

3. La sincronizzazione può essere verificata visualizzando `newest-snapshot` sul volume speculare.

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields newest-snapshot
source-path             destination-path        newest-snapshot
----------------------- ----------------------- ---------------
vserver1:jfsc1_oradata vserver1:vol_oradata hotbackup
```

4. Lo specchio può quindi essere rotto.

```
Cluster01::> snapmirror break -destination-path vserver1:vol_oradata
Operation succeeded: snapmirror break for destination
"vserver1:vol_oradata".
Cluster01::>
```

5. Montare il nuovo file system.con i file system basati su blocchi, le procedure precise variano in base al LVM in uso. È necessario configurare lo zoning FC o le connessioni iSCSI. Dopo aver stabilito la connettività ai LUN, comandi come Linux `pvscan` Potrebbe essere necessario per rilevare quali gruppi di volumi o LUN devono essere configurati correttamente per essere rilevati da ASM.

   In questo esempio viene utilizzato un semplice file system NFS. Questo file system può essere montato direttamente.

```
fas8060-nfs1:/vol_oradata        19922944   1639360   18283584   9%
/oradata
fas8060-nfs1:/vol_logs            9961472       128    9961344   1%
/logs
```

Creare un modello di creazione controlfile

Successivamente, è necessario creare un modello controlfile. Il `backup controlfile to trace` comando crea comandi di testo per ricreare un controlfile. In alcuni casi, questa funzione può risultare utile per ripristinare un database da un backup e viene spesso utilizzata con script che eseguono attività come la clonazione dei database.

1. L'output del comando seguente viene utilizzato per ricreare i file di controllo per il database migrato.

```
SQL> alter database backup controlfile to trace as '/tmp/waffle.ctrl';
Database altered.
```

2. Una volta creati i file di controllo, copiarli nel nuovo server.

```
[oracle@jfsc3 tmp]$ scp oracle@jfsc1:/tmp/waffle.ctrl /tmp/
oracle@jfsc1's password:
waffle.ctrl                                               100% 5199
5.1KB/s   00:00
```

File dei parametri di backup

Nel nuovo ambiente è necessario anche un file di parametri. Il metodo più semplice consiste nel creare un pfile dal file spfile o pfile corrente. In questo esempio, il database di origine utilizza un spfile.

```
SQL> create pfile='/tmp/waffle.tmp.pfile' from spfile;
File created.
```

Crea voce oratab

La creazione di una voce oratab è necessaria per il corretto funzionamento di utility come oraenv. Per creare una voce oratab, completare il passaggio seguente.

```
WAFFLE:/orabin/product/12.1.0/dbhome_1:N
```

Preparare la struttura delle directory

Se le directory richieste non sono già presenti, è necessario crearle oppure la procedura di avvio del database non riesce. Per preparare la struttura di directory, completare i seguenti requisiti minimi.

```
[oracle@jfsc3 ~]$ . oraenv
ORACLE_SID = [oracle] ? WAFFLE
The Oracle base has been set to /orabin
[oracle@jfsc3 ~]$ cd $ORACLE_BASE
[oracle@jfsc3 orabin]$ cd admin
[oracle@jfsc3 admin]$ mkdir WAFFLE
[oracle@jfsc3 admin]$ cd WAFFLE
[oracle@jfsc3 WAFFLE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

Aggiornamenti del file dei parametri

1. Per copiare il file dei parametri nel nuovo server, eseguire i seguenti comandi. La posizione predefinita è `$ORACLE_HOME/dbs` directory. In questo caso, il pfile può essere posizionato ovunque. Viene utilizzata solo come fase intermedia del processo di migrazione.

```
[oracle@jfsc3 admin]$ scp oracle@jfsc1:/tmp/waffle.tmp.pfile
$ORACLE_HOME/dbs/waffle.tmp.pfile
oracle@jfsc1's password:
waffle.pfile                                          100%  916
0.9KB/s   00:00
```

1. Modificare il file come richiesto. Ad esempio, se la posizione del log di archivio è stata modificata, il file pfile deve essere modificato per riflettere la nuova posizione. In questo esempio, vengono ricollocati solo i file di controllo, in parte per distribuirli tra i file system di log e di dati.

```
[root@jfsc1 tmp]# cat waffle.pfile
WAFFLE.__data_transfer_cache_size=0
WAFFLE.__db_cache_size=507510784
WAFFLE.__java_pool_size=4194304
WAFFLE.__large_pool_size=20971520
WAFFLE.__oracle_base='/orabin'#ORACLE_BASE set from environment
WAFFLE.__pga_aggregate_target=268435456
WAFFLE.__sga_target=805306368
WAFFLE.__shared_io_pool_size=29360128
WAFFLE.__shared_pool_size=234881024
WAFFLE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/WAFFLE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/oradata//WAFFLE/control01.ctl','/oradata//WAFFLE/contr
ol02.ctl'
*.control_files='/oradata/WAFFLE/control01.ctl','/logs/WAFFLE/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='WAFFLE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=WAFFLEXDB)'
*.log_archive_dest_1='LOCATION=/logs/WAFFLE/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

2. Al termine delle modifiche, creare un file spfile basato su questo file pfile.

```
SQL> create spfile from pfile='waffle.tmp.pfile';
File created.
```

Ricreare i file di controllo

In una fase precedente, l'output di `backup controlfile to trace` è stato copiato nel nuovo server. La parte specifica dell'uscita richiesta è la `controlfile recreation` comando. Queste informazioni si trovano nel file sotto la sezione contrassegnata `Set #1. NORESETLOGS`. Inizia con la linea `create controlfile reuse database` e dovrebbe includere la parola `noresetlogs`. Termina con il punto e virgola (;).

1. In questa procedura di esempio, il file viene letto come segue.

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/logs/WAFFLE/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/logs/WAFFLE/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/logs/WAFFLE/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

2. Modificare lo script come desiderato per riflettere la nuova posizione dei vari file. Ad esempio, alcuni file di dati noti per supportare un i/o elevato potrebbero essere reindirizzati a un file system su un Tier di storage dalle performance elevate. In altri casi, le modifiche possono essere apportate solo per motivi di amministrazione, ad esempio isolando i file di dati di un PDB in volumi dedicati.
3. In questo esempio, il `DATAFILE` stanza viene lasciata invariata, ma i log di redo vengono spostati in una nuova posizione in `/redo` piuttosto che condividere lo spazio con i log di archivio `/logs`.

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
SQL> CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
  2      MAXLOGFILES 16
  3      MAXLOGMEMBERS 3
  4      MAXDATAFILES 100
  5      MAXINSTANCES 8
  6      MAXLOGHISTORY 292
  7  LOGFILE
  8    GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  9    GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
 10    GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
 11  -- STANDBY LOGFILE
 12  DATAFILE
 13    '/oradata/WAFFLE/system01.dbf',
 14    '/oradata/WAFFLE/sysaux01.dbf',
 15    '/oradata/WAFFLE/undotbs01.dbf',
 16    '/oradata/WAFFLE/users01.dbf'
 17  CHARACTER SET WE8MSWIN1252
 18  ;
Control file created.
SQL>
```

Se i file sono posizionati in modo errato o i parametri non sono configurati correttamente, vengono generati errori che indicano ciò che deve essere corretto. Il database è montato, ma non è ancora aperto e non può essere aperto perché i file di dati in uso sono ancora contrassegnati come in modalità di backup a caldo. Per rendere il database coerente, è necessario applicare prima i registri di archiviazione.

Replica iniziale del registro

Per rendere coerenti i file di dati è necessaria almeno un'operazione di risposta del registro. Sono disponibili molte opzioni per la riproduzione dei registri. In alcuni casi, la posizione originale del log di archivio sul server originale può essere condivisa tramite NFS e la risposta del log può essere effettuata direttamente. In altri casi, è necessario copiare i registri di archivio.

Ad esempio, un semplice `scp` l'operazione può copiare tutti i log correnti dal server di origine al server di migrazione:

```
[oracle@jfsc3 arch]$ scp jfsc1:/logs/WAFFLE/arch/* ./
oracle@jfsc1's password:
1_22_912662036.dbf                                          100%   47MB
47.0MB/s   00:01
1_23_912662036.dbf                                          100%   40MB
40.4MB/s   00:00
1_24_912662036.dbf                                          100%   45MB
45.4MB/s   00:00
1_25_912662036.dbf                                          100%   41MB
40.9MB/s   00:01
1_26_912662036.dbf                                          100%   39MB
39.4MB/s   00:00
1_27_912662036.dbf                                          100%   39MB
38.7MB/s   00:00
1_28_912662036.dbf                                          100%   40MB
40.1MB/s   00:01
1_29_912662036.dbf                                          100%   17MB
16.9MB/s   00:00
1_30_912662036.dbf                                          100%  636KB
636.0KB/s   00:00
```

Riproduzione del registro iniziale

Una volta che i file si trovano nella posizione del log di archivio, è possibile riprodurli inviando il comando `recover database until cancel` seguito dalla risposta `AUTO` per riprodurre automaticamente tutti i registri disponibili.

```
SQL> recover database until cancel;
ORA-00279: change 382713 generated at 05/24/2016 09:00:54 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_23_912662036.dbf
ORA-00280: change 382713 for thread 1 is in sequence #23
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 405712 generated at 05/24/2016 15:01:05 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_24_912662036.dbf
ORA-00280: change 405712 for thread 1 is in sequence #24
ORA-00278: log file '/logs/WAFFLE/arch/1_23_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
ORA-00278: log file '/logs/WAFFLE/arch/1_30_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_31_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

La risposta finale del log di archivio riporta un errore, ma questo è normale. Il registro indica che `sqlplus` stava cercando un particolare file di registro e non lo ha trovato. Il motivo è, molto probabilmente, che il file di registro non esiste ancora.

Se il database di origine può essere arrestato prima di copiare i registri di archivio, questa operazione deve essere eseguita una sola volta. I log di archivio vengono copiati e riprodotti, quindi il processo può continuare direttamente con il processo di cutover che replica i log di ripristino critici.

Replica e riproduzione incrementale dei log

Nella maggior parte dei casi, la migrazione non viene eseguita immediatamente. Il completamento del processo di migrazione potrebbe richiedere alcuni giorni o addirittura settimane, pertanto i log devono essere inviati continuamente al database di replica e riprodotti. Pertanto, quando arriva il cutover, occorre trasferire e riprodurre minimi dati.

In questo modo è possibile eseguire script in molti modi diversi, ma uno dei metodi più diffusi è l'utilizzo di rsync, un'utilità comune di replica dei file. Il modo più sicuro per usare questa utility è configurarla come demone. Ad esempio, il `rsyncd.conf` file che segue mostra come creare una risorsa chiamata `waffle.arch` A cui si accede con le credenziali utente Oracle e a cui è mappato `/logs/WAFFLE/arch`. Soprattutto, la risorsa è impostata su sola lettura, consentendo la lettura dei dati di produzione, ma non l'alterazione.

```
[root@jfsc1 arch]# cat /etc/rsyncd.conf
[waffle.arch]
   uid=oracle
   gid=dba
   path=/logs/WAFFLE/arch
   read only = true
[root@jfsc1 arch]# rsync --daemon
```

Il seguente comando sincronizza la destinazione del log di archivio del nuovo server con la risorsa rsync `waffle.arch` sul server originale. Il `t` argomento in `rsync - potg` fa sì che l'elenco di file venga confrontato in base alla data e all'ora e che vengano copiati solo i nuovi file. Questo processo fornisce un aggiornamento incrementale del nuovo server. Questo comando può anche essere programmato in cron per essere eseguito regolarmente.

```
[oracle@jfsc3 arch]$ rsync -potg --stats --progress jfsc1::waffle.arch/*
/logs/WAFFLE/arch/
1_31_912662036.dbf
      650240 100%  124.02MB/s    0:00:00 (xfer#1, to-check=8/18)
1_32_912662036.dbf
     4873728 100%  110.67MB/s    0:00:00 (xfer#2, to-check=7/18)
1_33_912662036.dbf
     4088832 100%   50.64MB/s    0:00:00 (xfer#3, to-check=6/18)
1_34_912662036.dbf
     8196096 100%   54.66MB/s    0:00:00 (xfer#4, to-check=5/18)
1_35_912662036.dbf
    19376128 100%   57.75MB/s    0:00:00 (xfer#5, to-check=4/18)
1_36_912662036.dbf
       71680 100%  201.15kB/s    0:00:00 (xfer#6, to-check=3/18)
1_37_912662036.dbf
     1144320 100%    3.06MB/s    0:00:00 (xfer#7, to-check=2/18)
1_38_912662036.dbf
    35757568 100%   63.74MB/s    0:00:00 (xfer#8, to-check=1/18)
1_39_912662036.dbf
      984576 100%    1.63MB/s    0:00:00 (xfer#9, to-check=0/18)
Number of files: 18
Number of files transferred: 9
Total file size: 399653376 bytes
Total transferred file size: 75143168 bytes
Literal data: 75143168 bytes
Matched data: 0 bytes
File list size: 474
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 204
Total bytes received: 75153219
sent 204 bytes  received 75153219 bytes  150306846.00 bytes/sec
total size is 399653376  speedup is 5.32
```

Una volta ricevuti i registri, è necessario riprodurli. Gli esempi precedenti mostrano l'uso di sqlplus per l'esecuzione manuale `recover database until cancel`, un processo che può essere facilmente automatizzato. Nell'esempio illustrato viene utilizzato lo script descritto nella "Riproduci i registri sul database". Gli script accettano un argomento che specifica il database che richiede un'operazione di riproduzione. Ciò consente di utilizzare lo stesso script in una migrazione di più database.

```
[oracle@jfsc3 logs]$ ./replay.logs.pl WAFFLE
ORACLE_SID = [WAFFLE] ? The Oracle base remains unchanged with value
/orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu May 26 10:47:16 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 814256 generated at 05/26/2016 04:52:30 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_32_912662036.dbf
ORA-00280: change 814256 for thread 1 is in sequence #32
ORA-00278: log file '/logs/WAFFLE/arch/1_31_912662036.dbf' no longer
needed for
this recovery
ORA-00279: change 814780 generated at 05/26/2016 04:53:04 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_33_912662036.dbf
ORA-00280: change 814780 for thread 1 is in sequence #33
ORA-00278: log file '/logs/WAFFLE/arch/1_32_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
ORA-00278: log file '/logs/WAFFLE/arch/1_39_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

Cutover

Quando si è pronti per il passaggio al nuovo ambiente, è necessario eseguire una sincronizzazione finale che includa sia i registri di archivio che i registri di ripristino. Se la posizione originale del log di ripristino non è già nota, è possibile identificarla come segue:

```
SQL> select member from v$logfile;
MEMBER
--------------------------------------------------------------------------------
------
/logs/WAFFLE/redo/redo01.log
/logs/WAFFLE/redo/redo02.log
/logs/WAFFLE/redo/redo03.log
```

1. Arrestare il database di origine.
2. Eseguire una sincronizzazione finale dei registri di archivio sul nuovo server con il metodo desiderato.
3. I log di ripristino di origine devono essere copiati nel nuovo server. In questo esempio, i log di ripristino sono stati spostati in una nuova directory all'indirizzo `/redo`.

   ```
   [oracle@jfsc3 logs]$ scp jfsc1:/logs/WAFFLE/redo/* /redo/
   oracle@jfsc1's password:
   redo01.log
   100%   50MB  50.0MB/s   00:01
   redo02.log
   100%   50MB  50.0MB/s   00:00
   redo03.log
   100%   50MB  50.0MB/s   00:00
   ```

4. In questa fase, il nuovo ambiente di database contiene tutti i file necessari per portarlo nello stesso stato dell'origine. I registri di archivio devono essere riprodotti una volta finale.

```
SQL> recover database until cancel;
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

5. Al termine, i log di ripristino devono essere riprodotti. Se il messaggio `Media recovery complete` viene restituito, il processo è riuscito e i database sono sincronizzati e possono essere aperti.

```
SQL> recover database;
Media recovery complete.
SQL> alter database open;
Database altered.
```

Log shipping - da ASM a file system

In questo esempio viene illustrato l'utilizzo di Oracle RMAN per la migrazione di un database. È molto simile all'esempio precedente di distribuzione del log del file system, ma i file su ASM non sono visibili all'host. Le uniche opzioni per la migrazione dei dati presenti sui dispositivi ASM sono il riposizionamento del LUN ASM o l'utilizzo di Oracle RMAN per eseguire le operazioni di copia.

Sebbene RMAN sia un requisito per la copia dei file da Oracle ASM, l'utilizzo di RMAN non è limitato a ASM. RMAN può essere utilizzato per migrare da qualsiasi tipo di storage a qualsiasi altro tipo.

Questo esempio mostra il trasferimento di un database chiamato PANCAKE dallo storage ASM a un file system normale situato su un server diverso nei percorsi `/oradata` e. `/logs`.

Creare il backup del database

Il primo passo consiste nel creare un backup del database da migrare su un server alternativo. Poiché l'origine utilizza Oracle ASM, è necessario utilizzare RMAN. Un semplice backup RMAN può essere eseguito come segue. Questo metodo crea un backup con tag che può essere facilmente identificato da RMAN più avanti nella procedura.

Il primo comando definisce il tipo di destinazione per il backup e la posizione da utilizzare. Il secondo avvia il

backup dei soli file di dati.

```
RMAN> configure channel device type disk format '/rman/pancake/%U';
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters are successfully stored
RMAN> backup database tag 'ONTAP_MIGRATION';
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=251 device type=DISK
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/PANCAKE/system01.dbf
input datafile file number=00002 name=+ASM0/PANCAKE/sysaux01.dbf
input datafile file number=00003 name=+ASM0/PANCAKE/undotbs101.dbf
input datafile file number=00004 name=+ASM0/PANCAKE/users01.dbf
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1gr6c161_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:03
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1hr6c164_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

Backup controlfile

Un controlfile di backup è necessario più avanti nella procedura per `duplicate database` operazione.

```
RMAN> backup current controlfile format '/rman/pancake/ctrl.bkp';
Starting backup at 24-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/ctrl.bkp tag=TAG20160524T032651 comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

File dei parametri di backup

Nel nuovo ambiente è necessario anche un file di parametri. Il metodo più semplice consiste nel creare un pfile dal file spfile o pfile corrente. In questo esempio, il database di origine utilizza un spfile.

```
RMAN> create pfile='/rman/pancake/pfile' from spfile;
Statement processed
```

Script di ridenominazione file ASM

Diverse posizioni dei file attualmente definite nei file di controllo cambiano quando il database viene spostato. Lo script seguente crea uno script RMAN per semplificare il processo. Questo esempio mostra un database con un numero molto ridotto di file di dati, ma in genere i database contengono centinaia o addirittura migliaia di file di dati.

Questo script si trova in "Conversione da ASM a nome file system" e fa due cose.

In primo luogo, viene creato un parametro per ridefinire le posizioni del log di ripristino chiamate `log_file_name_convert`. Si tratta essenzialmente di un elenco di campi alternati. Il primo campo rappresenta la posizione di un registro di ripristino corrente, mentre il secondo campo rappresenta la posizione sul nuovo server. Il modello viene quindi ripetuto.

La seconda funzione consiste nel fornire un modello per la ridenominazione dei file di dati. Lo script esegue il ciclo dei file di dati, estrae le informazioni sul nome e sul numero del file e lo formatta come uno script RMAN. Quindi fa lo stesso con i file temporanei. Il risultato è un semplice script rman che può essere modificato come desiderato per assicurarsi che i file vengano ripristinati nella posizione desiderata.

```
SQL> @/rman/mk.rename.scripts.sql
Parameters for log file conversion:
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/NEW_PATH/redo01.log','+ASM0/PANCAKE/redo02.log',
'/NEW_PATH/redo02.log','+ASM0/PANCAKE/redo03.log', '/NEW_PATH/redo03.log'
rman duplication script:
run
{
set newname for datafile 1 to '+ASM0/PANCAKE/system01.dbf';
set newname for datafile 2 to '+ASM0/PANCAKE/sysaux01.dbf';
set newname for datafile 3 to '+ASM0/PANCAKE/undotbs101.dbf';
set newname for datafile 4 to '+ASM0/PANCAKE/users01.dbf';
set newname for tempfile 1 to '+ASM0/PANCAKE/temp01.dbf';
duplicate target database for standby backup location INSERT_PATH_HERE;
}
PL/SQL procedure successfully completed.
```

Acquisire l'output di questa schermata. Il `log_file_name_convert` il parametro viene inserito nel file pfile come descritto di seguito. Il file di dati RMAN rinominato e lo script duplicato devono essere modificati di conseguenza per posizionare i file di dati nelle posizioni desiderate. In questo esempio, sono tutti inseriti `/oradata/pancake`.

```
run
{
set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
set newname for datafile 4 to '/oradata/pancake/users.dbf';
set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
duplicate target database for standby backup location '/rman/pancake';
}
```

Preparare la struttura delle directory

Gli script sono quasi pronti per l'esecuzione, ma prima la struttura di directory deve essere in posizione. Se le directory richieste non sono già presenti, è necessario crearle oppure la procedura di avvio del database non riesce. L'esempio riportato di seguito riflette i requisiti minimi.

```
[oracle@jfsc2 ~]$ mkdir /oradata/pancake
[oracle@jfsc2 ~]$ mkdir /logs/pancake
[oracle@jfsc2 ~]$ cd /orabin/admin
[oracle@jfsc2 admin]$ mkdir PANCAKE
[oracle@jfsc2 admin]$ cd PANCAKE
[oracle@jfsc2 PANCAKE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

Crea voce oratab

Il seguente comando è necessario per il corretto funzionamento di utility come oraenv.

```
PANCAKE:/orabin/product/12.1.0/dbhome_1:N
```

Aggiornamenti dei parametri

Il file pfile salvato deve essere aggiornato per riflettere eventuali modifiche di percorso sul nuovo server. Le modifiche al percorso del file di dati vengono modificate dallo script di duplicazione RMAN e quasi tutti i database richiedono modifiche al `control_files` e. `log_archive_dest` parametri. Potrebbero inoltre essere presenti posizioni dei file di controllo che devono essere modificate e parametri quali `db_create_file_dest` Potrebbe non essere rilevante al di fuori di ASM. Prima di procedere, un DBA esperto deve esaminare attentamente le modifiche proposte.

In questo esempio, le modifiche principali sono le posizioni controlfile, la destinazione di archivio del registro e l'aggiunta di `log_file_name_convert` parametro.

```
PANCAKE.__data_transfer_cache_size=0
PANCAKE.__db_cache_size=545259520
PANCAKE.__java_pool_size=4194304
PANCAKE.__large_pool_size=25165824
PANCAKE.__oracle_base='/orabin'#ORACLE_BASE set from environment
PANCAKE.__pga_aggregate_target=268435456
PANCAKE.__sga_target=805306368
PANCAKE.__shared_io_pool_size=29360128
PANCAKE.__shared_pool_size=192937984
PANCAKE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/PANCAKE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='+ASM0/PANCAKE/control01.ctl','+ASM0/PANCAKE/control02.ctl
'
*.control_files='/oradata/pancake/control01.ctl','/logs/pancake/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='PANCAKE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=PANCAKEXDB)'
*.log_archive_dest_1='LOCATION=+ASM1'
*.log_archive_dest_1='LOCATION=/logs/pancake'
*.log_archive_format='%t_%s_%r.dbf'
'/logs/path/redo02.log'
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/logs/pancake/redo01.log', '+ASM0/PANCAKE/redo02.log',
'/logs/pancake/redo02.log', '+ASM0/PANCAKE/redo03.log',
'/logs/pancake/redo03.log'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

Dopo la conferma dei nuovi parametri, i parametri devono essere applicati. Esistono diverse opzioni, ma la maggior parte dei clienti crea un file spfile basato sul file pfile di testo.

```
bash-4.1$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0 Production on Fri Jan 8 11:17:40 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> create spfile from pfile='/rman/pancake/pfile';
File created.
```

Nomount di avvio

Il passaggio finale prima della replica del database consiste nel visualizzare i processi del database ma non nel montare i file. In questa fase, potrebbero manifestarsi problemi con spfile. Se il `startup nomount` comando non riesce a causa di un errore di parametro, è semplice chiudere, correggere il modello pfile, ricaricarlo come spfile, e riprovare.

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             373296240 bytes
Database Buffers          423624704 bytes
Redo Buffers                5455872 bytes
```

Duplicare il database

Il ripristino del backup RMAN precedente nella nuova posizione richiede più tempo rispetto ad altre fasi di questo processo. Il database deve essere duplicato senza modificare l'ID del database (DBID) o reimpostare i registri. Ciò impedisce l'applicazione dei registri, operazione necessaria per la sincronizzazione completa delle copie.

Connettersi al database con RMAN come aux ed eseguire il comando duplicato del database utilizzando lo script creato in un passaggio precedente.

```
[oracle@jfsc2 pancake]$ rman auxiliary /
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 03:04:56
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to auxiliary database: PANCAKE (not mounted)
RMAN> run
2> {
3> set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
4> set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
5> set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
6> set newname for datafile 4 to '/oradata/pancake/users.dbf';
7> set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
```

```
8> duplicate target database for standby backup location '/rman/pancake';
9> }
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting Duplicate Db at 24-MAY-16
contents of Memory Script:
{
   restore clone standby controlfile from  '/rman/pancake/ctrl.bkp';
}
executing Memory Script
Starting restore at 24-MAY-16
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
channel ORA_AUX_DISK_1: restoring control file
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:01
output file name=/oradata/pancake/control01.ctl
output file name=/logs/pancake/control02.ctl
Finished restore at 24-MAY-16
contents of Memory Script:
{
   sql clone 'alter database mount standby database';
}
executing Memory Script
sql statement: alter database mount standby database
released channel: ORA_AUX_DISK_1
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
contents of Memory Script:
{
   set newname for tempfile  1 to
 "/oradata/pancake/temp.dbf";
   switch clone tempfile all;
   set newname for datafile  1 to
 "/oradata/pancake/pancake.dbf";
   set newname for datafile  2 to
 "/oradata/pancake/sysaux.dbf";
   set newname for datafile  3 to
 "/oradata/pancake/undotbs1.dbf";
   set newname for datafile  4 to
 "/oradata/pancake/users.dbf";
   restore
   clone database
   ;
```

```
}
executing Memory Script
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/pancake/temp.dbf in control file
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting restore at 24-MAY-16
using channel ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: starting datafile backup set restore
channel ORA_AUX_DISK_1: specifying datafile(s) to restore from backup set
channel ORA_AUX_DISK_1: restoring datafile 00001 to
/oradata/pancake/pancake.dbf
channel ORA_AUX_DISK_1: restoring datafile 00002 to
/oradata/pancake/sysaux.dbf
channel ORA_AUX_DISK_1: restoring datafile 00003 to
/oradata/pancake/undotbs1.dbf
channel ORA_AUX_DISK_1: restoring datafile 00004 to
/oradata/pancake/users.dbf
channel ORA_AUX_DISK_1: reading from backup piece
/rman/pancake/1gr6c161_1_1
channel ORA_AUX_DISK_1: piece handle=/rman/pancake/1gr6c161_1_1
tag=ONTAP_MIGRATION
channel ORA_AUX_DISK_1: restored backup piece 1
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:07
Finished restore at 24-MAY-16
contents of Memory Script:
{
   switch clone datafile all;
}
executing Memory Script
datafile 1 switched to datafile copy
input datafile copy RECID=5 STAMP=912655725 file
name=/oradata/pancake/pancake.dbf
datafile 2 switched to datafile copy
input datafile copy RECID=6 STAMP=912655725 file
name=/oradata/pancake/sysaux.dbf
datafile 3 switched to datafile copy
input datafile copy RECID=7 STAMP=912655725 file
name=/oradata/pancake/undotbs1.dbf
datafile 4 switched to datafile copy
input datafile copy RECID=8 STAMP=912655725 file
name=/oradata/pancake/users.dbf
Finished Duplicate Db at 24-MAY-16
```

Replica iniziale del registro

A questo punto è necessario inviare le modifiche dal database di origine a una nuova posizione. In tal caso, potrebbe essere necessario eseguire una combinazione di operazioni. Il metodo più semplice sarebbe fare in modo che RMAN nel database di origine scriva i log di archivio in una connessione di rete condivisa. Se una posizione condivisa non è disponibile, un metodo alternativo consiste nell'utilizzare RMAN per scrivere su un file system locale e quindi utilizzare rcp o rsync per copiare i file.

In questo esempio, il `/rman` Directory è una condivisione NFS disponibile sia per il database originale che per quello migrato.

Una questione importante in questo caso è la `disk format` clausola. Il formato del disco del backup è `%h_%e_%a.dbf`, Che significa che è necessario utilizzare il formato del numero di thread, il numero di sequenza e l'ID di attivazione per il database. Anche se le lettere sono diverse, questa corrisponde alla `log_archive_format='%t_%s_%r.dbf` parametro nel pfile. Questo parametro specifica inoltre i log di archivio nel formato di numero di thread, numero di sequenza e ID di attivazione. Il risultato finale è che i backup del file di registro sull'origine utilizzano una convenzione di denominazione prevista dal database. In questo modo, vengono eseguite operazioni come `recover database` molto più semplice perché sqlplus anticipa correttamente i nomi dei log di archivio da riprodurre.

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/arch/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
released channel: ORA_DISK_1
RMAN> backup as copy archivelog from time 'sysdate-2';
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=373 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=70 STAMP=912658508
output file name=/rman/pancake/logship/1_54_912576125.dbf RECID=123
STAMP=912659482
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=29 STAMP=912654101
output file name=/rman/pancake/logship/1_41_912576125.dbf RECID=124
STAMP=912659483
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=33 STAMP=912654688
output file name=/rman/pancake/logship/1_45_912576125.dbf RECID=152
STAMP=912659514
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=36 STAMP=912654809
output file name=/rman/pancake/logship/1_47_912576125.dbf RECID=153
STAMP=912659515
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

Riproduzione del registro iniziale

Una volta che i file si trovano nella posizione del log di archivio, è possibile riprodurli inviando il comando `recover database until cancel` seguito dalla risposta `AUTO` per riprodurre automaticamente tutti i registri disponibili. Il file dei parametri sta attualmente indirizzando i log di archivio a. `/logs/archive`, Ma non corrisponde alla posizione in cui RMAN è stato utilizzato per salvare i registri. La posizione può essere reindirizzata temporaneamente come segue prima di ripristinare il database.

```
SQL> alter system set log_archive_dest_1='LOCATION=/rman/pancake/logship'
scope=memory;
System altered.
SQL> recover standby database until cancel;
ORA-00279: change 560224 generated at 05/24/2016 03:25:53 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_49_912576125.dbf
ORA-00280: change 560224 for thread 1 is in sequence #49
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 560353 generated at 05/24/2016 03:29:17 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_50_912576125.dbf
ORA-00280: change 560353 for thread 1 is in sequence #50
ORA-00278: log file '/rman/pancake/logship/1_49_912576125.dbf' no longer
needed
for this recovery
...
ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
ORA-00278: log file '/rman/pancake/logship/1_53_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_54_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

La risposta finale del log di archivio riporta un errore, ma questo è normale. L'errore indica che sqlplus stava cercando un particolare file di registro e non lo ha trovato. Il motivo è molto probabile che il file di registro non esista ancora.

Se il database di origine può essere arrestato prima di copiare i registri di archivio, questa operazione deve essere eseguita una sola volta. I log di archivio vengono copiati e riprodotti, quindi il processo può continuare direttamente con il processo di cutover che replica i log di ripristino critici.

Replica e riproduzione incrementale dei log

Nella maggior parte dei casi, la migrazione non viene eseguita immediatamente. Il completamento del processo di migrazione potrebbe richiedere alcuni giorni o addirittura settimane, pertanto i log devono essere inviati continuamente al database di replica e riprodotti. In questo modo si assicura che i dati minimi debbano essere trasferiti e riprodotti all'arrivo del cutover.

Questo processo può essere facilmente gestito tramite script. Ad esempio, è possibile pianificare il seguente comando nel database originale per assicurarsi che la posizione utilizzata per la spedizione dei log venga

aggiornata continuamente.

```
[oracle@jfsc1 pancake]$ cat copylogs.rman
configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
backup as copy archivelog from time 'sysdate-2';
```

```
[oracle@jfsc1 pancake]$ rman target / cmdfile=copylogs.rman
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 04:36:19
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: PANCAKE (DBID=3574534589)
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=369 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:22
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=124 STAMP=912659483
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:23
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_41_912576125.dbf
continuing other job steps, job failed will not be re-run
...
```

```
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:55
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:57
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
Recovery Manager complete.
```

Una volta ricevuti i registri, è necessario riprodurli. Gli esempi precedenti hanno mostrato l'uso di sqlplus per l'esecuzione manuale `recover database until cancel`, che può essere facilmente automatizzato. Nell'esempio illustrato viene utilizzato lo script descritto nella "Replay Logs on Standby Database". Lo script accetta un argomento che specifica il database che richiede un'operazione di riproduzione. Questo processo consente di utilizzare lo stesso script in una migrazione di più database.

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 04:47:10 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 562219 generated at 05/24/2016 04:15:08 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_55_912576125.dbf
ORA-00280: change 562219 for thread 1 is in sequence #55
ORA-00278: log file '/rman/pancake/logship/1_54_912576125.dbf' no longer
needed for this recovery
ORA-00279: change 562370 generated at 05/24/2016 04:19:18 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_56_912576125.dbf
ORA-00280: change 562370 for thread 1 is in sequence #56
ORA-00278: log file '/rman/pancake/logship/1_55_912576125.dbf' no longer
needed for this recovery
...
ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
ORA-00278: log file '/rman/pancake/logship/1_64_912576125.dbf' no longer
needed for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_65_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

Cutover

Quando si è pronti a passare al nuovo ambiente, è necessario eseguire una sincronizzazione finale. Quando si lavora con i normali file system, è facile assicurarsi che il database migrato sia sincronizzato al 100% rispetto all'originale, poiché i log di ripristino originali vengono copiati e riprodotti. Con ASM non esiste un buon modo per farlo. È possibile recuperare facilmente solo i registri di archivio. Per assicurarsi che i dati non vadano persi, è necessario eseguire con attenzione l'arresto finale del database originale.

1. In primo luogo, la base di dati deve essere chiusa, garantendo che non vengano apportate modifiche. Questa chiusura potrebbe includere la disattivazione delle operazioni pianificate, la chiusura dei listener e/o la chiusura delle applicazioni.
2. Una volta eseguita questa operazione, la maggior parte dei DBA crea una tabella fittizia da utilizzare come indicatore dell'arresto.
3. Forzare l'archiviazione di un registro per assicurarsi che la creazione della tabella fittizia sia registrata nei registri di archivio. A tale scopo, eseguire i seguenti comandi:

```
SQL> create table cutovercheck as select * from dba_users;
Table created.
SQL> alter system archive log current;
System altered.
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
```

4. Per copiare l'ultimo dei registri di archivio, eseguire i seguenti comandi. Il database deve essere disponibile ma non aperto.

```
SQL> startup mount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
Database mounted.
```

5. Per copiare i log di archivio, eseguire i seguenti comandi:

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=8 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:24
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:58
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:59:00
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
```

6. Infine, riprodurre i log di archivio rimanenti sul nuovo server.

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 05:00:53 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed
for thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 563629 generated at 05/24/2016 04:55:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_66_912576125.dbf
ORA-00280: change 563629 for thread 1 is in sequence #66
ORA-00278: log file '/rman/pancake/logship/1_65_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_66_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

7. In questa fase, replicare tutti i dati. Il database è pronto per essere convertito da un database di standby a un database operativo attivo e quindi aperto.

```
SQL> alter database activate standby database;
Database altered.
SQL> alter database open;
Database altered.
```

8. Verificare la presenza della tabella fittizia e poi rilasciarla.

```
SQL> desc cutovercheck
 Name                                      Null?    Type
 ----------------------------------------- --------
---------------------------
 USERNAME                                  NOT NULL VARCHAR2(128)
 USER_ID                                   NOT NULL NUMBER
 PASSWORD                                           VARCHAR2(4000)
 ACCOUNT_STATUS                            NOT NULL VARCHAR2(32)
 LOCK_DATE                                          DATE
 EXPIRY_DATE                                        DATE
 DEFAULT_TABLESPACE                        NOT NULL VARCHAR2(30)
 TEMPORARY_TABLESPACE                      NOT NULL VARCHAR2(30)
 CREATED                                   NOT NULL DATE
 PROFILE                                   NOT NULL VARCHAR2(128)
 INITIAL_RSRC_CONSUMER_GROUP                        VARCHAR2(128)
 EXTERNAL_NAME                                      VARCHAR2(4000)
 PASSWORD_VERSIONS                                  VARCHAR2(12)
 EDITIONS_ENABLED                                   VARCHAR2(1)
 AUTHENTICATION_TYPE                                VARCHAR2(8)
 PROXY_ONLY_CONNECT                                 VARCHAR2(1)
 COMMON                                             VARCHAR2(3)
 LAST_LOGIN                                         TIMESTAMP(9) WITH
TIME ZONE
 ORACLE_MAINTAINED                                  VARCHAR2(1)
SQL> drop table cutovercheck;
Table dropped.
```

Migrazione dei log di ripristino senza interruzioni

A volte, un database è organizzato correttamente in generale, ad eccezione dei registri di ripristino. Questo può accadere per molte ragioni, la più comune delle quali è correlata agli snapshot. Prodotti come SnapManager per Oracle, SnapCenter e il framework di gestione dello storage NetApp Snap Creator consentono il ripristino quasi istantaneo di un database, ma solo se vengono ripristinati i volumi dei file di dati. Se i log di redo condividono lo spazio con i file di dati, non è possibile eseguire la reversione in modo sicuro, poiché causerebbe la distruzione dei log di redo, probabilmente la perdita di dati. Pertanto, i log di ripristino devono essere spostati.

Questa procedura è semplice e può essere eseguita senza interruzioni.

Configurazione corrente del log di ripristino

1. Identificare il numero di gruppi di log di ripristino e i rispettivi numeri di gruppo.

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
rows selected.
```

2. Immettere le dimensioni dei registri di ripristino.

```
SQL> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
------------------------------------------------------------------------
--------
1 524288000
2 524288000
3 524288000
```

Creare nuovi registri

1. Per ogni log di ripristino, creare un nuovo gruppo con dimensioni e numero di membri corrispondenti.

```
SQL> alter database add logfile ('/newredo0/redo01a.log',
'/newredo1/redo01b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo02a.log',
'/newredo1/redo02b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo03a.log',
'/newredo1/redo03b.log') size 500M;
Database altered.
SQL>
```

2. Verificare la nuova configurazione.

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
--------------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
4 /newredo0/redo01a.log
4 /newredo1/redo01b.log
5 /newredo0/redo02a.log
5 /newredo1/redo02b.log
6 /newredo0/redo03a.log
6 /newredo1/redo03b.log
12 rows selected.
```

Rilasciare i vecchi registri

1. Rilasciare i vecchi registri (gruppi 1, 2 e 3).

```
SQL> alter database drop logfile group 1;
Database altered.
SQL> alter database drop logfile group 2;
Database altered.
SQL> alter database drop logfile group 3;
Database altered.
```

2. Se si verifica un errore che impedisce di rilasciare un registro attivo, forzare un passaggio al registro successivo per rilasciare il blocco e forzare un checkpoint globale. Fare riferimento al seguente esempio di questo processo. Il tentativo di rilasciare il gruppo di file di registro 2, che si trovava nella vecchia posizione, è stato negato perché in questo file di registro erano ancora presenti dati attivi.

```
SQL> alter database drop logfile group 2;
alter database drop logfile group 2
*
ERROR at line 1:
ORA-01623: log 2 is current log for instance NTAP (thread 1) - cannot
drop
ORA-00312: online log 2 thread 1: '/redo0/NTAP/redo02a.log'
ORA-00312: online log 2 thread 1: '/redo1/NTAP/redo02b.log'
```

3. Un’archiviazione dei log seguita da un punto di verifica consente di rilasciare il file di log.

```
SQL> alter system archive log current;
System altered.
SQL> alter system checkpoint;
System altered.
SQL> alter database drop logfile group 2;
Database altered.
```

4. Quindi, eliminare i log dal file system. Questo processo deve essere eseguito con estrema attenzione.

Copia dei dati dell’host

Come per la migrazione a livello di database, la migrazione nel layer host fornisce un approccio indipendente dal vendor di soluzioni di storage.

In altre parole, talvolta "basta copiare i file" è l’opzione migliore.

Sebbene questo approccio a bassa tecnologia possa sembrare troppo semplice, offre comunque vantaggi significativi in quanto non è richiesto alcun software speciale e i dati originali rimangono intatti in tutta sicurezza durante il processo. Il limite principale è rappresentato dal fatto che la migrazione dei dati di una copia file causa interruzioni, poiché il database deve essere arrestato prima dell’inizio dell’operazione di copia. Non esiste un buon modo per sincronizzare le modifiche all’interno di un file, quindi i file devono essere completamente disattivati prima che la copia abbia inizio.

Se l’arresto richiesto da un’operazione di copia non è desiderabile, l’opzione successiva migliore basata su host è sfruttare un Logical Volume Manager (LVM). Esistono molte opzioni LVM, tra cui Oracle ASM, tutte con funzionalità simili, ma anche con alcune limitazioni che è necessario tenere in considerazione. Nella maggior parte dei casi, la migrazione può essere eseguita senza downtime e interruzioni.

Copia da filesystem a filesystem

L’utilità di una semplice operazione di copia non deve essere sottovalutata. Si tratta di un processo altamente affidabile che non richiede particolari competenze su sistemi operativi, database o sistemi storage. Inoltre, è molto sicuro perché non influisce sui dati originali. In genere, un amministratore di sistema modifica i file system di origine in modo che vengano montati in sola lettura e quindi riavvia un server per garantire che nessun elemento possa danneggiare i dati correnti. Il processo di copia può essere eseguito tramite script per garantire che venga eseguito il più rapidamente possibile senza il rischio di errori dell’utente. Poiché il tipo di i/o è un semplice trasferimento sequenziale dei dati, risulta estremamente efficiente in termini di larghezza di banda.

Nell’esempio seguente viene illustrata un’opzione per una migrazione sicura e rapida.

Ambiente

L’ambiente da migrare è il seguente:

- File system attuali

```
ontap-nfs1:/host1_oradata       52428800  16196928  36231872  31%
/oradata
ontap-nfs1:/host1_logs          49807360    548032  49259328   2% /logs
```

- Nuovi file system

```
ontap-nfs1:/host1_logs_new      49807360       128  49807232   1%
/new/logs
ontap-nfs1:/host1_oradata_new   49807360       128  49807232   1%
/new/oradata
```

Panoramica

Un DBA può migrare il database chiudendo semplicemente il database e copiando i file. Tuttavia, se occorre migrare molti database o ridurre al minimo il downtime, il processo può essere facilmente gestito tramite script. L'utilizzo di script riduce inoltre la possibilità di errori da parte dell'utente.

Gli script di esempio illustrati automatizzano le seguenti operazioni:

- Chiusura del database in corso
- Conversione dei file system esistenti in uno stato di sola lettura
- Copiare tutti i dati dai file system di origine a quelli di destinazione, mantenendo tutte le autorizzazioni dei file
- Smontaggio dei file system vecchi e nuovi
- Rimontaggio dei nuovi file system negli stessi percorsi dei file system precedenti

Procedura

1. Arrestare il database.

```
[root@host1 current]# ./dbshut.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 15:58:48 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
```

2. Convertire i file system in sola lettura. Questa operazione può essere eseguita più rapidamente utilizzando uno script, come illustrato nella "Convertire il file system in sola lettura".

```
[root@host1 current]# ./mk.fs.readonly.pl /oradata
/oradata unmounted
/oradata mounted read-only
[root@host1 current]# ./mk.fs.readonly.pl /logs
/logs unmounted
/logs mounted read-only
```

3. Verificare che i file system siano ora di sola lettura.

```
ontap-nfs1:/host1_oradata on /oradata type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_logs on /logs type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

4. Sincronizzare il contenuto del file system con `rsync` comando.

```
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /oradata/ /new/oradata/
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
```

```
 10737426432 100%  153.50MB/s    0:01:06 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%   12.09MB/s    0:00:01 (xfer#2, to-check=9/13)
...
NTAP/undotbs02.dbf
  1073750016 100%  131.60MB/s    0:00:07 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    3.95MB/s    0:00:01 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  162204017.96 bytes/sec
total size is 18570092218  speedup is 1.00
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /logs/ /new/logs/
sending incremental file list
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%   95.98MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   49.45MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   44.68MB/s    0:00:01 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   68.03MB/s    0:00:00 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
```

```
sent 527098156 bytes  received 278 bytes  95836078.91 bytes/sec
total size is 527032832  speedup is 1.00
```

5. Smontare i vecchi file system e riposizionare i dati copiati. Questa operazione può essere eseguita più rapidamente utilizzando uno script, come illustrato nella "Sostituire il file system".

```
[root@host1 current]# ./swap.fs.pl /logs,/new/logs
/new/logs unmounted
/logs unmounted
Updated /logs mounted
[root@host1 current]# ./swap.fs.pl /oradata,/new/oradata
/new/oradata unmounted
/oradata unmounted
Updated /oradata mounted
```

6. Verificare che i nuovi file system siano in posizione.

```
ontap-nfs1:/host1_logs_new on /logs type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_oradata_new on /oradata type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

7. Avviare il database.

```
[root@host1 current]# ./dbstart.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 16:10:07 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
```

Cutover completamente automatizzato

Questo script di esempio accetta argomenti del SID del database seguiti da coppie di file system delimitate in comune. Per l'esempio sopra illustrato, il comando viene inviato come segue:

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
```

Quando viene eseguito, lo script di esempio tenta di eseguire la seguente sequenza. Termina se incontra un errore in qualsiasi fase:

1. Arrestare il database.
2. Convertire i file system correnti in stato di sola lettura.
3. Utilizzare ciascuna coppia di argomenti del file system delimitati da virgole e sincronizzare il primo file system con il secondo.
4. Smontare i file system precedenti.
5. Aggiornare `/etc/fstab` archiviare come segue:
   a. Creare un backup in `/etc/fstab.bak`.
   b. Annotare le voci precedenti per i file system precedenti e nuovi.
   c. Creare una nuova voce per il nuovo file system che utilizza il vecchio punto di montaggio.
6. Montare i file system.
7. Avviare il database.

Il testo seguente fornisce un esempio di esecuzione per questo script:

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:05:50 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
sending incremental file list
```

```
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%  185.40MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   81.34MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   70.42MB/s    0:00:00 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   47.08MB/s    0:00:01 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
sent 527098156 bytes  received 278 bytes  150599552.57 bytes/sec
total size is 527032832  speedup is 1.00
Succesfully replicated filesystem /logs to /new/logs
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
 10737426432 100%  176.55MB/s    0:00:58 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%    9.48MB/s    0:00:02 (xfer#2, to-check=9/13)
... NTAP/undotbs01.dbf
   309338112 100%   70.76MB/s    0:00:04 (xfer#9, to-check=2/13)
NTAP/undotbs02.dbf
  1073750016 100%  187.65MB/s    0:00:05 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    5.09MB/s    0:00:00 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
File list generation time: 0.001 seconds
```

```
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  177725933.55 bytes/sec
total size is 18570092218  speedup is 1.00
Succesfully replicated filesystem /oradata to /new/oradata
swap 0 /logs /new/logs
/new/logs unmounted
/logs unmounted
Mounted updated /logs
Swapped filesystem /logs for /new/logs
swap 1 /oradata /new/oradata
/new/oradata unmounted
/oradata unmounted
Mounted updated /oradata
Swapped filesystem /oradata for /new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:08:59 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
[root@host1 current]#
```

Migrazione Oracle ASM spfile e passwd

Una difficoltà nel completare la migrazione che coinvolge ASM è rappresentata dallo spfile specifico per ASM e dal file delle password. Per impostazione predefinita, questi file di metadati critici vengono creati nel primo gruppo di dischi ASM definito. Se un particolare gruppo di dischi ASM deve essere evacuato e rimosso, il file spfile e la password che governano l'istanza ASM deve essere riposizionato.

Un altro caso d'utilizzo in cui potrebbe essere necessario trasferire questi file è durante una distribuzione di software di gestione del database, come SnapManager per Oracle o il plug-in SnapCenter Oracle. Una delle funzionalità di questi prodotti è il ripristino rapido di un database ripristinando lo stato dei LUN ASM che ospitano i file di dati. Per eseguire questa operazione, è necessario portare il gruppo di dischi ASM offline prima di eseguire un ripristino. Questo non è un problema, purché i file di dati di un determinato database siano isolati in un gruppo di dischi ASM dedicato.

Quando il gruppo di dischi contiene anche il file ASM spfile/passwd, l'unico modo per mettere il gruppo di dischi in modalità non in linea è arrestare l'intera istanza ASM. Si tratta di un processo di interruzione, il che significa che il file spfile/passwd dovrebbe essere riposizionato.

Ambiente

1. SID database = TOAST
2. File di dati correnti su `+DATA`
3. File di log e file di controllo correnti attivati `+LOGS`
4. Nuovi gruppi di dischi ASM stabiliti come `+NEWDATA` e. `+NEWLOGS`

Posizioni dei file spfile/passwd ASM

Il trasferimento di questi file può essere eseguito senza interruzione delle attività. Tuttavia, per motivi di sicurezza, NetApp consiglia di arrestare l'ambiente del database in modo da poter essere certi che i file siano stati spostati e che la configurazione sia stata aggiornata correttamente. Questa procedura deve essere ripetuta se su un server sono presenti più istanze ASM.

Identificare le istanze ASM

Identificare le istanze ASM in base ai dati registrati in `oratab` file. Le istanze di ASM sono indicate dal simbolo +.

```
-bash-4.1$ cat /etc/oratab | grep '^+'
+ASM:/orabin/grid:N             # line added by Agent
```

Su questo server è presente un'istanza ASM denominata +ASM.

Assicurarsi che tutti i database siano chiusi

L'unico processo di smon visibile dovrebbe essere quello per l'istanza ASM in uso. La presenza di un altro processo di smon indica che un database è ancora in esecuzione.

```
-bash-4.1$ ps -ef | grep smon
oracle     857     1  0 18:26 ?        00:00:00 asm_smon_+ASM
```

L'unico processo di smon è l'istanza ASM stessa. Ciò significa che nessun altro database è in esecuzione ed è sicuro procedere senza il rischio di interrompere le operazioni del database.

Individuare i file

Identificare la posizione corrente del file spfile e della password di ASM utilizzando `spget` e. `pwget` comandi.

```
bash-4.1$ asmcmd
ASMCMD> spget
+DATA/spfile.ora
```

```
ASMCMD> pwget --asm
+DATA/orapwasm
```

I file si trovano entrambi alla base di `+DATA` gruppo di dischi.

Copiare i file

Copiare i file nel nuovo gruppo di dischi ASM con `spcopy` e. `pwcopy` comandi. Se il nuovo gruppo di dischi è stato creato di recente ed è attualmente vuoto, potrebbe essere necessario montarlo per primo.

```
ASMCMD> mount NEWDATA
```

```
ASMCMD> spcopy +DATA/spfile.ora +NEWDATA/spfile.ora
copying +DATA/spfile.ora -> +NEWDATA/spfilea.ora
```

```
ASMCMD> pwcopy +DATA/orapwasm +NEWDATA/orapwasm
copying +DATA/orapwasm -> +NEWDATA/orapwasm
```

I file sono stati copiati da `+DATA` a. `+NEWDATA`.

Aggiornare l'istanza ASM

L'istanza ASM deve ora essere aggiornata per riflettere la modifica della posizione. Il `spset` e. `pwset` I comandi aggiornano i metadati ASM richiesti per l'avvio del gruppo di dischi ASM.

```
ASMCMD> spset +NEWDATA/spfile.ora
ASMCMD> pwset --asm +NEWDATA/orapwasm
```

Attivare ASM utilizzando i file aggiornati

A questo punto, l'istanza ASM utilizza ancora le posizioni precedenti di questi file. L'istanza deve essere riavviata per forzare una rilettura dei file dalle nuove posizioni e per rilasciare i blocchi sui file precedenti.

```
-bash-4.1$ sqlplus / as sysasm
SQL> shutdown immediate;
ASM diskgroups volume disabled
ASM diskgroups dismounted
ASM instance shutdown
```

```
SQL> startup
ASM instance started
Total System Global Area 1140850688 bytes
Fixed Size                  2933400 bytes
Variable Size            1112751464 bytes
ASM Cache                  25165824 bytes
ORA-15032: not all alterations performed
ORA-15017: diskgroup "NEWDATA" cannot be mounted
ORA-15013: diskgroup "NEWDATA" is already mounted
```

Rimuovere i vecchi file spfile e password

Se la procedura è stata eseguita correttamente, i file precedenti non sono più bloccati e possono essere rimossi.

```
-bash-4.1$ asmcmd
ASMCMD> rm +DATA/spfile.ora
ASMCMD> rm +DATA/orapwasm
```

Copia da Oracle ASM a ASM

Oracle ASM è essenzialmente un volume manager e un file system combinati e leggeri. Poiché il file system non è facilmente visibile, è necessario utilizzare RMAN per eseguire operazioni di copia. Sebbene il processo di migrazione basato sulle copie sia sicuro e semplice, si traduce in un'interruzione. È possibile ridurre al minimo le interruzioni, ma non eliminarle completamente.

Se si desidera eseguire la migrazione senza interruzioni di un database basato su ASM, l'opzione migliore è sfruttare la capacità di ASM di riequilibrare le estensioni ASM nei nuovi LUN, eliminando al contempo i vecchi LUN. In genere, questo tipo di operazioni è sicuro e senza interruzioni, ma non offre alcun percorso di back-out. Se si riscontrano problemi di funzionamento o di prestazioni, l'unica opzione è quella di trasferire nuovamente i dati all'origine.

Questo rischio può essere evitato copiando il database nella nuova posizione piuttosto che spostare i dati, in modo che i dati originali non vengano toccati. Il database può essere completamente testato nella sua nuova posizione prima di entrare in funzione e il database originale è disponibile come opzione di fallback se vengono rilevati problemi.

Questa procedura è una delle numerose opzioni che interessano RMAN. È progettato per consentire un processo in due fasi in cui viene creato il backup iniziale e quindi sincronizzato successivamente tramite la riproduzione del registro. Questo processo è auspicabile per ridurre al minimo i tempi di inattività, in quanto consente al database di rimanere operativo e di distribuire i dati durante la copia di base iniziale.

Copia database

Oracle RMAN crea una copia di livello 0 (completa) del database di origine attualmente presente nel gruppo di dischi ASM `+DATA` alla nuova posizione su `+NEWDATA`.

```
-bash-4.1$ rman target /
Recovery Manager: Release 12.1.0.2.0 - Production on Sun Dec 6 17:40:03
2015
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: TOAST (DBID=2084313411)
RMAN> backup as copy incremental level 0 database format '+NEWDATA' tag
'ONTAP_MIGRATION';
Starting backup at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=302 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
...
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
output file name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
tag=ONTAP_MIGRATION RECID=5 STAMP=897759622
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWDATA/TOAST/BACKUPSET/2015_12_06/nnsnn0_ontap_migration_0.262.89
7759623 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

Forzare l'interruttore del registro di archiviazione

È necessario forzare un'opzione del log di archivio per assicurarsi che i log di archivio contengano tutti i dati necessari per rendere la copia completamente coerente. Senza questo comando, i dati chiave potrebbero essere ancora presenti nei log di ripristino.

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

Arrestare il database di origine

L'interruzione inizia in questa fase perché il database viene arrestato e inserito in una modalità di sola lettura ad accesso limitato. Per arrestare il database di origine, eseguire i seguenti comandi:

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                390073456 bytes
Database Buffers             406847488 bytes
Redo Buffers                   5455872 bytes
```

Backup ControlFile

È necessario eseguire il backup di controlfile nel caso in cui sia necessario interrompere la migrazione e ripristinare la posizione di archiviazione originale. Una copia del controlfile di backup non è richiesta al 100%, ma rende più semplice il processo di ripristino delle posizioni dei file di database nella posizione originale.

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=358 device type=DISK
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151206T174753 RECID=6
STAMP=897760073
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

Aggiornamenti dei parametri

Il file spfile corrente contiene riferimenti ai file di controllo nelle posizioni correnti all'interno del vecchio gruppo di dischi ASM. Deve essere modificato, il che è fatto facilmente modificando una versione pfile intermedia.

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

Aggiornare pfile

Aggiornare tutti i parametri che fanno riferimento ai vecchi gruppi di dischi ASM per riflettere i nuovi nomi dei gruppi di dischi ASM. Quindi salvare il file pfile aggiornato. Assicurarsi che il `db_create` parametri presenti.

Nell'esempio seguente, i riferimenti a. `+DATA` che sono stati modificati in `+NEWDATA` sono evidenziati in giallo. Due parametri chiave sono `db_create` parametri che creano nuovi file nella posizione corretta.

```
*.compatible='12.1.0.2.0'
*.control_files='+NEWLOGS/TOAST/CONTROLFILE/current.258.897683139'
*.db_block_size=8192
*. db_create_file_dest='+NEWDATA'
*. db_create_online_log_dest_1='+NEWLOGS'
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=+NEWLOGS'
*.log_archive_format='%t_%s_%r.dbf'
```

Aggiorna il file init.ora

La maggior parte dei database basati su ASM utilizza un `init.ora` file che si trova in `$ORACLE_HOME/dbs` Directory, che è un punto di spfile sul gruppo di dischi ASM. Questo file deve essere reindirizzato a una posizione sul nuovo gruppo di dischi ASM.

```
-bash-4.1$ cd $ORACLE_HOME/dbs
-bash-4.1$ cat initTOAST.ora
SPFILE='+DATA/TOAST/spfileTOAST.ora'
```

Modificare questo file come segue:

```
SPFILE=+NEWLOGS/TOAST/spfileTOAST.ora
```

Ricreazione del file dei parametri

spfile è ora pronto per essere popolato dai dati nel pfile modificato.

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

Avviare il database per iniziare a utilizzare il nuovo spfile

Avviare il database per assicurarsi che utilizzi ora il nuovo spfile creato e che eventuali ulteriori modifiche ai parametri di sistema siano registrate correttamente.

```
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                373296240 bytes
Database Buffers             423624704 bytes
Redo Buffers                   5455872 bytes
```

Ripristina controlfile

Il controlfile di backup creato da RMAN può anche essere ripristinato da RMAN direttamente nella posizione specificata nel nuovo spfile.

```
RMAN> restore controlfile from
'+DATA/TOAST/CONTROLFILE/current.258.897683139';
Starting restore at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=417 device type=DISK
channel ORA_DISK_1: copied control file copy
output file name=+NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
Finished restore at 06-DEC-15
```

Montare il database e verificare l'uso del nuovo controlfile.

```
RMAN> alter database mount;
using target database control file instead of recovery catalog
Statement processed
```

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
+NEWLOGS/TOAST/CONTROLFILE/cur
                                                 rent.273.897761061
```

Riproduzione del registro

Il database utilizza attualmente i file di dati nella vecchia posizione. Prima di poter utilizzare la copia, è necessario sincronizzarla. È trascorso del tempo durante il processo di copia iniziale e le modifiche sono state registrate principalmente nei registri di archivio. Queste modifiche vengono replicate come segue:

1. Eseguire un backup incrementale RMAN, che contiene i registri di archivio.

```
RMAN> backup incremental level 1 format '+NEWLOGS' for recover of copy
with tag 'ONTAP_MIGRATION' database;
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=62 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
input datafile file number=00002
name=+DATA/TOAST/DATAFILE/sysaux.260.897683143
input datafile file number=00003
name=+DATA/TOAST/DATAFILE/undotbs1.257.897683145
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/ncsnn1_ontap_migration_0.267.
897762697 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

2. Riprodurre nuovamente il registro.

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 06-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001
name=+NEWDATA/TOAST/DATAFILE/system.259.897759609
recovering datafile copy file number=00002
name=+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
recovering datafile copy file number=00003
name=+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
recovering datafile copy file number=00004
name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
channel ORA_DISK_1: reading from backup piece
+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.8977626
93
channel ORA_DISK_1: piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

Attivazione

Il controlfile ripristinato fa ancora riferimento ai file di dati nella posizione originale e contiene anche le informazioni di percorso per i file di dati copiati.

1. Per modificare i file di dati attivi, eseguire `switch database to copy` comando.

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/system.259.897759609"
datafile 2 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615"
datafile 3 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619"
datafile 4 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/users.258.897759623"
```

   I file di dati attivi sono ora i file di dati copiati, ma potrebbero comunque essere presenti modifiche nei log di ripristino finali.

2. Per riprodurre tutti i registri rimanenti, eseguire il `recover database` comando. Se il messaggio `media recovery complete` il processo è stato eseguito correttamente.

```
RMAN> recover database;
Starting recover at 06-DEC-15
using channel ORA_DISK_1
starting media recovery
media recovery complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

Questo processo ha modificato solo la posizione dei file di dati normali. I file di dati temporanei devono essere rinominati, ma non devono essere copiati perché sono solo temporanei. Il database è attualmente inattivo, pertanto non sono presenti dati attivi nei file di dati temporanei.

3. Per spostare i file di dati temporanei, identificarne prima la posizione.

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
------------------------------------------------------------------------
--------
1 +DATA/TOAST/TEMPFILE/temp.263.897683145
```

4. Spostare i file di dati temporanei utilizzando un comando RMAN che imposta il nuovo nome per ciascun file di dati. Con Oracle Managed Files (OMF), il nome completo non è necessario; il gruppo di dischi ASM è sufficiente. Quando il database viene aperto, OMF si collega alla posizione appropriata nel gruppo di dischi ASM. Per spostare i file, eseguire i seguenti comandi:

```
run {
set newname for tempfile 1 to '+NEWDATA';
switch tempfile all;
}
```

```
RMAN> run {
2> set newname for tempfile 1 to '+NEWDATA';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to +NEWDATA in control file
```

Migrazione dei log di ripristino

Il processo di migrazione è quasi completo, ma i log di ripristino si trovano ancora nel gruppo di dischi ASM originale. I log di ripristino non possono essere spostati direttamente. Viene invece creata una nuova serie di log di ripristino che viene aggiunta alla configurazione, seguita da una rimozione dei log precedenti.

1. Identificare il numero di gruppi di log di ripristino e i rispettivi numeri di gruppo.

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
------------------------------------------------------------------------------
--------
1 +DATA/TOAST/ONLINELOG/group_1.261.897683139
2 +DATA/TOAST/ONLINELOG/group_2.259.897683139
3 +DATA/TOAST/ONLINELOG/group_3.256.897683139
```

2. Immettere le dimensioni dei registri di ripristino.

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
------------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Per ogni log di ripristino, creare un nuovo gruppo con una configurazione corrispondente. Se non si utilizza OMF, è necessario specificare il percorso completo. Questo è anche un esempio che utilizza `db_create_online_log` parametri. Come mostrato in precedenza, questo parametro era impostato su +NEWLOGS. Questa configurazione consente di utilizzare i seguenti comandi per creare nuovi registri online senza dover specificare un percorso di file o un gruppo di dischi ASM specifico.

```
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
```

4. Aprire il database.

```
SQL> alter database open;
Database altered.
```

5. Rilasciare i vecchi registri.

```
RMAN> alter database drop logfile group 1;
Statement processed
```

6. Se si verifica un errore che impedisce di rilasciare un registro attivo, forzare un passaggio al registro

successivo per rilasciare il blocco e forzare un checkpoint globale. Di seguito è riportato un esempio. Il tentativo di rilasciare il gruppo di file di registro 3, che si trovava nella vecchia posizione, è stato negato perché in questo file di registro erano ancora presenti dati attivi. L'archiviazione di un registro dopo un punto di verifica consente di eliminare il file di registro.

```
RMAN> alter database drop logfile group 3;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 3 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 3 thread 1:
'+LOGS/TOAST/ONLINELOG/group_3.259.897563549'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 3;
Statement processed
```

7. Esaminare l'ambiente per assicurarsi che tutti i parametri basati sulla posizione siano aggiornati.

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

8. Nello script seguente viene illustrato come semplificare questo processo:

```
[root@host1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
+NEWDATA/TOAST/DATAFILE/system.259.897759609
+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
+NEWDATA/TOAST/DATAFILE/users.258.897759623
TOAST redo logs:
+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123
+NEWLOGS/TOAST/ONLINELOG/group_5.265.897763125
+NEWLOGS/TOAST/ONLINELOG/group_6.264.897763125
TOAST temp datafiles:
+NEWDATA/TOAST/TEMPFILE/temp.260.897763165
TOAST spfile
spfile                                string
+NEWDATA/spfiletoast.ora
TOAST key parameters
control_files +NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
log_archive_dest_1 LOCATION=+NEWLOGS
db_create_file_dest +NEWDATA
db_create_online_log_dest_1 +NEWLOGS
```

9. Se i gruppi di dischi ASM sono stati completamente evacuati, è possibile smontarli con `asmcmd`. Tuttavia, in molti casi i file appartenenti ad altri database o al file ASM spfile/passwd potrebbero essere ancora presenti.

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

Copia da Oracle ASM al file system

La procedura di copia da Oracle ASM a file system è molto simile alla procedura di copia da ASM a ASM, con vantaggi e restrizioni simili. La differenza principale è la sintassi dei vari comandi e parametri di configurazione quando si utilizza un file system visibile anziché un gruppo di dischi ASM.

Copia database

Oracle RMAN viene utilizzato per creare una copia di livello 0 (completa) del database di origine attualmente presente nel gruppo di dischi ASM `+DATA` alla nuova posizione su `/oradata`.

```
RMAN> backup as copy incremental level 0 database format
'/oradata/TOAST/%U' tag 'ONTAP_MIGRATION';
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=377 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-
1_01r5fhjg tag=ONTAP_MIGRATION RECID=1 STAMP=911722099
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-
2_02r5fhjo tag=ONTAP_MIGRATION RECID=2 STAMP=911722106
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt tag=ONTAP_MIGRATION RECID=3 STAMP=911722113
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/oradata/TOAST/cf_D-TOAST_id-2098173325_04r5fhk5
tag=ONTAP_MIGRATION RECID=4 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting datafile copy
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-
4_05r5fhk6 tag=ONTAP_MIGRATION RECID=5 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/oradata/TOAST/06r5fhk7_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
```

Forzare l’interruttore del registro di archiviazione

È necessario forzare lo switch del log di archivio per assicurarsi che i log di archivio contengano tutti i dati necessari per rendere la copia completamente coerente. Senza questo comando, i dati chiave potrebbero essere ancora presenti nei log di ripristino. Per forzare un’opzione del log di archivio, eseguire il comando seguente:

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

Arrestare il database di origine

L'interruzione inizia in questa fase perché il database viene arrestato e inserito in una modalità di sola lettura ad accesso limitato. Per arrestare il database di origine, eseguire i seguenti comandi:

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

Backup ControlFile

Eseguire il backup dei file di controllo nel caso in cui sia necessario interrompere la migrazione e ripristinare la posizione di archiviazione originale. Una copia del controlfile di backup non è richiesta al 100%, ma rende più semplice il processo di ripristino delle posizioni dei file di database nella posizione originale.

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 08-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151208T194540 RECID=30
STAMP=897939940
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 08-DEC-15
```

Aggiornamenti dei parametri

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

Aggiornare pfile

Tutti i parametri che fanno riferimento ai vecchi gruppi di dischi ASM devono essere aggiornati e, in alcuni casi, eliminati quando non sono più rilevanti. Aggiornarli per riflettere i nuovi percorsi del file system e salvare il file pfile aggiornato. Assicurarsi che sia elencato il percorso di destinazione completo. Per aggiornare questi parametri, eseguire i seguenti comandi:

```
*.audit_file_dest='/orabin/admin/TOAST/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/logs/TOAST/arch/control01.ctl','/logs/TOAST/redo/control
02.ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=/logs/TOAST/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

Disattivare il file init.ora originale

Questo file si trova in `$ORACLE_HOME/dbs` Ed è in genere in un pfile che funge da puntatore a spfile sul gruppo di dischi ASM. Per assicurarsi che spfile originale non sia più utilizzato, rinominarlo. Non eliminarlo, tuttavia, perché questo file è necessario se la migrazione deve essere interrotta.

```
[oracle@jfsc1 ~]$ cd $ORACLE_HOME/dbs
[oracle@jfsc1 dbs]$ cat initTOAST.ora
SPFILE='+ASM0/TOAST/spfileTOAST.ora'
[oracle@jfsc1 dbs]$ mv initTOAST.ora initTOAST.ora.prev
[oracle@jfsc1 dbs]$
```

Ricreazione del file dei parametri

Questa è la fase finale del trasferimento di spfile. Il file spfile originale non viene più utilizzato e il database viene avviato (ma non montato) utilizzando il file intermedio. Il contenuto di questo file può essere scritto nella nuova posizione spfile come segue:

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

Avviare il database per iniziare a utilizzare il nuovo spfile

È necessario avviare il database per rilasciare i blocchi sul file intermedio e avviare il database utilizzando solo il nuovo file spfile. L'avvio del database dimostra inoltre che la nuova posizione di spfile è corretta e che i suoi dati sono validi.

```
RMAN> shutdown immediate;
Oracle instance shut down
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

Ripristina controlfile

È stato creato un controlfile di backup nel percorso `/tmp/TOAST.ctrl` nelle fasi precedenti della procedura. Il nuovo spfile definisce le posizioni controlfile come /`logfs/TOAST/ctrl/ctrlfile1.ctrl` e. `/logfs/TOAST/redo/ctrlfile2.ctrl`. Tuttavia, tali file non esistono ancora.

1. Questo comando ripristina i dati controlfile nei percorsi definiti in spfile.

   ```
   RMAN> restore controlfile from '/tmp/TOAST.ctrl';
   Starting restore at 13-MAY-16
   using channel ORA_DISK_1
   channel ORA_DISK_1: copied control file copy
   output file name=/logs/TOAST/arch/control01.ctl
   output file name=/logs/TOAST/redo/control02.ctl
   Finished restore at 13-MAY-16
   ```

2. Eseguire il comando mount in modo che i file di controllo vengano rilevati correttamente e contengano dati validi.

   ```
   RMAN> alter database mount;
   Statement processed
   released channel: ORA_DISK_1
   ```

   Per convalidare `control_files` eseguire il seguente comando:

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
/logs/TOAST/arch/control01.ctl
                                                 ,
/logs/TOAST/redo/control02.c
                                                 tl
```

Riproduzione del registro

Il database sta attualmente utilizzando i file di dati nella vecchia posizione. Prima di poter utilizzare la copia, è necessario sincronizzare i file di dati. È trascorso del tempo durante il processo di copia iniziale e le modifiche sono state registrate principalmente nei registri di archivio. Queste modifiche vengono replicate nei due passaggi seguenti.

1. Eseguire un backup incrementale RMAN, che contiene i registri di archivio.

```
RMAN>  backup incremental level 1 format '/logs/TOAST/arch/%U' for
recover of copy with tag 'ONTAP_MIGRATION' database;
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=124 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/logs/TOAST/arch/09r5fj8i_1_1 tag=ONTAP_MIGRATION
comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

2. Riprodurre i registri.

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 13-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
recovering datafile copy file number=00002 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
recovering datafile copy file number=00003 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
recovering datafile copy file number=00004 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
channel ORA_DISK_1: reading from backup piece
/logs/TOAST/arch/09r5fj8i_1_1
channel ORA_DISK_1: piece handle=/logs/TOAST/arch/09r5fj8i_1_1
tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

Attivazione

Il controlfile ripristinato fa ancora riferimento ai file di dati nella posizione originale e contiene anche le informazioni di percorso per i file di dati copiati.

1. Per modificare i file di dati attivi, eseguire `switch database to copy` comando:

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSTEM_FNO-1_01r5fhjg"
datafile 2 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSAUX_FNO-2_02r5fhjo"
datafile 3 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt"
datafile 4 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-USERS_FNO-4_05r5fhk6"
```

2. Sebbene i file di dati debbano essere completamente coerenti, è necessario eseguire un passaggio finale per riprodurre le modifiche rimanenti registrate nei registri di ripristino online. Utilizzare `recover database` comando per riprodurre queste modifiche e rendere la copia identica al 100% all'originale. Tuttavia, la copia non è ancora aperta.

```
RMAN> recover database;
Starting recover at 13-MAY-16
using channel ORA_DISK_1
starting media recovery
archived log for thread 1 with sequence 28 is already on disk as file
+ASM0/TOAST/redo01.log
archived log file name=+ASM0/TOAST/redo01.log thread=1 sequence=28
media recovery complete, elapsed time: 00:00:00
Finished recover at 13-MAY-16
```

Spostare i file di dati temporanei

1. Identificare la posizione dei file di dati temporanei ancora in uso sul gruppo di dischi originale.

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
------------------------------------------------------------------------
--------
1 +ASM0/TOAST/temp01.dbf
```

2. Per spostare i file di dati, eseguire i seguenti comandi. Se ci sono molti tempfile, utilizzare un editor di testo per creare il comando RMAN e quindi tagliarlo e incollarlo.

```
RMAN> run {
2> set newname for tempfile 1 to '/oradata/TOAST/temp01.dbf';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/TOAST/temp01.dbf in control file
```

Migrazione dei log di ripristino

Il processo di migrazione è quasi completo, ma i log di ripristino si trovano ancora nel gruppo di dischi ASM originale. I log di ripristino non possono essere spostati direttamente. Al contrario, viene creata e aggiunta alla configurazione una nuova serie di log di ripristino, in seguito a una perdita dei vecchi log.

1. Identificare il numero di gruppi di log di ripristino e i rispettivi numeri di gruppo.

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
------------------------------------------------------------------------------
--------
1 +ASM0/TOAST/redo01.log
2 +ASM0/TOAST/redo02.log
3 +ASM0/TOAST/redo03.log
```

2. Immettere le dimensioni dei registri di ripristino.

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
------------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Per ogni log di ripristino, creare un nuovo gruppo utilizzando le stesse dimensioni del gruppo di log di ripristino corrente utilizzando la nuova posizione del file system.

```
RMAN> alter database add logfile '/logs/TOAST/redo/log00.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log01.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log02.rdo' size
52428800;
Statement processed
```

4. Rimuovere i vecchi gruppi di file di registro che si trovano ancora nell'archivio precedente.

```
RMAN> alter database drop logfile group 4;
Statement processed
RMAN> alter database drop logfile group 5;
Statement processed
RMAN> alter database drop logfile group 6;
Statement processed
```

5. Se si verifica un errore che blocca l'eliminazione di un registro attivo, forzare un passaggio al registro successivo per rilasciare il blocco e forzare un punto di verifica globale. Di seguito è riportato un esempio. Il tentativo di rilasciare il gruppo di file di registro 3, che si trovava nella vecchia posizione, è stato negato

perché in questo file di registro erano ancora presenti dati attivi. L'archiviazione dei log seguita da un punto di verifica consente l'eliminazione dei file di log.

```
RMAN> alter database drop logfile group 4;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 4 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 4 thread 1:
'+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 4;
Statement processed
```

6. Esaminare l'ambiente per assicurarsi che tutti i parametri basati sulla posizione siano aggiornati.

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

7. Nel seguente script viene illustrato come semplificare questo processo.

```
[root@jfsc1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
TOAST redo logs:
/logs/TOAST/redo/log00.rdo
/logs/TOAST/redo/log01.rdo
/logs/TOAST/redo/log02.rdo
TOAST temp datafiles:
/oradata/TOAST/temp01.dbf
TOAST spfile
spfile                                  string
/orabin/product/12.1.0/dbhome_
                                                        1/dbs/spfileTOAST.ora
TOAST key parameters
control_files /logs/TOAST/arch/control01.ctl,
/logs/TOAST/redo/control02.ctl
log_archive_dest_1 LOCATION=/logs/TOAST/arch
```

8. Se i gruppi di dischi ASM sono stati completamente evacuati, è possibile smontarli con `asmcmd`. In molti casi, i file appartenenti ad altri database o al file ASM spfile/passwd possono essere ancora presenti.

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

Procedura di pulizia del file di dati

Il processo di migrazione potrebbe generare file di dati con sintassi lunga o criptica, a seconda del modo in cui è stato utilizzato Oracle RMAN. Nell'esempio illustrato, il backup è stato eseguito con il formato file di `/oradata/TOAST/%U`. `%U` Indica che RMAN deve creare un nome univoco predefinito per ciascun file di dati. Il risultato è simile a quanto illustrato nel testo seguente. I nomi tradizionali dei file di dati sono incorporati nei nomi. Questo può essere ripulito utilizzando l'approccio basato su script illustrato nella "Pulitura della migrazione ASM".

```
[root@jfsc1 current]# ./fixuniquenames.pl TOAST
#sqlplus Commands
shutdown immediate;
startup mount;
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/system.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/sysaux.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt /oradata/TOAST/undotbs1.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
/oradata/TOAST/users.dbf
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSTEM_FNO-1_01r5fhjg' to '/oradata/TOAST/system.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSAUX_FNO-2_02r5fhjo' to '/oradata/TOAST/sysaux.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
UNDOTBS1_FNO-3_03r5fhjt' to '/oradata/TOAST/undotbs1.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
USERS_FNO-4_05r5fhk6' to '/oradata/TOAST/users.dbf';
alter database open;
```

Ribilanciamento di Oracle ASM

Come indicato in precedenza, è possibile eseguire la migrazione trasparente di un gruppo di dischi Oracle ASM in un nuovo sistema di storage utilizzando il processo di ribilanciamento. Riassumendo, il processo di ribilanciamento richiede l'aggiunta di LUN di dimensioni uguali al gruppo esistente di LUN, seguita da un'operazione di disgregazione del LUN precedente. Oracle ASM riposiziona automaticamente i dati sottostanti nel nuovo storage in un layout ottimale e, al termine, rilascia i vecchi LUN.

Il processo di migrazione utilizza un i/o sequenziale efficiente e non causa generalmente un'interruzione delle performance, ma la velocità di migrazione può essere rallentata quando necessario.

Identificazione dei dati da migrare

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
SQL> select group_number||' '||name from v$asm_diskgroup;
1 NEWDATA
```

Creazione di nuovi LUN

Creare nuovi LUN delle stesse dimensioni e impostare l'appartenenza a utenti e gruppi come richiesto. I LUN devono essere visualizzati come `CANDIDATE` dischi.

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
 0 0 /dev/mapper/3600a098038303537762b47594c31586b CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315869 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315858 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c31586a CANDIDATE
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
```

Aggiungere nuovi LUN

Anche se è possibile eseguire tutte le operazioni di aggiunta e rilascio, in genere è più semplice aggiungere nuovi LUN in due passaggi. Innanzitutto, aggiungere i nuovi LUN al gruppo di dischi. Questo passaggio comporta la migrazione di metà delle estensioni dai LUN ASM correnti ai nuovi LUN.

La potenza di riequilibrio indica la velocità di trasferimento dei dati. Più alto è il numero, più alto è il parallelismo del trasferimento dei dati. La migrazione viene eseguita con efficienti operazioni di i/o sequenziali che hanno scarse probabilità di causare problemi di performance. Tuttavia, se lo si desidera, il potere di riequilibrio di una migrazione in corso può essere regolato con `alter diskgroup [name] rebalance power [level]` comando. Le migrazioni tipiche utilizzano un valore di 5.

```
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586b' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315869' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315858' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586a' rebalance power 5;
Diskgroup altered.
```

Funzionamento del monitor

È possibile monitorare e gestire un'operazione di ribilanciamento in più modi. Per questo esempio è stato utilizzato il comando seguente.

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

Una volta completata la migrazione, non vengono segnalate operazioni di ribilanciamento.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

LUN meno recenti

La migrazione è ormai a metà strada. Potrebbe essere opportuno eseguire alcuni test delle prestazioni di base per assicurarsi che l'ambiente sia sano. Dopo la conferma, è possibile spostare i dati rimanenti eliminando i vecchi LUN. Tenere presente che ciò non determina il rilascio immediato dei LUN. L'operazione di rilascio indica ad Oracle ASM di riposizionare prima le estensioni e quindi rilasciare il LUN.

```
sqlplus / as sysasm
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0000 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0001 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0002 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0003 rebalance power 5;
Diskgroup altered.
```

Funzionamento del monitor

L'operazione di ribilanciamento può essere monitorata e gestita in più modi. Per questo esempio è stato utilizzato il seguente comando:

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

Una volta completata la migrazione, non vengono segnalate operazioni di ribilanciamento.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

Rimuovere i vecchi LUN

Prima di rimuovere i vecchi LUN dal gruppo di dischi, è necessario eseguire un controllo finale dello stato dell'intestazione. Dopo il rilascio di un LUN da ASM, non viene più elencato un nome e lo stato dell'intestazione viene elencato come `FORMER`. Questo indica che questi LUN possono essere rimossi in modo sicuro dal sistema.

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NAME||''||GROUP_NUMBER||''||TOTAL_MB||''||PATH||''||HEADER_STATUS
--------------------------------------------------------------------------
------
 0 0 /dev/mapper/3600a098038303537762b47594c315863 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315864 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315861 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315862 FORMER
NEWDATA_0005 1 10240 /dev/mapper/3600a098038303537762b47594c315869 MEMBER
NEWDATA_0007 1 10240 /dev/mapper/3600a098038303537762b47594c31586a MEMBER
NEWDATA_0004 1 10240 /dev/mapper/3600a098038303537762b47594c31586b MEMBER
NEWDATA_0006 1 10240 /dev/mapper/3600a098038303537762b47594c315858 MEMBER
8 rows selected.
```

Migrazione LVM

La procedura qui presentata mostra i principi di una migrazione basata su LVM di un gruppo di volumi chiamato `datavg`. Gli esempi sono tratti da Linux LVM, ma i principi si applicano ugualmente a AIX, HP-UX e VxVM. I comandi precisi possono variare.

1. Identificare i LUN attualmente presenti in `datavg` gruppo di volumi.

```
[root@host1 ~]# pvdisplay -C | grep datavg
  /dev/mapper/3600a098038303537762b47594c31582f datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31585a datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c315859 datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31586c datavg lvm2 a--  10.00g
10.00g
```

2. Creazione di nuovi LUN di dimensioni fisiche identiche o leggermente superiori e definizione di volumi fisici.

```
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315864
  Physical volume "/dev/mapper/3600a098038303537762b47594c315864"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315863
  Physical volume "/dev/mapper/3600a098038303537762b47594c315863"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315862
  Physical volume "/dev/mapper/3600a098038303537762b47594c315862"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315861
  Physical volume "/dev/mapper/3600a098038303537762b47594c315861"
successfully created
```

3. Aggiungere i nuovi volumi al gruppo di volumi.

```
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315864
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315863
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315862
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315861
  Volume group "datavg" successfully extended
```

4. Eseguire il `pvmove` Comando per spostare le estensioni di ogni LUN corrente nel nuovo LUN. Il `-i [seconds]` l'argomento controlla l'avanzamento dell'operazione.

```
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31582f
/dev/mapper/3600a098038303537762b47594c315864
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 14.2%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 28.4%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 42.5%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 57.1%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 72.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 87.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31585a
/dev/mapper/3600a098038303537762b47594c315863
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 14.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 29.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 44.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 60.1%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 75.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 90.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c315859
/dev/mapper/3600a098038303537762b47594c315862
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 14.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 29.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 45.5%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 61.1%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 76.6%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 91.7%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31586c
/dev/mapper/3600a098038303537762b47594c315861
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 15.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 30.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 46.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 61.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 77.2%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 92.3%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 100.0%
```

5. Una volta completato questo processo, rimuovere i LUN precedenti dal gruppo di volumi utilizzando `vgreduce` comando. Se l'operazione ha esito positivo, è ora possibile rimuovere il LUN dal sistema in modo sicuro.

```
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31582f
Removed "/dev/mapper/3600a098038303537762b47594c31582f" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31585a
  Removed "/dev/mapper/3600a098038303537762b47594c31585a" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c315859
  Removed "/dev/mapper/3600a098038303537762b47594c315859" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31586c
  Removed "/dev/mapper/3600a098038303537762b47594c31586c" from volume
group "datavg"
```

Importazione LUN esterne

Pianificazione

Le procedure per migrare LE risorse SAN utilizzando FLI sono documentate in NetApp "Documentazione per l'importazione dei LUN esteri di ONTAP" .

Dal punto di vista del database e dell'host, non sono necessarie operazioni speciali. Dopo l'aggiornamento delle zone FC e la disponibilità dei LUN su ONTAP, LVM dovrebbe essere in grado di leggere i metadati LVM dai LUN. Inoltre, i gruppi di volumi sono pronti per l'uso senza ulteriori passaggi di configurazione. Rari casi, gli ambienti potrebbero includere file di configurazione con hard-code e riferimenti allo storage array precedente. Ad esempio, un sistema Linux che includeva `/etc/multipath.conf` Le regole che fanno riferimento a un WWN di un dato dispositivo devono essere aggiornate per riflettere le modifiche introdotte da FLI.

> Fare riferimento alla matrice di compatibilità NetApp per informazioni sulle configurazioni supportate. Se il proprio ambiente non è incluso, contattare il rappresentante NetApp per assistenza.

Questo esempio mostra la migrazione di LUN ASM e LVM ospitati su un server Linux. FLI è supportato su altri sistemi operativi e, sebbene i comandi sul lato host possano differire, i principi sono gli stessi e le procedure ONTAP sono identiche.

Identificare i LUN LVM

La prima fase della preparazione consiste nell'identificare i LUN da migrare. Nell'esempio mostrato qui, due file system basati su SAN sono montati su `/orabin` e. `/backups`.

```
[root@host1 ~]# df -k
Filesystem                          1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root                52403200   8811464  43591736  17% /
devtmpfs                             65882776         0  65882776   0% /dev
...
fas8060-nfs-public:/install         199229440 119368128  79861312  60%
/install
/dev/mapper/sanvg-lvorabin           20961280  12348476   8612804  59%
/orabin
/dev/mapper/sanvg-lvbackups          73364480  62947536  10416944  86%
/backups
```

Il nome del gruppo di volumi può essere estratto dal nome del dispositivo, che utilizza il formato (nome del gruppo di volumi)-(nome del volume logico). In questo caso, viene chiamato il gruppo di volumi `sanvg`.

Il `pvdisplay` Il comando può essere utilizzato come segue per identificare i LUN che supportano questo gruppo di volumi. In questo caso, sono presenti 10 LUN che compongono il `sanvg` gruppo di volumi.

```
[root@host1 ~]# pvdisplay -C -o pv_name,pv_size,pv_fmt,vg_name
  PV                                             PSize   VG
  /dev/mapper/3600a0980383030445424487556574266  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574267  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574268  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574269  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426a  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426b  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426c  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426d  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426e  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426f  10.00g sanvg
  /dev/sda2                                     278.38g rhel
```

Identificare i LUN ASM

Anche i LUN ASM devono essere migrati. Per ottenere il numero di LUN e percorsi LUN da sqlplus come utente sysasm, eseguire il comando seguente:

```
SQL> select path||' '||os_mb from v$asm_disk;
PATH||''||OS_MB
--------------------------------------------------------------------------------
------
/dev/oracleasm/disks/ASM0 10240
/dev/oracleasm/disks/ASM9 10240
/dev/oracleasm/disks/ASM8 10240
/dev/oracleasm/disks/ASM7 10240
/dev/oracleasm/disks/ASM6 10240
/dev/oracleasm/disks/ASM5 10240
/dev/oracleasm/disks/ASM4 10240
/dev/oracleasm/disks/ASM1 10240
/dev/oracleasm/disks/ASM3 10240
/dev/oracleasm/disks/ASM2 10240
10 rows selected.
SQL>
```

Modifiche alla rete FC

L'ambiente corrente contiene 20 LUN da migrare. Aggiornare la SAN corrente in modo che ONTAP possa accedere ai LUN correnti. I dati non sono ancora stati migrati, ma ONTAP deve leggere le informazioni di configurazione dalle LUN correnti per creare la nuova home page per quei dati.

Almeno una porta HBA sul sistema AFF/FAS deve essere configurata come porta Initiator. Inoltre, le zone FC devono essere aggiornate in modo che ONTAP possa accedere alle LUN sullo storage array esterno. Alcuni storage array hanno configurato il masking dei LUN, che limita i WWN che possono accedere a una determinata LUN. In tal caso, è necessario aggiornare anche il masking dei LUN per garantire l'accesso ai WWN di ONTAP.

Al termine di questa operazione, ONTAP dovrebbe essere in grado di visualizzare l'array di archiviazione esterno con `storage array show` comando. Il campo chiave restituito è il prefisso utilizzato per identificare il LUN esterno sul sistema. Nell'esempio seguente, i LUN dell'array esterno `FOREIGN_1` Appare in ONTAP usando il prefisso di `FOR-1`.

Identificare un array esterno

```
Cluster01::> storage array show -fields name,prefix
name          prefix
------------- ------
FOREIGN_1     FOR-1
Cluster01::>
```

Identificare i LUN esterni

I LUN possono essere elencati passando l' `array-name` al `storage disk show` comando. I dati restituiti vengono referenziati più volte durante la procedura di migrazione.

```
Cluster01::> storage disk show -array-name FOREIGN_1 -fields disk,serial
disk     serial-number
-------- -------------
FOR-1.1  800DT$HuVWBX
FOR-1.2  800DT$HuVWBZ
FOR-1.3  800DT$HuVWBW
FOR-1.4  800DT$HuVWBY
FOR-1.5  800DT$HuVWB/
FOR-1.6  800DT$HuVWBa
FOR-1.7  800DT$HuVWBd
FOR-1.8  800DT$HuVWBb
FOR-1.9  800DT$HuVWBc
FOR-1.10 800DT$HuVWBe
FOR-1.11 800DT$HuVWBf
FOR-1.12 800DT$HuVWBg
FOR-1.13 800DT$HuVWBi
FOR-1.14 800DT$HuVWBh
FOR-1.15 800DT$HuVWBj
FOR-1.16 800DT$HuVWBk
FOR-1.17 800DT$HuVWBm
FOR-1.18 800DT$HuVWBl
FOR-1.19 800DT$HuVWBo
FOR-1.20 800DT$HuVWBn
20 entries were displayed.
Cluster01::>
```

Registrare LUN di array esterni come candidati di importazione

Le LUN esterne vengono inizialmente classificate come qualsiasi tipo di LUN specifico. Prima di poter importare i dati, i LUN devono essere contrassegnati come esterni e quindi come candidati al processo di importazione. Questo passaggio viene completato passando il numero di serie a. `storage disk modify`, come illustrato nell'esempio seguente. Si noti che questa procedura etichetta solo il LUN come estraneo all'interno di ONTAP. Nessun dato viene scritto nella LUN esterna stessa.

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign true
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign true
Cluster01::*>
```

Creazione di volumi per l'hosting di LUN migrati

Per ospitare le LUN migrate è necessario un volume. La configurazione esatta dei volumi dipende dal piano generale per sfruttare le funzionalità di ONTAP. In questo esempio, i LUN ASM vengono posizionati in un volume e i LUN LVM in un secondo volume. In questo modo, puoi gestire le LUN come gruppi indipendenti per scopi come il tiering, la creazione di snapshot o l'impostazione di controlli della qualità del servizio.

Impostare `snapshot-policy `to `none`. Il processo di migrazione può comportare un notevole ricambio dei dati. Pertanto, potrebbe verificarsi un notevole aumento del consumo di spazio se le istantanee vengono create accidentalmente perché i dati indesiderati vengono acquisiti nelle istantanee.

```
Cluster01::> volume create -volume new_asm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1152] Job succeeded: Successful
Cluster01::> volume create -volume new_lvm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1153] Job succeeded: Successful
Cluster01::>
```

Creare LUN ONTAP

Una volta creati i volumi, è necessario creare i nuovi LUN. In genere, la creazione di un LUN richiede all'utente di specificare tali informazioni come la dimensione LUN, ma in questo caso l'argomento del disco esterno viene passato al comando. Di conseguenza, ONTAP replica i dati di configurazione LUN correnti dal numero di serie specificato. Utilizza inoltre la geometria del LUN e i dati della tabella delle partizioni per regolare l'allineamento delle LUN e stabilire prestazioni ottimali.

In questo passaggio, i numeri di serie devono essere referenziati rispetto all'array esterno per assicurarsi che il LUN esterno corretto corrisponda al nuovo LUN corretto.

```
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN0 -ostype
linux -foreign-disk 800DT$HuVWBW
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN1 -ostype
linux -foreign-disk 800DT$HuVWBX
Created a LUN of size 10g (10737418240)
...
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN8 -ostype
linux -foreign-disk 800DT$HuVWBn
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN9 -ostype
linux -foreign-disk 800DT$HuVWBo
Created a LUN of size 10g (10737418240)
```

Creare relazioni di importazione

I LUN sono stati creati ma non sono configurati come destinazione di replica. Prima di eseguire questo passaggio, i LUN devono essere messi offline. Questo passaggio aggiuntivo è progettato per proteggere i dati dagli errori dell'utente. Se ONTAP consentisse di eseguire una migrazione su un LUN online, rischierebbe di provocare la sovrascrittura dei dati attivi con un errore tipografico. Questa fase aggiuntiva, che obbliga l'utente a portare un LUN offline, consente di verificare se viene utilizzato il LUN di destinazione corretto come destinazione della migrazione.

```
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN0
Warning: This command will take LUN "/vol/new_asm/LUN0" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN1
Warning: This command will take LUN "/vol/new_asm/LUN1" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
...
Warning: This command will take LUN "/vol/new_lvm/LUN8" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_lvm/LUN9
Warning: This command will take LUN "/vol/new_lvm/LUN9" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
```

Una volta che i LUN sono offline, è possibile stabilire la relazione di importazione passando il numero di serie del LUN esterno a. `lun import create` comando.

```
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN0
-foreign-disk 800DT$HuVWBW
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN1
-foreign-disk 800DT$HuVWBX
...
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN8
-foreign-disk 800DT$HuVWBn
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN9
-foreign-disk 800DT$HuVWBo
Cluster01::*>
```

Una volta stabilite tutte le relazioni di importazione, è possibile riportare online i LUN.

```
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN9
Cluster01::*>
```

Crea gruppo iniziatore

Un gruppo iniziatore (igroup) fa parte dell'architettura di mascheramento LUN di ONTAP. Un LUN appena creato non è accessibile a meno che non venga concesso per la prima volta l'accesso a un host. A tale scopo, creare un igroup in cui siano elencati i nomi WWN FC o iSCSI Initiator a cui è necessario concedere l'accesso. Al momento della scrittura del report, FLI era supportato solo per LUN FC. Tuttavia, la conversione in post-migrazione iSCSI è un'attività semplice, come illustrato nella "Conversione protocollo".

In questo esempio, viene creato un igroup che contiene due WWN corrispondenti alle due porte disponibili sull'HBA dell'host.

```
Cluster01::*> igroup create linuxhost -protocol fcp -ostype linux
-initiator 21:00:00:0e:1e:16:63:50 21:00:00:0e:1e:16:63:51
```

Mappare nuovi LUN all'host

Dopo la creazione di igroup, i LUN vengono quindi mappati all'igroup definito. Questi LUN sono disponibili solo per i WWN inclusi in questo igroup. In questa fase del processo di migrazione, NetApp presume che l'host non sia stato sottoposto a zoning in ONTAP. Questo è importante perché se l'host è contemporaneamente collegato all'array esterno e al nuovo sistema ONTAP, vi è il rischio che su ogni array possano essere rilevati LUN con lo stesso numero di serie. Questa situazione potrebbe causare malfunzionamenti del multipath o danni ai dati.

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
Cluster01::*>
```

Cutover

Una parte delle interruzioni durante l'importazione di LUN esterne è inevitabile a causa della necessità di modificare la configurazione di rete FC. Tuttavia, l'interruzione non deve durare più a lungo del tempo necessario per riavviare l'ambiente di database e

aggiornare lo zoning FC per passare dalla connettività FC dell'host al ONTAP.

Questo processo può essere riassunto come segue:

1. Quiete di tutta l'attività LUN sui LUN esterni.
2. Reindirizzare le connessioni FC dell'host al nuovo sistema ONTAP.
3. Attivare il processo di importazione.
4. Rilevare nuovamente i LUN.
5. Riavviare il database.

Non è necessario attendere il completamento del processo di migrazione. Non appena inizia la migrazione di un determinato LUN, questo è disponibile su ONTAP e può fornire dati durante il processo di copia dei dati. Tutte le letture vengono passate alla LUN esterna e tutte le scritture vengono scritte in modo sincrono su entrambi gli array. L'operazione di copia è molto veloce e l'overhead del reindirizzamento del traffico FC è minimo, per cui qualsiasi impatto sulle performance deve essere transitorio e minimo. In caso di problemi, è possibile ritardare il riavvio dell'ambiente fino al completamento del processo di migrazione e all'eliminazione delle relazioni di importazione.

Chiudere il database

Il primo passo per chiudere l'ambiente in questo esempio è arrestare il database.

```
[oracle@host1 bin]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base remains unchanged with value /orabin
[oracle@host1 bin]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, Automatic Storage Management, OLAP, Advanced
Analytics
and Real Application Testing options
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
SQL>
```

Chiudere i servizi di rete

Uno dei file system basati su SAN oggetto della migrazione include anche i servizi Oracle ASM. La disattivazione dei LUN sottostanti richiede lo smontaggio dei file system, il che a sua volta significa l'arresto di tutti i processi con file aperti su questo file system.

```
[oracle@host1 bin]$ ./crsctl stop has -f
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'host1'
CRS-2673: Attempting to stop 'ora.evmd' on 'host1'
CRS-2673: Attempting to stop 'ora.DATA.dg' on 'host1'
CRS-2673: Attempting to stop 'ora.LISTENER.lsnr' on 'host1'
CRS-2677: Stop of 'ora.DATA.dg' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.asm' on 'host1'
CRS-2677: Stop of 'ora.LISTENER.lsnr' on 'host1' succeeded
CRS-2677: Stop of 'ora.evmd' on 'host1' succeeded
CRS-2677: Stop of 'ora.asm' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.cssd' on 'host1'
CRS-2677: Stop of 'ora.cssd' on 'host1' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'host1' has completed
CRS-4133: Oracle High Availability Services has been stopped.
[oracle@host1 bin]$
```

Smontare i file system

Se tutti i processi vengono arrestati, l'operazione umount ha esito positivo. Se l'autorizzazione viene negata, è necessario che sul file system sia presente un processo con blocco. Il `fuser` command può aiutare a identificare questi processi.

```
[root@host1 ~]# umount /orabin
[root@host1 ~]# umount /backups
```

Disattivare i gruppi di volumi

Una volta smontati tutti i file system di un dato gruppo di volumi, è possibile disattivare il gruppo di volumi.

```
[root@host1 ~]# vgchange --activate n sanvg
  0 logical volume(s) in volume group "sanvg" now active
[root@host1 ~]#
```

Modifiche alla rete FC

È ora possibile aggiornare le zone FC per rimuovere tutti gli accessi dall'host all'array esterno e stabilire l'accesso a ONTAP.

Avviare il processo di importazione

Per avviare i processi di importazione LUN, eseguire `lun import start` comando.

```
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN0
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN1
...
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN8
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN9
Cluster01::lun import*>
```

Monitorare l'avanzamento dell'importazione

L'operazione di importazione può essere monitorata con `lun import show` comando. Come illustrato di seguito, è in corso l'importazione di tutte le LUN da 20 GB, il che significa che i dati sono ora accessibili tramite ONTAP, anche se l'operazione di copia dei dati continua a proseguire.

```
Cluster01::lun import*> lun import show -fields path,percent-complete
vserver   foreign-disk path              percent-complete
--------- ------------ ----------------- ----------------
vserver1  800DT$HuVWB/ /vol/new_asm/LUN4 5
vserver1  800DT$HuVWBW /vol/new_asm/LUN0 5
vserver1  800DT$HuVWBX /vol/new_asm/LUN1 6
vserver1  800DT$HuVWBY /vol/new_asm/LUN2 6
vserver1  800DT$HuVWBZ /vol/new_asm/LUN3 5
vserver1  800DT$HuVWBa /vol/new_asm/LUN5 4
vserver1  800DT$HuVWBb /vol/new_asm/LUN6 4
vserver1  800DT$HuVWBc /vol/new_asm/LUN7 4
vserver1  800DT$HuVWBd /vol/new_asm/LUN8 4
vserver1  800DT$HuVWBe /vol/new_asm/LUN9 4
vserver1  800DT$HuVWBf /vol/new_lvm/LUN0 5
vserver1  800DT$HuVWBg /vol/new_lvm/LUN1 4
vserver1  800DT$HuVWBh /vol/new_lvm/LUN2 4
vserver1  800DT$HuVWBi /vol/new_lvm/LUN3 3
vserver1  800DT$HuVWBj /vol/new_lvm/LUN4 3
vserver1  800DT$HuVWBk /vol/new_lvm/LUN5 3
vserver1  800DT$HuVWBl /vol/new_lvm/LUN6 4
vserver1  800DT$HuVWBm /vol/new_lvm/LUN7 3
vserver1  800DT$HuVWBn /vol/new_lvm/LUN8 2
vserver1  800DT$HuVWBo /vol/new_lvm/LUN9 2
20 entries were displayed.
```

Se è necessario un processo offline, ritardare il riscoperta o il riavvio dei servizi finché il comando non `lun import show` indica che la migrazione è stata completata correttamente. È quindi possibile completare il processo di migrazione come descritto in "Importazione di LUN esterne - completamento".

Se hai bisogno di una migrazione online, procedi con il rilevamento dei LUN nella nuova sede e attiva i servizi.

Eseguire la scansione delle modifiche al dispositivo SCSI

Nella maggior parte dei casi, l'opzione più semplice per ritrovare nuove LUN è riavviare l'host. In questo modo, si rimuovono automaticamente i vecchi dispositivi obsoleti, si rilevano correttamente tutti i nuovi LUN e si creano dispositivi associati come i dispositivi multipathing. L'esempio qui mostra una procedura completamente online a scopo dimostrativo.

Attenzione: Prima di riavviare un host, assicurarsi che tutte le voci in `/etc/fstab` Il riferimento alle risorse SAN migrate verrà commentato. Se questa operazione non viene eseguita e si verificano problemi con l'accesso LUN, il sistema operativo potrebbe non avviarsi. Questa situazione non danneggia i dati. Tuttavia, può essere molto scomodo avviare in modalità rescue o in una modalità simile e correggere `/etc/fstab` In modo che il sistema operativo possa essere avviato per consentire la risoluzione dei problemi.

I LUN della versione di Linux utilizzata in questo esempio possono essere rianalizzati con `rescan-scsi-bus.sh` comando. Se il comando viene eseguito correttamente, nell'output viene visualizzato ogni percorso LUN. L'output può essere difficile da interpretare, ma, se la configurazione di zoning e igroup era corretta, molti LUN dovrebbero apparire che includono un `NETAPP` stringa fornitore.

```
[root@host1 /]# rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 0 2 0 0 ...
OLD: Host: scsi0 Channel: 02 Id: 00 Lun: 00
      Vendor: LSI      Model: RAID SAS 6G 0/1  Rev: 2.13
      Type:   Direct-Access                    ANSI SCSI revision: 05
Scanning host 1 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 1 0 0 0 ...
OLD: Host: scsi1 Channel: 00 Id: 00 Lun: 00
      Vendor: Optiarc  Model: DVD RW AD-7760H  Rev: 1.41
      Type:   CD-ROM                           ANSI SCSI revision: 05
Scanning host 2 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 3 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 4 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 5 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 6 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 7 for  all SCSI target IDs, all LUNs
 Scanning for device 7 0 0 10 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 10
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 11 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 11
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 12 ...
...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 18
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 9 0 1 19 ...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 19
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
0 new or changed device(s) found.
0 remapped or resized device(s) found.
0 device(s) removed.
```

Verificare la presenza di dispositivi multipercorso

Il processo di rilevamento LUN attiva anche la ricreazione dei dispositivi multipath, ma è noto che il driver multipathing Linux presenta problemi occasionali. L'output di `multipath -ll` dovrebbe essere controllato per verificare che l'output sia come previsto. Per esempio, l'uscita seguente mostra dispositivi multipercorso associati a A. `NETAPP` stringa fornitore. Ciascun dispositivo dispone di quattro percorsi, di cui due con priorità 50 e due con priorità 10. Anche se l'output esatto può variare con diverse versioni di Linux, questo risultato

sembra come previsto.

Fare riferimento alla documentazione delle utilità host per la versione di Linux utilizzata per verificare che `/etc/multipath.conf` le impostazioni sono corrette.

```
[root@host1 /]# multipath -ll
3600a098038303558735d493762504b36 dm-5 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:4  sdat 66:208 active ready running
| `- 9:0:1:4  sdbn 68:16  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:4  sdf  8:80   active ready running
  `- 9:0:0:4  sdz  65:144 active ready running
3600a098038303558735d493762504b2d dm-10 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:8  sdax 67:16  active ready running
| `- 9:0:1:8  sdbr 68:80  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:8  sdj  8:144  active ready running
  `- 9:0:0:8  sdad 65:208 active ready running
...
3600a098038303558735d493762504b37 dm-8 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:5  sdau 66:224 active ready running
| `- 9:0:1:5  sdbo 68:32  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:5  sdg  8:96   active ready running
  `- 9:0:0:5  sdaa 65:160 active ready running
3600a098038303558735d493762504b4b dm-22 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:19 sdbi 67:192 active ready running
| `- 9:0:1:19 sdcc 69:0   active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:19 sdu  65:64  active ready running
  `- 9:0:0:19 sdao 66:128 active ready running
```

Riattivare il gruppo di volumi LVM

Se i LUN LVM sono stati rilevati correttamente, l' `vgchange --activate y` il comando dovrebbe riuscire. Questo è un buon esempio del valore di un volume manager logico. Una modifica del WWN di una LUN o anche di un numero di serie non è importante perché i metadati del gruppo di volumi vengono scritti sul LUN stesso.

Il sistema operativo ha eseguito la scansione dei LUN e ha rilevato una piccola quantità di dati scritti sul LUN che lo identifica come volume fisico appartenente a. `sanvg volumegroup`. Successivamente, ha costruito tutti i dispositivi necessari. È sufficiente riattivare il gruppo di volumi.

```
[root@host1 /]# vgchange --activate y sanvg
  Found duplicate PV fpCzdLTuKfy2xDZjai1NliJh3TjLUBiT: using
/dev/mapper/3600a098038303558735d493762504b46 not /dev/sdp
  Using duplicate PV /dev/mapper/3600a098038303558735d493762504b46 from
subsystem DM, ignoring /dev/sdp
  2 logical volume(s) in volume group "sanvg" now active
```

Rimontare i file system

Dopo la riattivazione del gruppo di volumi, i file system possono essere montati con tutti i dati originali intatti. Come indicato in precedenza, i file system sono completamente operativi anche se la replica dei dati è ancora attiva nel gruppo back.

```
[root@host1 /]# mount /orabin
[root@host1 /]# mount /backups
[root@host1 /]# df -k
Filesystem                          1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root                52403200   8837100  43566100  17% /
devtmpfs                             65882776         0  65882776   0% /dev
tmpfs                                 6291456        84   6291372   1%
/dev/shm
tmpfs                                65898668      9884  65888784   1% /run
tmpfs                                65898668         0  65898668   0%
/sys/fs/cgroup
/dev/sda1                              505580    224828    280752  45% /boot
fas8060-nfs-public:/install          199229440 119368256  79861184  60%
/install
fas8040-nfs-routable:/snapomatic      9961472     30528   9930944   1%
/snapomatic
tmpfs                                13179736        16  13179720   1%
/run/user/42
tmpfs                                13179736         0  13179736   0%
/run/user/0
/dev/mapper/sanvg-lvorabin           20961280  12357456   8603824  59%
/orabin
/dev/mapper/sanvg-lvbackups          73364480  62947536  10416944  86%
/backups
```

Ripetere la scansione per i dispositivi ASM

I dispositivi ASMlib dovrebbero essere stati riselezionati al momento della nuova scansione dei dispositivi SCSI. La riscoperta può essere verificata online riavviando ASMlib e quindi eseguendo la scansione dei dischi.

Questa fase è pertinente solo alle configurazioni ASM in cui viene utilizzato ASMlib.

Attenzione: Se non viene utilizzato ASMlib, il `/dev/mapper` i dispositivi dovrebbero essere stati ricreati automaticamente. Tuttavia, le autorizzazioni potrebbero non essere corrette. È necessario impostare autorizzazioni speciali sui dispositivi sottostanti per ASM in assenza di ASMlib. Questa operazione viene solitamente eseguita tramite voci speciali in entrambi `/etc/multipath.conf` oppure `udev` o eventualmente in entrambi i set di regole. È possibile che questi file debbano essere aggiornati per riflettere le modifiche apportate all'ambiente in termini di numeri WWN o di serie per assicurarsi che i dispositivi ASM dispongano ancora delle autorizzazioni corrette.

In questo esempio, il riavvio di ASMlib e la scansione dei dischi mostrano gli stessi 10 LUN ASM dell'ambiente originale.

```
[root@host1 /]# oracleasm exit
Unmounting ASMlib driver filesystem: /dev/oracleasm
Unloading module "oracleasm": oracleasm
[root@host1 /]# oracleasm init
Loading module "oracleasm": oracleasm
Configuring "oracleasm" to use device physical block size
Mounting ASMlib driver filesystem: /dev/oracleasm
[root@host1 /]# oracleasm scandisks
Reloading disk partitions: done
Cleaning any stale ASM disks...
Scanning system for ASM disks...
Instantiating disk "ASM0"
Instantiating disk "ASM1"
Instantiating disk "ASM2"
Instantiating disk "ASM3"
Instantiating disk "ASM4"
Instantiating disk "ASM5"
Instantiating disk "ASM6"
Instantiating disk "ASM7"
Instantiating disk "ASM8"
Instantiating disk "ASM9"
```

Riavviare i servizi di rete

Ora che i dispositivi LVM e ASM sono online e disponibili, è possibile riavviare i servizi grid.

```
[root@host1 /]# cd /orabin/product/12.1.0/grid/bin
[root@host1 bin]# ./crsctl start has
```

Riavviare il database

Dopo aver riavviato i servizi di griglia, è possibile avviare il database. Potrebbe essere necessario attendere alcuni minuti affinché i servizi ASM diventino completamente disponibili prima di provare ad avviare il database.

```
[root@host1 bin]# su - oracle
[oracle@host1 ~]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base has been set to /orabin
[oracle@host1 ~]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> startup
ORACLE instance started.
Total System Global Area 3221225472 bytes
Fixed Size                  4502416 bytes
Variable Size            1207962736 bytes
Database Buffers         1996488704 bytes
Redo Buffers               12271616 bytes
Database mounted.
Database opened.
SQL>
```

Completamento

Dal punto di vista dell'host, la migrazione è completa, ma l'i/o viene ancora servito dall'array esterno fino a quando le relazioni di importazione non vengono eliminate.

Prima di eliminare le relazioni, è necessario confermare che il processo di migrazione è completo per tutte le LUN.

```
Cluster01::*> lun import show -vserver vserver1 -fields foreign-
disk,path,operational-state
vserver   foreign-disk path              operational-state
--------- ------------ ----------------- -----------------
vserver1 800DT$HuVWB/ /vol/new_asm/LUN4 completed
vserver1 800DT$HuVWBW /vol/new_asm/LUN0 completed
vserver1 800DT$HuVWBX /vol/new_asm/LUN1 completed
vserver1 800DT$HuVWBY /vol/new_asm/LUN2 completed
vserver1 800DT$HuVWBZ /vol/new_asm/LUN3 completed
vserver1 800DT$HuVWBa /vol/new_asm/LUN5 completed
vserver1 800DT$HuVWBb /vol/new_asm/LUN6 completed
vserver1 800DT$HuVWBc /vol/new_asm/LUN7 completed
vserver1 800DT$HuVWBd /vol/new_asm/LUN8 completed
vserver1 800DT$HuVWBe /vol/new_asm/LUN9 completed
vserver1 800DT$HuVWBf /vol/new_lvm/LUN0 completed
vserver1 800DT$HuVWBg /vol/new_lvm/LUN1 completed
vserver1 800DT$HuVWBh /vol/new_lvm/LUN2 completed
vserver1 800DT$HuVWBi /vol/new_lvm/LUN3 completed
vserver1 800DT$HuVWBj /vol/new_lvm/LUN4 completed
vserver1 800DT$HuVWBk /vol/new_lvm/LUN5 completed
vserver1 800DT$HuVWBl /vol/new_lvm/LUN6 completed
vserver1 800DT$HuVWBm /vol/new_lvm/LUN7 completed
vserver1 800DT$HuVWBn /vol/new_lvm/LUN8 completed
vserver1 800DT$HuVWBo /vol/new_lvm/LUN9 completed
20 entries were displayed.
```

Elimina relazioni di importazione

Al termine del processo di migrazione, eliminare la relazione di migrazione. Dopo aver fatto ciò, l'i/o viene servito esclusivamente dalle unità su ONTAP.

```
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN9
```

Annullare la registrazione di LUN esterne

Infine, modificare il disco per rimuovere `is-foreign` designazione.

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign false
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign false
Cluster01::*>
```

Conversione del protocollo

La modifica del protocollo utilizzato per accedere a un LUN è un requisito comune.

In alcuni casi, fa parte di una strategia globale di migrazione dei dati nel cloud. TCP/IP è il protocollo del cloud e il passaggio da FC a iSCSI facilita la migrazione in vari ambienti cloud. In altri casi, iSCSI potrebbe essere desiderabile per sfruttare i costi ridotti di un IP SAN. A volte, una migrazione potrebbe utilizzare un protocollo diverso come misura temporanea. Ad esempio, se un array esterno e LUN basati su ONTAP non possono coesistere sugli stessi HBA, è possibile utilizzare LUN iSCSI abbastanza a lungo da copiare i dati dal vecchio array. Dopo la rimozione dei vecchi LUN dal sistema, è possibile riconvertirli in FC.

La seguente procedura illustra la conversione da FC a iSCSI, ma i principi generali si applicano a una conversione da iSCSI a FC inversa.

Installare iSCSI Initiator

La maggior parte dei sistemi operativi include un iniziatore iSCSI software per impostazione predefinita, ma se non è incluso, può essere facilmente installato.

```
[root@host1 /]# yum install -y iscsi-initiator-utils
Loaded plugins: langpacks, product-id, search-disabled-repos,
subscription-
              : manager
Resolving Dependencies
--> Running transaction check
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.el7 will be
updated
--> Processing Dependency: iscsi-initiator-utils = 6.2.0.873-32.el7 for
package: iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7 will be
an update
--> Running transaction check
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.el7 will
be updated
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
will be an update
```

```
--> Finished Dependency Resolution
Dependencies Resolved
================================================================================
===
 Package                         Arch   Version              Repository
Size
================================================================================
===
Updating:
 iscsi-initiator-utils           x86_64 6.2.0.873-32.0.2.el7 ol7_latest 416
k
Updating for dependencies:
 iscsi-initiator-utils-iscsiuio x86_64 6.2.0.873-32.0.2.el7 ol7_latest  84
k
Transaction Summary
================================================================================
===
Upgrade  1 Package (+1 Dependent package)
Total download size: 501 k
Downloading packages:
No Presto metadata available for ol7_latest
(1/2): iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_6 | 416 kB   00:00
(2/2): iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2. |  84 kB   00:00
--------------------------------------------------------------------------------
---
Total                                           2.8 MB/s | 501 kB
00:00Cluster01
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Updating   : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
1/4
  Updating   : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
2/4
  Cleanup    : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
3/4
  Cleanup    : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
rhel-7-server-eus-rpms/7Server/x86_64/productid       | 1.7 kB     00:00
rhel-7-server-rpms/7Server/x86_64/productid           | 1.7 kB     00:00
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
1/4
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
2/4
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
```

```
3/4
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
Updated:
  iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7
Dependency Updated:
  iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
Complete!
[root@host1 /]#
```

Identificare il nome dell'iniziatore iSCSI

Durante il processo di installazione viene generato un nome iSCSI initiator univoco. Su Linux, si trova in `/etc/iscsi/initiatorname.iscsi` file. Questo nome viene utilizzato per identificare l'host sulla SAN IP.

```
[root@host1 /]# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1992-05.com.redhat:497bd66ca0
```

Creare un nuovo gruppo iniziatore

Un gruppo iniziatore (igroup) fa parte dell'architettura di mascheramento LUN di ONTAP. Un LUN appena creato non è accessibile a meno che non venga concesso per la prima volta l'accesso a un host. Questa operazione viene eseguita creando un igroup che elenca i nomi WWN FC o iniziatori iSCSI che richiedono l'accesso.

In questo esempio, viene creato un igroup che contiene l'iniziatore iSCSI dell'host Linux.

```
Cluster01::*> igroup create -igroup linuxiscsi -protocol iscsi -ostype
linux -initiator iqn.1994-05.com.redhat:497bd66ca0
```

Chiudere l'ambiente

Prima di modificare il protocollo LUN, è necessario disattivare completamente i LUN. Tutti i database di uno dei LUN da convertire devono essere chiusi, i file system devono essere dismontati e i gruppi di volumi devono essere disattivati. Se si utilizza ASM, assicurarsi che il gruppo di dischi ASM sia smontato e chiudere tutti i servizi della griglia.

Rimuovere la mappatura dei LUN dalla rete FC

Una volta terminate completamente le LUN, rimuovere le mappature dall'igroup FC originale.

```
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
```

Eseguire nuovamente il mapping dei LUN alla rete IP

Concedere l'accesso a ogni LUN al nuovo gruppo di iniziatori basati su iSCSI.

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxiscsi
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxiscsi
Cluster01::*>
```

Rilevamento delle destinazioni iSCSI

Il rilevamento iSCSI richiede due fasi. La prima è scoprire le destinazioni, che non è la stessa cosa per scoprire un LUN. Il `iscsiadm` il comando mostrato di seguito verifica il gruppo di portali specificato da `-p argument` Memorizza un elenco di tutti gli indirizzi IP e le porte che offrono servizi iSCSI. In questo caso, vi sono quattro indirizzi IP con servizi iSCSI sulla porta predefinita 3260.

> Il completamento di questo comando può richiedere alcuni minuti se non è possibile raggiungere uno qualsiasi degli indirizzi IP di destinazione.

```
[root@host1 ~]# iscsiadm -m discovery -t st -p fas8060-iscsi-public1
10.63.147.197:3260,1033 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
10.63.147.198:3260,1034 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.203:3260,1030 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.202:3260,1029 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
```

Rilevamento delle LUN iSCSI

Dopo aver rilevato le destinazioni iSCSI, riavviare il servizio iSCSI per rilevare i LUN iSCSI disponibili e creare i dispositivi associati, ad esempio i dispositivi multipath o ASMlib.

```
[root@host1 ~]# service iscsi restart
Redirecting to /bin/systemctl restart  iscsi.service
```

Riavviare l'ambiente

Riavviare l'ambiente riattivando i gruppi di volumi, rimontando i file system, riavviando i servizi RAC e così via. Per precauzione, NetApp consiglia di riavviare il server al termine del processo di conversione, per assicurarsi che tutti i file di configurazione siano corretti e che tutti i dispositivi obsoleti vengano rimossi.

Attenzione: Prima di riavviare un host, assicurarsi che tutte le voci in `/etc/fstab` Il riferimento alle risorse SAN migrate verrà commentato. Se questa operazione non viene eseguita e si verificano problemi con l'accesso LUN, il risultato può essere un sistema operativo che non si avvia. Questo problema non danneggia i dati. Tuttavia, può essere molto scomodo avviare in modalità rescue o una modalità simile e corretta `/etc/fstab` In modo che il sistema operativo possa essere avviato per consentire l'avvio delle operazioni di risoluzione dei problemi.

Script di esempio

Gli script presentati sono forniti come esempi di come eseguire lo script di varie attività del sistema operativo e del database. Vengono forniti così come sono. Se è necessario supporto per una procedura particolare, contattare NetApp o un rivenditore NetApp.

Arresto del database

Lo script Perl seguente prende un singolo argomento del SID Oracle e chiude un database. Può essere eseguito come utente Oracle o come root.

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
77 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF4
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
`;};
print @out;
if ("@out" =~ /ORACLE instance shut down/) {
print "$oraclesid shut down\n";
exit 0;}
elsif ("@out" =~ /Connected to an idle instance/) {
print "$oraclesid already shut down\n";
exit 0;}
else {
print "$oraclesid failed to shut down\n";
exit 1;}
```

Avvio del database

Lo script Perl seguente prende un singolo argomento del SID Oracle e chiude un database. Può essere eseguito come utente Oracle o come root.

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
'
`;}
else {
@out=`. oraenv << EOF3
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
`;};
print @out;
if ("@out" =~ /Database opened/) {
print "$oraclesid started\n";
exit 0;}
elsif ("@out" =~ /cannot start already-running ORACLE/) {
print "$oraclesid already started\n";
exit 1;}
else {
78 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
print "$oraclesid failed to start\n";
exit 1;}
```

Convertire il file system in sola lettura

Lo script seguente prende un argomento del file system e tenta di smontarlo e rimontarlo in modalità di sola lettura. Questa operazione è utile durante i processi di migrazione in cui un file system deve essere mantenuto disponibile per replicare i dati e deve essere protetto contro danni accidentali.

```
#! /usr/bin/perl
use strict;
#use warnings;
my $filesystem=$ARGV[0];
my @out=`umount '$filesystem'`;
if ($? == 0) {
 print "$filesystem unmounted\n";
 @out = `mount -o ro '$filesystem'`;
 if ($? == 0) {
 print "$filesystem mounted read-only\n";
 exit 0;}}
else {
 print "Unable to unmount $filesystem\n";
 exit 1;}
print @out;
```

Sostituire il file system

L'esempio di script riportato di seguito viene utilizzato per sostituire un file system con un altro. Poiché modifica il file ``/etc/fstab, deve essere eseguito come root. Accetta un singolo argomento delimitato da virgole per i file system vecchi e nuovi.

1. Per sostituire il file system, eseguire lo script seguente:

```
#! /usr/bin/perl
use strict;
#use warnings;
my $oldfs;
my $newfs;
my @oldfstab;
my @newfstab;
my $source;
my $mountpoint;
my $leftover;
my $oldfstabentry='';
my $newfstabentry='';
my $migratedfstabentry='';
($oldfs, $newfs) = split (',',$ARGV[0]);
open(my $filehandle, '<', '/etc/fstab') or die "Could not open
/etc/fstab\n";
while (my $line = <$filehandle>) {
 chomp $line;
 ($source, $mountpoint, $leftover) = split(/[ , ]/,$line, 3);
 if ($mountpoint eq $oldfs) {
 $oldfstabentry = "#Removed by swap script $source $oldfs $leftover";}
```

```
 elsif ($mountpoint eq $newfs) {
 $newfstabentry = "#Removed by swap script $source $newfs $leftover";
 $migratedfstabentry = "$source $oldfs $leftover";}
 else {
 push (@newfstab, "$line\n")}}
79 Migration of Oracle Databases to NetApp Storage Systems © 2021
NetApp, Inc. All rights reserved
push (@newfstab, "$oldfstabentry\n");
push (@newfstab, "$newfstabentry\n");
push (@newfstab, "$migratedfstabentry\n");
close($filehandle);
if ($oldfstabentry eq ''){
 die "Could not find $oldfs in /etc/fstab\n";}
if ($newfstabentry eq ''){
 die "Could not find $newfs in /etc/fstab\n";}
my @out=`umount '$newfs'`;
if ($? == 0) {
 print "$newfs unmounted\n";}
else {
 print "Unable to unmount $newfs\n";
 exit 1;}
@out=`umount '$oldfs'`;
if ($? == 0) {
 print "$oldfs unmounted\n";}
else {
 print "Unable to unmount $oldfs\n";
 exit 1;}
system("cp /etc/fstab /etc/fstab.bak");
open ($filehandle, ">", '/etc/fstab') or die "Could not open /etc/fstab
for writing\n";
for my $line (@newfstab) {
 print $filehandle $line;}
close($filehandle);
@out=`mount '$oldfs'`;
if ($? == 0) {
 print "Mounted updated $oldfs\n";
 exit 0;}
else{
 print "Unable to mount updated $oldfs\n";
 exit 1;}
exit 0;
```

Come esempio di utilizzo di questo script, si supponga che i dati in `/oradata` viene migrato in `/neworadata` e. `/logs` viene migrato in `/newlogs`. Uno dei metodi più semplici per eseguire questa attività consiste nell'utilizzare una semplice operazione di copia dei file per riportare la nuova periferica al punto di montaggio originale.

2. Si supponga che i file system vecchi e nuovi siano presenti in `/etc/fstab` archiviare come segue:

```
cluster01:/vol_oradata /oradata nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
cluster01:/vol_logs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /newlogs nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
```

3. Quando viene eseguito, questo script smonta il file system corrente e lo sostituisce con il nuovo:

```
[root@jfsc3 scripts]# ./swap.fs.pl /oradata,/neworadata
/neworadata unmounted
/oradata unmounted
Mounted updated /oradata
[root@jfsc3 scripts]# ./swap.fs.pl /logs,/newlogs
/newlogs unmounted
/logs unmounted
Mounted updated /logs
```

4. Lo script aggiorna anche `/etc/fstab` file di conseguenza. Nell'esempio illustrato, sono incluse le seguenti modifiche:

```
#Removed by swap script cluster01:/vol_oradata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_logs /logs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_newlogs /newlogs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0
0
```

Migrazione automatizzata del database

In questo esempio viene illustrato l'utilizzo di script di arresto, avvio e sostituzione del file system per automatizzare completamente la migrazione.

```
#! /usr/bin/perl
```

```
use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my @oldfs;
my @newfs;
my $x=1;
while ($x < scalar(@ARGV)) {
   ($oldfs[$x-1], $newfs[$x-1]) = split (',',$ARGV[$x]);
   $x+=1;}
my @out=`./dbshut.pl '$oraclesid'`;
print @out;
if ($? ne 0) {
   print "Failed to shut down database\n";
   exit 0;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`./mk.fs.readonly.pl '$oldfs[$x]'`;
   if ($? ne 0) {
      print "Failed to make filesystem $oldfs[$x] readonly\n";
      exit 0;}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`rsync -rlpogt --stats --progress --exclude='.snapshot'
'$oldfs[$x]/' '/$newfs[$x]/'`;
   print @out;
   if ($? ne 0) {
      print "Failed to copy filesystem $oldfs[$x] to $newfs[$x]\n";
      exit 0;}
   else {
      print "Succesfully replicated filesystem $oldfs[$x] to
$newfs[$x]\n";}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   print "swap $x $oldfs[$x] $newfs[$x]\n";
   my @out=`./swap.fs.pl '$oldfs[$x],$newfs[$x]'`;
   print @out;
   if ($? ne 0) {
      print "Failed to swap filesystem $oldfs[$x] for $newfs[$x]\n";
      exit 1;}
   else {
      print "Swapped filesystem $oldfs[$x] for $newfs[$x]\n";}
   $x+=1;}
my @out=`./dbstart.pl '$oraclesid'`;
print @out;
```

Visualizzare le posizioni dei file

Questo script raccoglie una serie di parametri critici del database e li stampa in un formato di facile lettura. Questo script può essere utile quando si esaminano i layout dei dati. Inoltre, lo script può essere modificato per altri usi.

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;};
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print "$oraclesid datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select member from v\\\\\$logfile;');
print "$oraclesid redo logs:\n";
for $line (@out) {
```

```
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name from v\\\\\$tempfile;');
print "$oraclesid temp datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('show parameter spfile;');
print "$oraclesid spfile\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name||\' \'||value from v\\\\\$parameter where
isdefault=\'FALSE\';');
print "$oraclesid key parameters\n";
for $line (@out) {
        chomp($line);
        if ($line =~ /control_files/) {print "$line\n";}
        if ($line =~ /db_create/) {print "$line\n";}
        if ($line =~ /db_file_name_convert/) {print "$line\n";}
        if ($line =~ /log_archive_dest/) {print "$line\n";}}
        if ($line =~ /log_file_name_convert/) {print "$line\n";}
        if ($line =~ /pdb_file_name_convert/) {print "$line\n";}
        if ($line =~ /spfile/) {print "$line\n";}
print "\n";
```

Pulitura della migrazione ASM

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
```

```
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;}
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print @out;
print "shutdown immediate;\n";
print "startup mount;\n";
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "host mv $line $1$newname\n";}}
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "alter database rename file '$line' to
'$1$newname';\n";}}
print "alter database open;\n";
print "\n";
```

Conversione del nome da ASM a file system

```
set serveroutput on;
set wrap off;
declare
    cursor df is select file#, name from v$datafile;
    cursor tf is select file#, name from v$tempfile;
    cursor lf is select member from v$logfile;
    firstline boolean := true;
begin
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('Parameters for log file conversion:');
    dbms_output.put_line(CHR(13));
    dbms_output.put('*.log_file_name_convert = ');
    for lfrec in lf loop
        if (firstline = true) then
            dbms_output.put('''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*/','') || '''');
        else
            dbms_output.put(',''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*/','') || '''');
        end if;
        firstline:=false;
    end loop;
    dbms_output.put_line(CHR(13));
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('rman duplication script:');
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('run');
    dbms_output.put_line('{');
    for dfrec in df loop
        dbms_output.put_line('set newname for datafile ' ||
            dfrec.file# || ' to ''' || dfrec.name ||''';');
    end loop;
    for tfrec in tf loop
        dbms_output.put_line('set newname for tempfile ' ||
            tfrec.file# || ' to ''' || tfrec.name ||''';');
    end loop;
    dbms_output.put_line('duplicate target database for standby backup
location INSERT_PATH_HERE;');
    dbms_output.put_line('}');
end;
/
```

Riprodurre i log sul database

Questo script accetta un singolo argomento di un SID Oracle per un database in modalità mount e tenta di riprodurre tutti i log di archivio attualmente disponibili.

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
84 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
`;
}
print @out;
```

Riprodurre i registri sul database di standby

Questo script è identico allo script precedente, tranne che è progettato per un database di standby.

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
`;
}
print @out;
```

Note aggiuntive

Ottimizzazione delle prestazioni e benchmarking

Il test accurato delle performance dello storage del database è un argomento estremamente complicato. Richiede la comprensione dei seguenti problemi:

- IOPS e throughput
- La differenza tra le operazioni i/o in primo piano e in background
- L'effetto della latenza sul database
- Numerose impostazioni del sistema operativo e di rete che influiscono sulle performance dello storage

Inoltre, occorre prendere in considerazione attività che non riguardano i database di storage. Esiste un punto in cui l'ottimizzazione delle performance dello storage non produce vantaggi utili perché le performance dello storage non sono più un fattore limitante per le performance.

La maggior parte dei clienti che utilizzano database sceglie ora gli array all-flash, il che crea alcune

considerazioni aggiuntive. Ad esempio, prendi in considerazione il test delle performance su un sistema AFF A900 a due nodi:

- Con un rapporto di lettura/scrittura di 80/20:1, due nodi A900 possono fornire oltre 1M IOPS di database casuali prima che la latenza attraversi anche il contrassegno 150µs. Questo ben oltre le attuali richieste di performance della maggior parte dei database è difficile prevedere il miglioramento previsto. Lo storage verrebbe ampiamente cancellato come collo di bottiglia.
- La larghezza di banda della rete è una fonte sempre più comune di limitazioni delle prestazioni. Ad esempio, le soluzioni su disco a rotazione sono spesso dei colli di bottiglia per le performance dei database perché la latenza i/o è molto elevata. Quando un array all-flash rimuove le limitazioni di latenza, spesso la barriera passa alla rete. Si tratta di un aspetto particolarmente interessante nel caso di ambienti virtualizzati e sistemi blade in cui è difficile visualizzare la vera connettività di rete. Ciò può complicare il test delle performance se il sistema di storage stesso non può essere pienamente utilizzato a causa di limitazioni della larghezza di banda.
- Generalmente, il confronto delle performance di un array all-flash con un array contenente dischi rotanti non è possibile a causa dell'aumento drastico della latenza degli array all-flash. I risultati dei test in genere non sono significativi.
- Il confronto delle performance di picco degli IOPS con un array all-flash spesso non è un test utile, in quanto i database non sono limitati dall'i/o dello storage Ad esempio, si supponga che un array sia in grado di sostenere 500K IOPS casuali, mentre un altro possa sostenere 300K KB. La differenza è irrilevante nel mondo reale se un database impiega il 99% del suo tempo per l'elaborazione della CPU. I carichi di lavoro non utilizzano mai le funzionalità complete dello storage array. Al contrario, le funzionalità degli IOPS di picco potrebbero essere critiche in una piattaforma di consolidamento in cui si prevede che lo storage array venga caricato alle proprie funzionalità di picco.
- In qualsiasi test dello storage, si tiene sempre in considerazione sia la latenza che gli IOPS. Molti storage array sul mercato dichiarano livelli estremi di IOPS, ma la latenza rende quegli IOPS inutili a tali livelli. La destinazione tipica degli array all-flash è il contrassegno 1ms. Un approccio migliore al test non consiste nel misurare gli IOPS massimi possibili, ma nel determinare quanti IOPS può supportare uno storage array prima che la latenza media sia superiore a 1ms ms.

Oracle Automatic workload Repository e benchmarking

Il gold standard per i confronti delle performance Oracle è un report Oracle Automatic workload Repository (AWR).

Esistono diversi tipi di rapporti AWR. Da un punto di vista dello storage, un report generato dall'esecuzione di `awrrpt.sql` È il comando più completo e utile, in quanto è destinato a una specifica istanza del database e include alcuni istogrammi dettagliati che suddividono gli eventi i/o dello storage in base alla latenza.

Il confronto fra due array delle performance implica l'esecuzione idealmente dello stesso carico di lavoro su ciascun array e la produzione di un report AWR che punta esattamente al carico di lavoro. Nel caso di un carico di lavoro con esecuzione molto lunga, è possibile utilizzare un singolo rapporto AWR con un tempo trascorso che comprende il tempo di inizio e di fine, ma è preferibile suddividere i dati AWR come rapporti multipli. Ad esempio, se un processo batch è stato eseguito dalla mezzanotte alle 6, creare una serie di rapporti AWR di un'ora dalle 1:1 alle 2:00 e così via.

In altri casi, è necessario ottimizzare una query molto breve. L'opzione migliore è un report AWR basato su uno snapshot AWR creato all'inizio della query e un secondo snapshot AWR creato al termine della query. Il server di database dovrebbe essere altrimenti silenzioso per ridurre al minimo l'attività in background che potrebbe oscurare l'attività della query in analisi.

Laddove i report AWR non sono disponibili, i report statspack Oracle sono una buona alternativa. Contengono la maggior parte delle stesse statistiche i/o di un rapporto AWR.

Oracle AWR e risoluzione dei problemi

Un report AWR è anche lo strumento più importante per analizzare un problema di prestazioni.

Come per il benchmarking, il troubleshooting delle performance richiede la misurazione precisa di un determinato carico di lavoro. Quando possibile, fornisci dati AWR quando segnali un problema di performance al centro di supporto NetApp o quando lavori con un account team NetApp o partner in merito a una nuova soluzione.

Quando si forniscono i dati AWR, considerare i seguenti requisiti:

- Eseguire `awrrpt.sql` per generare il report. L'output può essere di testo o HTML.
- Se si utilizzano Oracle Real Application Clusters (RAC), generare report AWR per ciascuna istanza del cluster.
- Indicare l'ora specifica in cui si è verificato il problema. Il tempo massimo accettabile trascorso di un rapporto AWR è generalmente di un'ora. Se un problema persiste per più ore o richiede un'operazione multi-ora, ad esempio un processo batch, fornire più rapporti AWR di un'ora che coprono l'intero periodo da analizzare.
- Se possibile, regolare l'intervallo dell'istantanea AWR su 15 minuti. Questa impostazione consente di eseguire un'analisi più dettagliata. Ciò richiede anche ulteriori esecuzioni di `awrrpt.sql` per fornire un report per ogni intervallo di 15 minuti.
- Se il problema è una query in esecuzione molto breve, fornire un report AWR basato su uno snapshot AWR creato all'inizio dell'operazione e un secondo snapshot AWR creato al termine dell'operazione. Il server di database dovrebbe essere altrimenti silenzioso per ridurre al minimo l'attività in background che oscurrebbe l'attività dell'operazione in analisi.
- Se viene segnalato un problema di prestazioni in determinati momenti ma non in altri, fornire dati AWR aggiuntivi che dimostrino buone prestazioni per il confronto.

calibra_io

Il `calibrate_io` command non deve mai essere utilizzato per testare, confrontare o eseguire il benchmark dei sistemi storage. Come indicato nella documentazione di Oracle, questa procedura calibra le funzionalità i/o dello storage.

La calibrazione non è la stessa del benchmarking. Lo scopo di questo comando è di emettere i/o per aiutare a calibrare le operazioni di database e migliorarne l'efficienza ottimizzando il livello di i/o inviato all'host. Poiché il tipo di i/o eseguito da `calibrate_io` L'operazione non rappresenta l'i/o effettivo dell'utente del database, i risultati non sono prevedibili e spesso non sono nemmeno riproducibili.

SLOB2

SLOB2, il Silly Little Oracle Benchmark, è diventato lo strumento preferito per la valutazione delle prestazioni del database. È stato sviluppato da Kevin Closson ed è disponibile su "https://kevinclosson.net/slob/". Occorrono pochi minuti per installare e configurare, oltre a utilizzare un database Oracle effettivo per generare schemi di i/o su una tablespace definibile dall'utente. È una delle poche opzioni di test disponibili in grado di saturare un array all-flash con l'i/O. È utile anche per generare livelli molto inferiori di i/o per simulare carichi di lavoro di storage che sono IOPS bassi ma sensibili alla latenza.

Panca di rotazione

Swingbench può essere utile per testare le prestazioni del database, ma è estremamente difficile utilizzare Swingbench in un modo che mette a dura prova lo storage. NetApp non ha riscontrato test da Swingbench che hanno dato i/o sufficienti per essere un carico significativo su qualsiasi array AFF. In casi limitati, è possibile utilizzare Order Entry Test (OET) per valutare lo storage dal punto di vista della latenza. Ciò può essere utile in situazioni in cui un database ha una dipendenza di latenza nota per determinate query. Assicurarsi che l'host e la rete siano configurati correttamente per realizzare i potenziali di latenza di un array all-flash.

HammerDB

HammerDB è uno strumento di test del database che simula, tra gli altri, i benchmark TPC-C e TPC-H. La creazione di un set di dati di dimensioni sufficienti per eseguire correttamente un test può richiedere molto tempo, ma può rivelarsi uno strumento efficace per valutare le prestazioni delle applicazioni OLTP e di data warehouse.

Orion

Lo strumento Oracle Orion è stato comunemente utilizzato con Oracle 9, ma non è stato mantenuto per garantire la compatibilità con le modifiche in vari sistemi operativi host. Viene raramente utilizzato con Oracle 10 o Oracle 11 a causa di incompatibilità con il sistema operativo e la configurazione dello storage.

Oracle ha riscritto lo strumento e viene installato per impostazione predefinita con Oracle 12c. Sebbene questo prodotto sia stato migliorato e utilizzi molte delle stesse chiamate utilizzate da un database Oracle reale, non utilizza esattamente lo stesso percorso di codice o lo stesso comportamento i/o utilizzato da Oracle. Ad esempio, la maggior parte degli i/o Oracle viene eseguita in modo sincrono, il che significa che il database si arresta finché l'i/o non viene completato quando l'operazione i/o viene completata in primo piano. Il semplice flooding di un sistema storage con i/o casuali non rappresenta una riproduzione di i/o Oracle reali e non offre un metodo diretto per confrontare gli array di storage o misurare l'effetto delle modifiche alla configurazione.

Detto questo, ci sono alcuni casi d'utilizzo per Orion, come la misurazione generale delle massime prestazioni possibili di una particolare configurazione host-rete-storage, o per misurare lo stato di un sistema storage. Con un test accurato, è possibile ideare test Orion utilizzabili per confrontare gli storage array o valutare l'effetto di una modifica della configurazione, a condizione che i parametri includano la considerazione di IOPS, throughput e latenza e cercare di replicare fedelmente un carico di lavoro realistico.

NFSv3 serrature obsolete

Se un server di database Oracle si blocca, potrebbe essersi verificato un problema con blocchi NFS obsoleti al riavvio. Questo problema può essere evitato prestando particolare attenzione alla configurazione della risoluzione dei nomi sul server.

Questo problema si verifica perché la creazione di un blocco e la cancellazione di un blocco utilizzano due metodi di risoluzione dei nomi leggermente diversi. Sono coinvolti due processi: Network Lock Manager (NLM) e il client NFS. NLM utilizza `uname n` per determinare il nome host, mentre la `rpc.statd` usi di processo `gethostbyname()`. Questi nomi host devono corrispondere affinché il sistema operativo elimini correttamente i blocchi obsoleti. Ad esempio, l'host potrebbe cercare i blocchi di proprietà di `dbserver5`, ma i blocchi sono stati registrati dall'host come `dbserver5.mydomain.org`. Se `gethostbyname()` non restituisce lo stesso valore di `uname -a`, quindi il processo di rilascio del blocco non ha avuto esito positivo.

Il seguente script di esempio verifica se la risoluzione dei nomi è completamente coerente:

```
#! /usr/bin/perl
$uname=`uname -n`;
chomp($uname);
($name, $aliases, $addrtype, $length, @addrs) = gethostbyname $uname;
print "uname -n yields: $uname\n";
print "gethostbyname yields: $name\n";
```

Se `gethostbyname` non corrisponde `uname`, è probabile che siano presenti blocchi obsoleti. Ad esempio, questo risultato rivela un potenziale problema:

```
uname -n yields: dbserver5
gethostbyname yields: dbserver5.mydomain.org
```

La soluzione viene generalmente trovata modificando l'ordine in cui gli host vengono visualizzati `/etc/hosts`. Ad esempio, si supponga che il file hosts includa questa voce:

```
10.156.110.201  dbserver5.mydomain.org dbserver5 loghost
```

Per risolvere il problema, modificare l'ordine di visualizzazione del nome di dominio completo e del nome host breve:

```
10.156.110.201  dbserver5 dbserver5.mydomain.org loghost
```

`gethostbyname()` ora restituisce il breve `dbserver5` nome host, che corrisponde all'output di `uname`. I blocchi vengono quindi cancellati automaticamente dopo un arresto anomalo del server.

Verifica dell'allineamento di WAFL

Il corretto allineamento dell'WAFL è fondamentale per garantire buone prestazioni. Sebbene ONTAP gestisca blocchi in 4KB unità, questo fatto non significa che ONTAP esegua tutte le operazioni in 4KB unità. Infatti, ONTAP supporta operazioni a blocchi di diverse dimensioni, ma la contabilità sottostante è gestita da WAFL in 4KB unità.

Il termine "allineamento" si riferisce al modo in cui l'i/o Oracle corrisponde a queste unità 4KB. Per ottenere prestazioni ottimali è necessario che un blocco Oracle 8KB risieda su due blocchi fisici da 4KB WAFL su un'unità. Se un blocco è sfalsato di 2KB, questo blocco risiede su metà di un blocco 4KB, un blocco 4KB completo separato e quindi sulla metà di un terzo blocco 4KB. Questa disposizione causa un peggioramento delle prestazioni.

L'allineamento non è un problema con i file system NAS. I file di dati Oracle sono allineati all'inizio del file in base alle dimensioni del blocco Oracle. Pertanto, le dimensioni dei blocchi di 8KB, 16KB e 32KB sono sempre allineate. Tutte le operazioni di blocco sono sfalsate dall'inizio del file in unità di 4 kilobyte.

I LUN, al contrario, contengono generalmente qualche tipo di intestazione del driver o metadati del file system all'inizio che creano un offset. L'allineamento è raramente un problema nei sistemi operativi moderni, perché

questi sistemi operativi sono progettati per unità fisiche che potrebbero utilizzare un settore 4KB nativo, che richiede anche l'allineamento dell'i/o ai confini del 4KB per ottenere prestazioni ottimali.

Ci sono, tuttavia, alcune eccezioni. È possibile che un database sia stato migrato da un sistema operativo meno recente non ottimizzato per i/o 4KB o che un errore utente durante la creazione della partizione abbia causato un offset che non è in unità di 4KB.

I seguenti esempi sono specifici per Linux, ma la procedura può essere adattata per qualsiasi sistema operativo.

Allineato

L'esempio seguente mostra un controllo dell'allineamento su un singolo LUN con una singola partizione.

Innanzitutto, creare la partizione che utilizza tutte le partizioni disponibili sul disco.

```
[root@host0 iscsi]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
disklabel
Building a new DOS disklabel with disk identifier 0xb97f94c1.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.
The device presents a logical sector size that is smaller than
the physical sector size. Aligning to a physical sector (or optimal
I/O) size boundary is recommended, or performance may be impacted.
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-10240, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-10240, default 10240):
Using default value 10240
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
[root@host0 iscsi]#
```

L'allineamento può essere controllato matematicamente con il seguente comando:

```
[root@host0 iscsi]# fdisk -u -l /dev/sdb
Disk /dev/sdb: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 4096 bytes
I/O size (minimum/optimal): 4096 bytes / 65536 bytes
Disk identifier: 0xb97f94c1
   Device Boot      Start         End      Blocks   Id  System
/dev/sdb1              32    20971519    10485744   83  Linux
```

L'output mostra che le unità sono 512 byte, e l'inizio della partizione è 32 unità. Si tratta di un totale di 32 x 512 = 16.834 byte, ovvero un multiplo intero di 4KB blocchi WAFL. Questa partizione è allineata correttamente.

Per verificare il corretto allineamento, attenersi alla seguente procedura:

1. Identificare l'UUID (Universal Unique Identifier) del LUN.

```
FAS8040SAP::> lun show -v /vol/jfs_luns/lun0
              Vserver Name: jfs
                  LUN UUID: ed95d953-1560-4f74-9006-85b352f58fcd
                    Mapped: mapped`                `
```

2. Immettere la shell del nodo sul controller ONTAP.

```
FAS8040SAP::> node run -node FAS8040SAP-02
Type 'exit' or 'Ctrl-D' to return to the CLI
FAS8040SAP-02> set advanced
set not found.  Type '?' for a list of commands
FAS8040SAP-02> priv set advanced
Warning: These advanced commands are potentially dangerous; use
         them only when directed to do so by NetApp
         personnel.
```

3. Avviare le raccolte statistiche sull'UUID di destinazione identificato nel primo passaggio.

```
FAS8040SAP-02*> stats start lun:ed95d953-1560-4f74-9006-85b352f58fcd
Stats identifier name is 'Ind0xffffff08b9536188'
FAS8040SAP-02*>
```

4. Eseguire alcuni i/O. È importante utilizzare `iflag` Argomento per assicurarsi che i/o sia sincrono e non bufferizzato.

Prestare molta attenzione con questo comando. Inversione del `if` e. `of` gli argomenti distruggono i dati.

```
[root@host0 iscsi]# dd if=/dev/sdb1 of=/dev/null iflag=dsync count=1000
bs=4096
1000+0 records in
1000+0 records out
4096000 bytes (4.1 MB) copied, 0.0186706 s, 219 MB/s
```

5. Arrestare le statistiche e visualizzare l'istogramma di allineamento. Tutti i i/o devono trovarsi in `.0` Bucket, che indica i/o allineato al limite di un blocco 4KB.

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff08b9536188
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:186%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
```

Disallineato

L'esempio seguente mostra i/o disallineati:

1. Creare una partizione che non si allinea a un confine 4KB. Questo non è il comportamento predefinito sui sistemi operativi moderni.

```
[root@host0 iscsi]# fdisk -u /dev/sdb
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First sector (32-20971519, default 32): 33
Last sector, +sectors or +size{K,M,G} (33-20971519, default 20971519):
Using default value 20971519
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
```

2. La partizione è stata creata con un offset a 33 settori anziché con il valore predefinito 32. Ripetere la procedura descritta in "Allineato". L'istogramma viene visualizzato come segue:

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:136%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_partial_blocks:31%
```

Il disallineamento è chiaro. L'i/o rientra principalmente in* *`.1` benna, che corrisponde all'offset previsto. Quando la partizione è stata creata, è stata spostata di 512 byte più avanti nel dispositivo rispetto al valore predefinito ottimizzato, il che significa che l'istogramma è spostato di 512 byte.

Inoltre, il `read_partial_blocks` Le statistiche sono diverse da zero, il che significa che è stato eseguito l'i/o che non ha riempito l'intero blocco da 4KB KB.

Ripristina la logging

Le procedure qui spiegate sono applicabili ai file di dati. I log di ripristino e gli archivi di Oracle hanno modelli di i/o diversi. Ad esempio, il redo logging è una sovrascrittura circolare di un singolo file. Se si utilizza la dimensione predefinita del blocco da 512 byte, le statistiche di scrittura sono simili a queste:

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-4f74-
9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.0:12%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.1:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.3:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.4:13%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.5:6%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.6:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.7:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_partial_blocks:85%
```

L'i/o viene distribuito in tutti i bucket di istogramma, ma non si tratta di un problema di prestazioni. Velocità di redo-logging estremamente elevate potrebbero, tuttavia, trarre vantaggio dall'utilizzo di dimensioni del blocco di 4KB KB. In questo caso, è consigliabile assicurarsi che i LUN di redo-logging siano allineati correttamente. Tuttavia, questo non è importante per le buone prestazioni come l'allineamento dei file dati.

PostgreSQL

Panoramica

PostgreSQL viene fornito con varianti che includono PostgreSQL, PostgreSQL Plus ed EDB Postgres Advanced Server (ECAS). PostgreSQL viene in genere distribuito come database back-end per applicazioni multi-Tier. È supportato da pacchetti middleware comuni (come PHP, Java, Python, Tcl/TK, ODBC, E JDBC) ed è stata storicamente una scelta popolare per i sistemi di gestione di database open-source. ONTAP è una scelta eccellente per l'esecuzione di database PostgreSQL per la sua affidabilità, prestazioni elevate ed efficienza di gestione dei dati.

Questa documentazione su ONTAP e il database PostgreSQL sostituisce il database *TR-4770: PostgreSQL precedentemente pubblicato sulle Best practice di ONTAP.*

Con la crescita esponenziale dei dati, la gestione dei dati diventa più complessa per le aziende. Questa complessità aumenta i costi di licenza, operativi, di supporto e di manutenzione. Per ridurre il TCO complessivo, considerare il passaggio da database commerciali a open-source con storage back-end affidabile e dalle performance elevate.

ONTAP è una piattaforma ideale, perché ONTAP è letteralmente progettato per i database. Sono state create numerose funzionalità come le ottimizzazioni della latenza io random per la qualità del servizio avanzata fino alle funzionalità FlexClone di base per rispondere specificamente alle esigenze dei carichi di lavoro dei database.

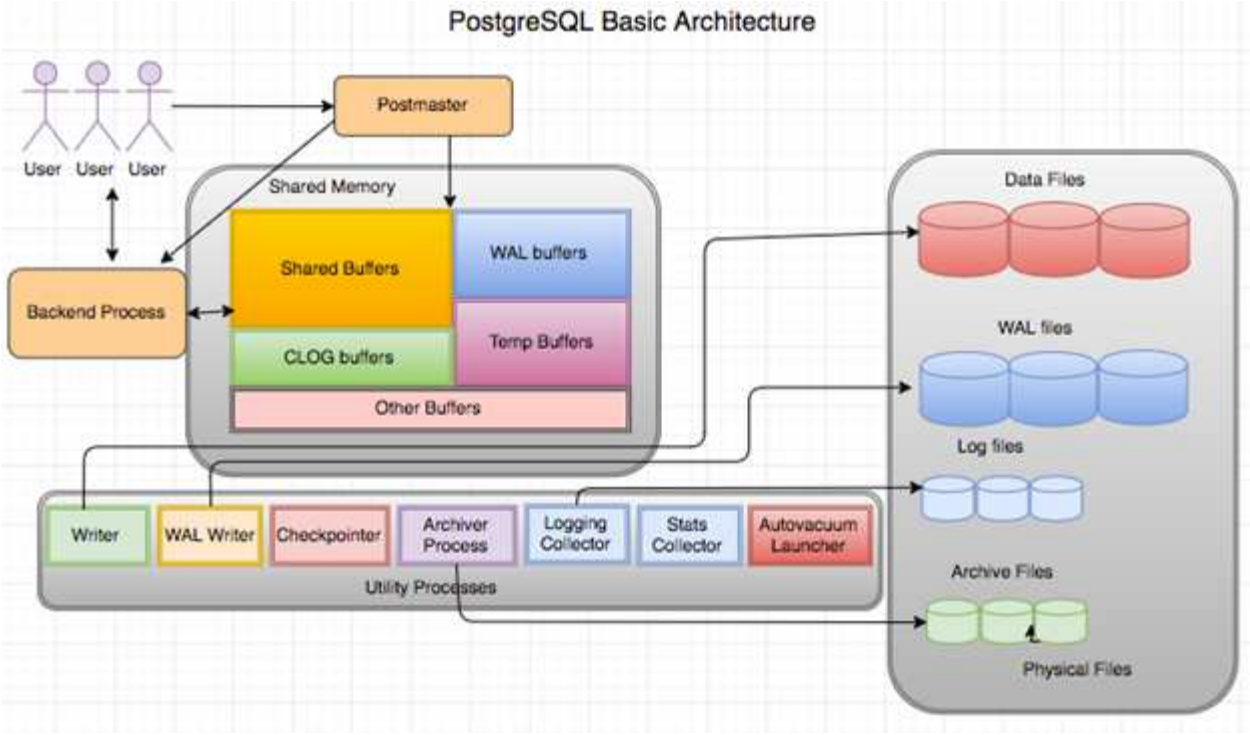
Funzioni aggiuntive come gli aggiornamenti senza interruzioni, (inclusa la sostituzione dello storage) garantiscono la disponibilità dei database critici. Puoi anche disporre di un disaster recovery istantaneo per ambienti di grandi dimensioni tramite MetroCluster o selezionare database tramite la sincronizzazione attiva di SnapMirror.

Soprattutto, ONTAP offre prestazioni senza pari con la possibilità di dimensionare la soluzione in base alle proprie esigenze specifiche. I nostri sistemi high-end possono offrire oltre 1M IOPS con latenze misurate in microsecondi, ma se ti servono solo 100K IOPS, puoi dimensionare al meglio la tua soluzione storage con un controller più piccolo che esegue ancora lo stesso sistema operativo per lo storage.

Configurazione del database

Architettura

PostgreSQL è un RDBMS basato su architettura client e server. Un'istanza di PostgreSQL è nota come cluster di database, ovvero una raccolta di database anziché una raccolta di server.



Un database PostgreSQL contiene tre elementi principali: Il postmaster, il front-end (client) e il back-end Il client invia richieste al postmaster con informazioni quali il protocollo IP e il database a cui connettersi. Il postmaster autentica la connessione e la passa al processo back-end per ulteriori comunicazioni. Il processo back-end esegue la query e invia i risultati direttamente al front-end (client).

Un'istanza PostgreSQL si basa su un modello multiprocesso anziché su un modello multithread. Genera più processi per diversi processi e ogni processo ha una propria funzionalità. I processi principali includono il processo client, il processo di scrittura WAL, il processo di scrittura in background e il processo di checkpointer:

- Quando un processo client (in primo piano) invia richieste di lettura o scrittura all'istanza PostgreSQL, non legge o scrive dati direttamente sul disco. Innanzitutto, memorizza i dati nei buffer condivisi e nei buffer WAL (Write-ahead logging).
- Un processo di scrittura WAL manipola il contenuto dei buffer condivisi e dei buffer WAL da scrivere nei registri WAL. I registri WAL sono in genere registri di transazioni di PostgreSQL e vengono scritti in sequenza. Pertanto, per migliorare i tempi di risposta dal database, PostgreSQL scrive prima nei registri delle transazioni e riconosce il client.
- Per impostare il database in uno stato coerente, il processo di scrittura in background verifica periodicamente la presenza di pagine sporche nel buffer condiviso. Quindi, scarica i dati sui file di dati che sono memorizzati su volumi NetApp o LUN.
- Anche il processo checkpointer viene eseguito periodicamente (meno frequentemente del processo in background) e impedisce qualsiasi modifica ai buffer. Segnala al processo di scrittura WAL di scrivere e svuotare il record del punto di verifica alla fine dei registri WAL memorizzati sul disco NetApp. Segnala inoltre al processo di scrittura in background di scrivere e scaricare tutte le pagine sporche sul disco.

Parametri di inizializzazione

È possibile creare un nuovo cluster di database utilizzando `initdb` programma. An `initdb` script crea i file di dati, le tabelle di sistema e i database dei modelli (template0 e template1) che definiscono il cluster.

Il database dei modelli rappresenta un database di stock. Contiene le definizioni per le tabelle di sistema, le viste standard, le funzioni e i tipi di dati. `pgdata` funge da argomento per il `initdb` script che specifica la posizione del cluster di database.

Tutti gli oggetti di database in PostgreSQL sono gestiti internamente dai rispettivi OID. Le tabelle e gli indici sono inoltre gestiti da singoli OID. Le relazioni tra gli oggetti del database e i rispettivi OID vengono memorizzate in tabelle di cataloghi di sistema appropriate, a seconda del tipo di oggetto. Ad esempio, gli OID dei database e delle tabelle heap vengono memorizzati in `pg_database` e `pg_class, rispettivamente. È possibile determinare gli OID eseguendo query sul client PostgreSQL.

Ogni database ha le proprie tabelle e i file di indice che sono limitati a 1GB. Ogni tabella ha due file associati, rispettivamente con il suffisso `_fsm` e. `_vm`. Sono indicate come mappa dello spazio libero e mappa di visibilità. Questi file memorizzano le informazioni sulla capacità di spazio libero e hanno visibilità su ogni pagina del file di tabella. Gli indici hanno solo mappe di spazio libero individuali e non hanno mappe di visibilità.

Il `pg_xlog/pg_wal` la directory contiene i registri write-ahead. I registri write-ahead sono utilizzati per migliorare l'affidabilità e le performance del database. Ogni volta che si aggiorna una riga in una tabella, PostgreSQL scrive prima la modifica nel registro write-ahead e successivamente scrive le modifiche alle pagine di dati effettive su un disco. Il `pg_xlog` la directory di solito contiene diversi file, ma initdb crea solo il primo. I file aggiuntivi vengono aggiunti in base alle necessità. Ciascun file xlog è lungo 16MB MB.

Impostazioni

Esistono diverse configurazioni di ottimizzazione PostgreSQL che possono migliorare le prestazioni.

I parametri più comunemente utilizzati sono i seguenti:

- `max_connections = <num>`: Il numero massimo di connessioni al database da avere contemporaneamente. Utilizzare questo parametro per limitare lo scambio sul disco e l'interruzione delle prestazioni. A seconda delle esigenze dell'applicazione, è anche possibile regolare questo parametro per le impostazioni del pool di connessione.
- `shared_buffers = <num>`: Il metodo più semplice per migliorare le prestazioni del server di database. Il valore predefinito è basso per la maggior parte dei componenti hardware moderni. Durante l'implementazione viene impostato su circa il 25% della RAM disponibile sul sistema. Questa impostazione di parametro varia in base al funzionamento con determinate istanze di database; potrebbe essere necessario aumentare e diminuire i valori per prova ed errore. Tuttavia, l'impostazione di un livello elevato potrebbe degradare le prestazioni.
- `effective_cache_size = <num>`: Questo valore indica all'ottimizzatore di PostgreSQL la quantità di memoria disponibile per la memorizzazione nella cache dei dati e aiuta a determinare se utilizzare un indice. Un valore maggiore aumenta la probabilità di utilizzare un indice. Questo parametro deve essere impostato sulla quantità di memoria allocata a. `shared_buffers` più la quantità di cache del sistema operativo disponibile. Spesso questo valore corrisponde a più del 50% della memoria di sistema totale.
- `work_mem = <num>`: Questo parametro controlla la quantità di memoria da utilizzare nelle operazioni di ordinamento e nelle tabelle hash. Se si esegue un ordinamento pesante nell'applicazione, potrebbe essere necessario aumentare la quantità di memoria, ma prestare attenzione. Non si tratta di un parametro a livello di sistema, ma di un parametro per operazione. Se una query complessa contiene diverse operazioni di ordinamento, utilizza più unità di memoria work_mem e più backend potrebbero farlo contemporaneamente. Questa query può spesso indurre il server di database a effettuare lo swap se il valore è troppo grande. Questa opzione era precedentemente chiamata sort_mem nelle versioni precedenti di PostgreSQL.

- `fsync = <boolean> (on or off)`: Questo parametro determina se tutte le pagine WAL devono essere sincronizzate su disco utilizzando fsync() prima che venga eseguito il commit di una transazione. Disattivandolo a volte si possono migliorare le prestazioni di scrittura e attivandolo si aumenta la protezione dal rischio di danneggiamento quando il sistema si blocca.
- `checkpoint_timeout`: Il processo del punto di verifica elimina i dati sottoposti a commit sul disco. Ciò comporta numerose operazioni di lettura/scrittura su disco. Il valore è impostato in secondi e valori inferiori riducono il tempo di recupero da crash e valori crescenti possono ridurre il carico sulle risorse di sistema riducendo le chiamate al punto di verifica. In base alla criticità dell'applicazione, all'utilizzo, alla disponibilità del database, impostare il valore di checkpoint_timeout.
- `commit_delay = <num>` e. `commit_siblings = <num>`: Queste opzioni vengono utilizzate insieme per migliorare le prestazioni scrivendo più transazioni che vengono effettuate contemporaneamente. Se ci sono diversi oggetti commit_siblings attivi nel momento in cui la transazione è in fase di commit, il server attende Commit_delay microsecondi per tentare di eseguire più transazioni contemporaneamente.
- `max_worker_processes / max_parallel_workers`: Configurare il numero ottimale di lavoratori per i processi. Max_Parallel_Workers corrisponde al numero di CPU disponibili. A seconda della progettazione dell'applicazione, le query potrebbero richiedere un numero minore di lavoratori per le operazioni parallele. È meglio mantenere lo stesso valore per entrambi i parametri, ma regolare il valore dopo la verifica.
- `random_page_cost = <num>`: Questo valore controlla il modo in cui PostgreSQL visualizza le letture del disco non sequenziali. Un valore più elevato indica che PostgreSQL è più probabile che utilizzi una scansione sequenziale invece di una scansione di indice, indicando che il server dispone di dischi veloci modificare questa impostazione dopo aver valutato altre opzioni come l'ottimizzazione basata su piano, l'aspirazione, l'indicizzazione per modificare query o schemi.
- `effective_io_concurrency = <num>`: Questo parametro imposta il numero di operazioni di i/o su disco simultanee che PostgreSQL tenta di eseguire contemporaneamente. L'aumento di questo valore aumenta il numero di operazioni di i/o che una singola sessione PostgreSQL tenta di avviare in parallelo. L'intervallo consentito è compreso tra 1 e 1.000 o zero per disattivare l'emissione di richieste i/o asincrone. Attualmente, questa impostazione influisce solo sulle scansioni bitmap heap. I dischi a stato solido (SSD) e altro storage basato su memoria (NVMe) possono spesso elaborare molte richieste simultanee, cosicché il valore migliore può essere centinaia.

Consultare la documentazione di PostgreSQL per un elenco completo dei parametri di configurazione di PostgreSQL.

TOAST

TOAST è l'acronimo di OVERSIZED-Attribute Storage Technique. PostgreSQL utilizza una dimensione di pagina fissa (in genere 8KB) e non consente alle tuple di occupare più pagine. Pertanto, non è possibile memorizzare direttamente valori di campo grandi. Quando si tenta di memorizzare una riga che supera queste dimensioni, TOAST suddivide i dati delle colonne di grandi dimensioni in "pezzi" più piccoli e li memorizza in una tabella TOAST.

I valori elevati degli attributi tostati vengono estratti (se selezionati) solo quando il set di risultati viene inviato al client. La tabella stessa è molto più piccola e può contenere più righe nella cache buffer condivisa di quanto non possa fare senza alcuna archiviazione out-of-line (TOAST).

VUOTO

Nelle normali operazioni PostgreSQL, le tuple eliminate o rese obsolete da un aggiornamento non vengono fisicamente rimosse dalla tabella; rimangono presenti fino all'esecuzione di VACUUM. Pertanto, è necessario eseguire il VUOTO periodicamente, soprattutto nelle tabelle aggiornate di frequente. Lo spazio occupato deve quindi essere recuperato per essere riutilizzato da nuove righe, per evitare di esaurire lo spazio su disco. Tuttavia, non restituisce lo spazio al sistema operativo.

Lo spazio libero all'interno di una pagina non è frammentato. L'ASPIRAPOLVERE riscrive l'intero blocco, comprimendo in modo efficiente le righe rimanenti e lasciando un singolo blocco contiguo di spazio libero in una pagina.

Al contrario, VACUUM FULL comprime attivamente le tabelle scrivendo una versione completamente nuova del file di tabella senza spazio morto. Questa azione riduce al minimo le dimensioni della tabella, ma può richiedere molto tempo. Richiede inoltre ulteriore spazio su disco per la nuova copia della tabella fino al completamento dell'operazione. L'obiettivo del VUOTO DI routine è di evitare l'attività di VUOTO PIENO. Questo processo non solo mantiene le tabelle alla loro dimensione minima, ma mantiene anche l'utilizzo costante dello spazio su disco.

Tablespace

Due tablespace vengono create automaticamente al momento dell'inizializzazione del cluster di database.

Il `pg_global` tablespace viene utilizzato per i cataloghi di sistema condivisi. Il `pg_default` tablespace è la tablespace predefinita dei database template1 e template0. Se la partizione o il volume su cui il cluster è stato inizializzato esaurisce lo spazio e non può essere esteso, è possibile creare uno spazio di tabella in un'altra partizione ed utilizzarlo fino a quando il sistema non può essere riconfigurato.

Un indice molto utilizzato può essere collocato su un disco veloce e altamente disponibile, come un dispositivo a stato solido. Inoltre, una tabella che memorizza i dati archiviati utilizzati raramente o non critici per le prestazioni può essere archiviata su un sistema su disco meno costoso e più lento, come le unità SAS o SATA.

Gli spazi di tabella fanno parte del cluster di database e non possono essere trattati come una raccolta autonoma di file di dati. Dipendono dai metadati contenuti nella directory dei dati principale e pertanto non possono essere collegati a un cluster di database diverso o sottoposti a backup individuale. Analogamente, se si perde uno spazio di tabella (a causa dell'eliminazione dei file, del guasto del disco e così via), il cluster del database potrebbe diventare illeggibile o non avviarsi. Posizionando una tablespace su un file system temporaneo come un disco RAM si rischia l'affidabilità dell'intero cluster.

Una volta creato, è possibile utilizzare un tablespace da qualsiasi database se l'utente richiedente dispone di privilegi sufficienti. PostgreSQL utilizza collegamenti simbolici per semplificare l'implementazione di tablespace. PostgreSQL aggiunge una riga al `pg_tablespace` Tabella (una tavola a livello di cluster) e assegna un nuovo identificatore di oggetto (OID) a quella riga. Infine, il server utilizza l'OID per creare un collegamento simbolico tra il cluster e la directory specificata. La directory `$PGDATA/pg_tblspc` contiene collegamenti simbolici che puntano a ciascuno degli spazi di tabella non incorporati definiti nel cluster.

Configurazione dello storage

NFS

I database PostgreSQL possono essere ospitati su filesystem NFSv3 o NFSv4. L'opzione migliore dipende da fattori esterni al database.

Per esempio, il comportamento di bloccaggio di NFSv4 può essere preferibile in certi ambienti raggruppati. (Vedere "qui" per ulteriori informazioni)

In caso contrario, la funzionalità del database dovrebbe essere quasi identica, incluse le prestazioni. L'unico requisito è l'uso di `hard` opzione di montaggio. Questo è necessario per garantire che i timeout software non producano errori io irreversibili.

Se si sceglie NFSv4 come protocollo, NetApp consiglia di utilizzare NFSv4,1. Nel NFSv4,1 sono stati apportati alcuni miglioramenti funzionali al protocollo NFSv4 che migliorano la resilienza rispetto al NFSv4,0.

Utilizzare le seguenti opzioni di montaggio per i carichi di lavoro generali del database:

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=65536,wsize=65536
```

Se si prevede un io sequenziale pesante, le dimensioni del trasferimento NFS possono essere aumentate come descritto nella sezione seguente.

Dimensioni trasferimento NFS

Per impostazione predefinita, ONTAP limita le dimensioni i/o NFS a 64K.

L'i/o casuale con la maggior parte delle applicazioni e dei database utilizza blocchi di dimensioni molto inferiori, ben al di sotto del limite massimo di 64K KB. L'i/o a blocchi di grandi dimensioni è solitamente a parallelismo, pertanto anche il massimo di 64K Gbps non costituisce un limite all'ottenimento della massima larghezza di banda.

Ci sono alcuni carichi di lavoro in cui il massimo di 64K crea un limite. In particolare, le operazioni single-threaded, come l'operazione di backup o ripristino o la scansione di un database completa della tabella, vengono eseguite più velocemente e in modo più efficiente se il database è in grado di eseguire un numero di i/o inferiore ma maggiore. Le dimensioni ottimali per la gestione i/o per ONTAP sono 256K KB.

Le dimensioni massime di trasferimento per una SVM ONTAP possono essere modificate come segue:

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

> Non diminuire mai la dimensione massima di trasferimento consentita su ONTAP al di sotto del valore rsize/wsize dei file system NFS attualmente montati. In alcuni sistemi operativi, ciò può causare blocchi o addirittura danni ai dati. Ad esempio, se i client NFS sono attualmente impostati su un valore rsize/wsize di 65536, la dimensione massima di trasferimento ONTAP potrebbe essere regolata tra 65536 e 1048576 senza alcun effetto perché i client stessi sono limitati. La riduzione della dimensione massima di trasferimento inferiore a 65536 GB può danneggiare la disponibilità o i dati.

Una volta aumentata la dimensione di trasferimento a livello ONTAP, si utilizzeranno le seguenti opzioni di montaggio:

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=262144,wsize=262144
```

NFSv3 tabelle slot TCP

Se NFSv3 viene usato con Linux, è fondamentale impostare correttamente le tabelle degli slot TCP.

Le tabelle degli slot TCP sono l'equivalente di NFSv3 della profondità della coda degli HBA (host Bus Adapter). Queste tabelle controllano il numero di operazioni NFS che possono essere in sospeso in qualsiasi momento. Il valore predefinito è di solito 16, che è troppo basso per ottenere prestazioni ottimali. Il problema opposto si verifica sui kernel Linux più recenti, che possono aumentare automaticamente il limite della tabella degli slot TCP a un livello che satura il server NFS con le richieste.

Per prestazioni ottimali e per evitare problemi di prestazioni, regolare i parametri del kernel che controllano le tabelle degli slot TCP.

Eseguire `sysctl -a | grep tcp.*.slot_table` e osservare i seguenti parametri:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tutti i sistemi Linux dovrebbero includere `sunrpc.tcp_slot_table_entries`, ma solo alcuni includono `sunrpc.tcp_max_slot_table_entries`. Entrambi devono essere impostati su 128.

> La mancata impostazione di questi parametri può avere effetti significativi sulle prestazioni. In alcuni casi, le prestazioni sono limitate poiché il sistema operativo linux non fornisce i/o sufficienti In altri casi, le latenze i/o aumentano quando il sistema operativo linux tenta di emettere più i/o di quanto possa essere gestito.

SAN

I database PostgreSQL con SAN sono generalmente ospitati su filesystem xfs, ma altri possono essere utilizzati se supportati dal fornitore del sistema operativo

Mentre un singolo LUN può generalmente supportare fino a 100K IOPS, i database io-intensive richiedono generalmente l'utilizzo di LVM con lo striping.

Striping LVM

Prima dell'era dei dischi flash, era stato utilizzato lo striping per superare i limiti di performance dei dischi rotanti. Ad esempio, se un sistema operativo deve eseguire un'operazione di lettura a 1MB bit, la lettura di 1MB GB di dati da un'unica unità richiederebbe un'ampia ricerca e lettura della testina dell'unità poiché il sistema 1MB viene trasferito lentamente. Se quei 1MB TB di dati sono stati suddivisi in 8 LUN, il sistema operativo potrebbe emettere otto operazioni di lettura 128K in parallelo, riducendo il tempo necessario per completare il trasferimento da 1MB GB.

Lo striping con dischi rotanti era più difficile perché lo schema di i/o doveva essere noto in anticipo. Se lo striping non è stato regolato correttamente per i modelli i/o reali, le configurazioni con striping potrebbero danneggiare le prestazioni. Con i database Oracle, e in particolare con le configurazioni all-flash, lo striping è molto più semplice da configurare ed è stato dimostrato che le performance risultano notevolmente migliorate.

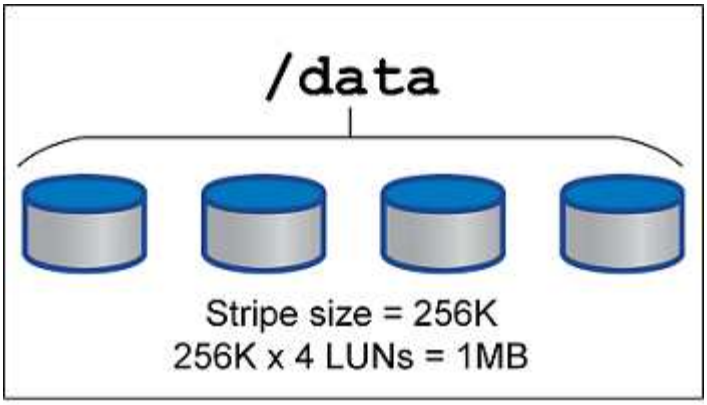
Per impostazione predefinita, i gestori di volume logici, come lo stripe di Oracle ASM, ma il sistema operativo LVM nativo non lo fanno. Alcune di esse collegano più LUN insieme come un dispositivo concatenato, il che

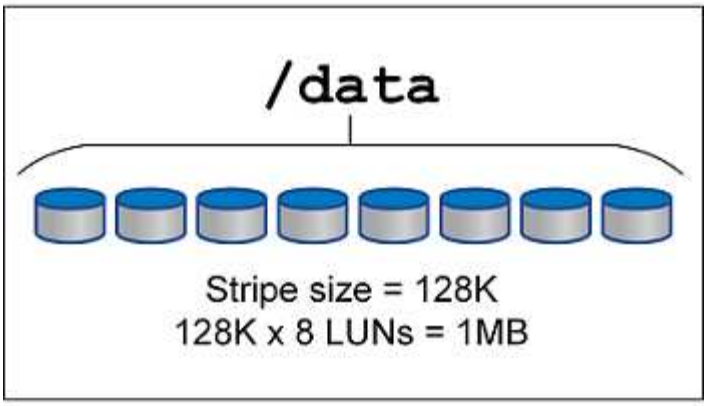
comporta file di dati che esistono su un solo dispositivo LUN. Ciò causa punti caldi. Le altre implementazioni LVM sono impostate per impostazione predefinita su estensioni distribuite. Questo è simile allo striping, ma è più grossolano. I LUN nel gruppo di volumi vengono suddivisi in porzioni di grandi dimensioni, chiamate estensioni e generalmente misurati in molti megabyte, e i volumi logici vengono quindi distribuiti tra tali estensioni. Il risultato è un i/o casuale per un file dovrebbe essere ben distribuito tra i LUN, ma le operazioni i/o sequenziali non sono così efficienti come potrebbero essere.

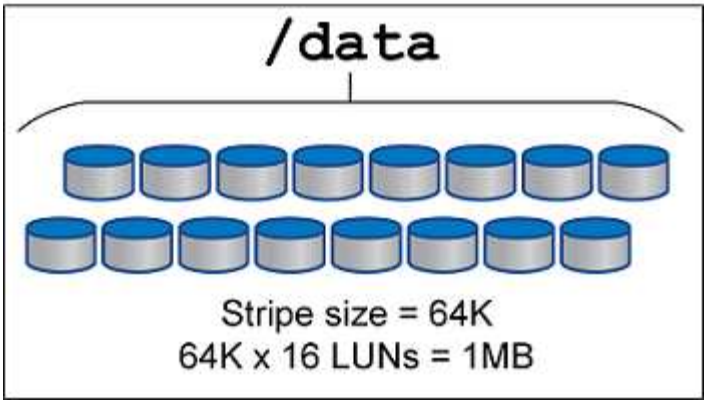
L'i/o delle applicazioni che richiedono elevate performance è quasi sempre (a) in unità delle dimensioni dei blocchi di base o (b) un megabyte.

L'obiettivo principale di una configurazione con striping è quello di garantire che l'i/o a file singolo possa essere eseguito come una singola unità, mentre l'i/o a blocchi multipli, di dimensioni pari a 1MB GB, può essere parallelizzato in modo uniforme tra tutti i LUN del volume con striping. Ciò significa che la dimensione dello stripe non deve essere inferiore alla dimensione del blocco del database e che la dimensione dello stripe moltiplicata per il numero di LUN deve essere 1MB.

La figura seguente mostra tre possibili opzioni per la regolazione delle dimensioni dello stripe e della larghezza. Il numero di LUN viene selezionato per soddisfare i requisiti di prestazioni come descritto sopra, ma in tutti i casi i dati totali all'interno di uno stripe singolo sono 1MB.







Protezione dei dati

Protezione DDTA nativa

Uno degli aspetti principali della progettazione dello storage è la protezione dei volumi PostgreSQL. I clienti possono proteggere i database PostgreSQL utilizzando l'approccio dump o i backup del file system. In questa sezione vengono illustrati i diversi approcci per il backup di singoli database o dell'intero cluster.

Sono disponibili tre approcci per il backup dei dati PostgreSQL:

- Dump di SQL Server
- Backup a livello di file system
- Archiviazione continua

L'idea alla base del metodo dump di SQL Server è generare un file con comandi di SQL Server che, quando viene restituito al server, può ricreare il database così come era al momento del dump. PostgreSQL fornisce i programmi di utilità `pg_dump` e. `pg_dump_all` per la creazione di backup singolo e a livello di cluster. Questi dump sono logici e non contengono informazioni sufficienti per essere utilizzati da WAL Replay.

Una strategia di backup alternativa consiste nell'utilizzare il backup a livello di file system, in cui gli amministratori copiano direttamente i file utilizzati da PostgreSQL per memorizzare i dati nel database. Questo metodo viene eseguito in modalità non in linea: Il database o il cluster devono essere chiusi. Un'altra alternativa è quella di utilizzare `pg_basebackup` Per eseguire il backup hot streaming del database PostgreSQL.

Snapshot

I backup basati su snapshot con PostgreSQL richiedono la configurazione di snapshot per file di dati, file WAL e file WAL archiviati per garantire un ripristino completo o point-in-time.

Per i database PostgreSQL, il tempo medio di backup con gli snapshot è compreso tra pochi secondi e pochi minuti. Questa velocità di backup è da 60 a 100 volte più veloce di `pg_basebackup` e altri approcci di backup basati sul file system.

Le snapshot sullo storage NetApp possono essere coerenti con il crash e con l'applicazione. Viene creato uno snapshot coerente con i crash sullo storage senza chiudere il database, mentre uno snapshot coerente con l'applicazione viene creato mentre il database è in modalità backup. NetApp garantisce inoltre che le snapshot successive siano backup incrementali perenni, per promuovere il risparmio dello storage e l'efficienza della rete.

Poiché le snapshot sono rapide e non influiscono sulle prestazioni del sistema, è possibile pianificare snapshot multiple ogni giorno invece di creare un unico backup giornaliero come avviene con l'altra tecnologia di backup in streaming. Quando è necessaria un'operazione di ripristino e ripristino, il downtime del sistema viene ridotto da due caratteristiche principali:

- La tecnologia di recovery di dati NetApp SnapRestore consente di eseguire l'operazione di ripristino in pochi secondi.
- Obiettivi di recovery point (RPO) aggressivi richiedono l'applicazione di un numero inferiore di log dei database e un'accelerazione del recovery in avanti.

Per eseguire il backup di PostgreSQL, è necessario assicurarsi che i volumi di dati siano protetti contemporaneamente con WAL (gruppo di coerenza) e i registri archiviati. Mentre si utilizza la tecnologia Snapshot per copiare i file WAL, assicurarsi di eseguire `pg_stop` Per svuotare tutte le voci WAL che devono essere archiviate. Se si svuotano le voci WAL durante il ripristino, sarà sufficiente arrestare il database, smontare o eliminare la directory dei dati esistente ed eseguire un'operazione SnapRestore sull'archiviazione. Al termine del ripristino, è possibile montare il sistema e riportarlo allo stato corrente. Per il ripristino point-in-time, è anche possibile ripristinare i registri WAL e di archivio; quindi PostgreSQL decide il punto più coerente e lo recupera automaticamente.

I gruppi di coerenza sono una funzionalità di ONTAP e sono consigliati quando ci sono più volumi montati su una singola istanza o su un database con tablespace multiple. Uno snapshot del gruppo di coerenza garantisce che tutti i volumi siano raggruppati e protetti. È possibile gestire in modo efficiente un gruppo di coerenza da ONTAP System Manager, clonandolo per creare una copia dell'istanza di un database a scopo di test o sviluppo.

Software per la data Protection

Il plug-in NetApp SnapCenter per i database PostgreSQL, combinato con le tecnologie Snapshot e NetApp FlexClone, offre diversi vantaggi, tra cui:

- Backup e ripristino rapidi.
- Cloni efficienti in termini di spazio.
- La capacità di creare un sistema di disaster recovery rapido ed efficace.

> Potresti preferire scegliere i partner di backup premium di NetApp come Veeam Software e CommVault nelle seguenti circostanze:
>
> - Gestire i carichi di lavoro in un ambiente eterogeneo
> - Memorizzazione dei backup su cloud o nastro per una conservazione a lungo termine
> - Supporto per un'ampia gamma di versioni e tipi di sistema operativo

Il plug-in SnapCenter per PostgreSQL è un plugin supportato dalla comunità e la configurazione e la documentazione sono disponibili nell'archivio automazione di NetApp. Tramite SnapCenter, l'utente può eseguire il backup di database, clonare e ripristinare i dati in remoto.

SAP

VMware

VMware vSphere con ONTAP

VMware vSphere con ONTAP

ONTAP è stata una soluzione storage leader per VMware vSphere e, più di recente, per gli ambienti Cloud Foundation dalla sua introduzione nel moderno data center nel 2002. Continua a introdurre funzionalità innovative che semplificano la gestione e riducono i costi.

In questo documento viene presentata la soluzione ONTAP per vSphere, che mette in evidenza le più recenti informazioni sui prodotti e le Best practice per semplificare la distribuzione, ridurre i rischi e semplificare la gestione.

> Questa documentazione sostituisce i report tecnici precedentemente pubblicati *TR-4597: VMware vSphere for ONTAP*

Le Best practice integrano altri documenti come guide ed elenchi di compatibilità. Sono sviluppati in base a test di laboratorio e a un'ampia esperienza sul campo da parte di tecnici e clienti NetApp. Potrebbero non essere le uniche pratiche supportate che funzionano in ogni ambiente, ma sono generalmente le soluzioni più semplici che soddisfano le esigenze della maggior parte dei clienti.

Questo documento si concentra sulle funzionalità delle versioni recenti di ONTAP (9.x) in esecuzione su vSphere 7,0 o versioni successive. Vedere "Tool di matrice di interoperabilità (IMT)" e "Guida alla compatibilità VMware" per i dettagli relativi a versioni specifiche.

Perché scegliere ONTAP per VMware vSphere?

I clienti scelgono con fiducia ONTAP per vSphere sia per le soluzioni di storage SAN che NAS. La nuova architettura di storage disaggregato semplificata, presente negli ultimi All SAN Array, offre un'esperienza semplificata familiare agli amministratori di storage SAN, pur mantenendo la maggior parte delle integrazioni e delle funzionalità dei sistemi ONTAP tradizionali. I sistemi ONTAP forniscono un'eccezionale protezione snapshot e solidi strumenti di gestione. Trasferendo le funzioni su un archivio dedicato, ONTAP massimizza le risorse host, riduce i costi e mantiene prestazioni ottimali. Inoltre, i carichi di lavoro possono essere facilmente migrati tramite Storage vMotion su VMFS, NFS o vVols.

I vantaggi dell'utilizzo di ONTAP per vSphere

Sono molti i motivi per cui decine di migliaia di clienti hanno scelto ONTAP come soluzione storage per vSphere, ad esempio un sistema storage unificato che supporta protocolli SAN e NAS, solide funzionalità di protezione dei dati che utilizzano snapshot efficienti in termini di spazio e molti strumenti per aiutarti a gestire i dati delle applicazioni. L'utilizzo di un sistema storage separato dall'hypervisor consente di trasferire molte funzioni e massimizzare l'investimento nei sistemi host vSphere. Questo approccio non solo garantisce che le risorse host siano incentrate sui carichi di lavoro delle applicazioni, ma evita anche effetti casuali sulle performance delle applicazioni derivanti dalle operazioni di storage.

L’utilizzo di ONTAP insieme a vSphere è un’ottima combinazione che consente di ridurre le spese per l’hardware host e il software VMware. Puoi anche proteggere i tuoi dati a un costo inferiore mantenendo prestazioni elevate e costanti. Poiché i carichi di lavoro virtualizzati sono mobili, è possibile esplorare diversi approcci utilizzando Storage vMotion per spostare le VM tra datastore VMFS, NFS o vVols , tutti sullo stesso sistema di storage.

Ecco i fattori chiave che i clienti apprezzano oggi:

- **Archiviazione unificata.** I sistemi che eseguono ONTAP sono unificati in diversi modi significativi. Originariamente, questo approccio si riferiva sia ai protocolli NAS che SAN e ONTAP continua a essere una piattaforma leader per SAN, oltre alla sua forza originaria in NAS. Nel mondo vSphere, questo approccio potrebbe anche significare un sistema unificato per l’infrastruttura desktop virtuale (VDI) insieme all’infrastruttura server virtuale (VSI). I sistemi che eseguono ONTAP sono in genere meno costosi per VSI rispetto ai tradizionali array aziendali e tuttavia dispongono di funzionalità avanzate di efficienza di archiviazione per gestire VDI nello stesso sistema. ONTAP unifica inoltre una varietà di supporti di archiviazione, dagli SSD ai SATA, e può estenderli facilmente al cloud. Non è necessario acquistare un sistema operativo di archiviazione per le prestazioni, un altro per gli archivi e un altro ancora per il cloud. ONTAP li collega tutti insieme.
- **All SAN Array (ASA).** I sistemi ONTAP ASA più recenti (a partire dai modelli A1K, A90, A70, A50, A30 e A20) sono costruiti su una nuova architettura dello storage in grado di eliminare il tradizionale paradigma dello storage ONTAP per la gestione degli aggregati e dei volumi. Poiché non ci sono condivisioni di file system, non c’è bisogno di volumi! Tutto lo storage collegato a una coppia ha viene trattato come una SAZ (Storage Availability zone) comune, all’interno della quale i LUN e i namespace NVMe vengono forniti come "Storage Units" (SUS). I più recenti sistemi ASA sono progettati per essere semplici da gestire e con un’esperienza familiare per gli amministratori dello storage SAN. Questa nuova architettura è ideale per gli ambienti vSphere, poiché consente una facile gestione delle risorse storage e fornisce un’esperienza semplificata per gli amministratori dello storage SAN. L’architettura ASA supporta anche la più recente tecnologia NVMe over Fabrics (NVMe-of), che offre performance e scalabilità ancora superiori per i workload vSphere.
- **Tecnologia Snapshot.** ONTAP è stata la prima azienda a offrire la tecnologia Snapshot per la protezione dei dati, che rimane la più avanzata del settore. Questo approccio efficiente in termini di spazio alla data Protection è stato esteso per supportare le API VMware vSphere per l’integrazione degli array (VAAI). Questa integrazione ti consente di sfruttare le funzionalità snapshot di ONTAP per le operazioni di backup e ripristino, riducendo l’impatto sul tuo ambiente di produzione. Questo approccio consente inoltre di utilizzare le snapshot per un rapido recovery delle macchine virtuali, riducendo tempo e lavoro necessari per il ripristino dei dati. Inoltre, la tecnologia snapshot di ONTAP è integrata con le soluzioni VLSR (Live Site Recovery Manager) di VMware, in precedenza Site Recovery Manager [SRM], per fornire una strategia di protezione dei dati completa per il vostro ambiente virtualizzato.
- **Gestione basata su policy di archiviazione e volumi virtuali.** NetApp è stata uno dei primi partner di progettazione di VMware nello sviluppo di vSphere Virtual Volumes (vVols), fornendo input architettonici e supporto iniziale per vVols e VMware vSphere API for Storage Awareness (VASA). Questo approccio non solo ha introdotto la gestione granulare dello storage delle VM in VMFS, ma ha anche supportato l’automazione del provisioning dello storage tramite una gestione basata su policy di storage. Questo approccio consente agli architetti di storage di progettare pool di storage con diverse funzionalità che possono essere facilmente utilizzate dagli amministratori delle VM. ONTAP è leader nel settore dello storage in termini di scala vVol, supportando centinaia di migliaia di vVols in un singolo cluster, mentre i fornitori di array aziendali e di array flash più piccoli supportano solo alcune migliaia vVols per array. NetApp sta inoltre guidando l’evoluzione della gestione granulare delle VM con le sue funzionalità future.
- **Efficienza di archiviazione.** Sebbene NetApp sia stata la prima a fornire la deduplicazione per i carichi di lavoro di produzione, questa innovazione non è stata né la prima né l’ultima in questo settore. Tutto è iniziato con gli snapshot, un meccanismo di protezione dei dati efficiente in termini di spazio e senza effetti sulle prestazioni, insieme alla tecnologia FlexClone per creare istantaneamente copie di lettura/scrittura delle VM per uso in produzione e backup. NetApp ha continuato a fornire funzionalità inline, tra cui

deduplicazione, compressione e deduplicazione a blocchi zero, per ottenere il massimo spazio di archiviazione dai costosi SSD. ONTAP ha inoltre aggiunto la possibilità di impacchettare operazioni di I/O e file più piccoli in un blocco di disco utilizzando la compattazione. La combinazione di queste funzionalità ha portato i clienti a ottenere risparmi fino a 5:1 per VSI e fino a 30:1 per VDI. La nuova generazione di sistemi ONTAP include anche la compressione e la deduplicazione accelerate dall'hardware, che possono migliorare ulteriormente l'efficienza di archiviazione e ridurre i costi. Questo approccio consente di archiviare più dati in meno spazio, riducendo il costo complessivo di archiviazione e migliorando le prestazioni. NetApp è così sicura delle sue capacità di efficienza di storage che offre un collegamento: https://www.netapp.com/pdf.html?item=/media/79014-ng-937-Efficiency-Guarantee-Customer-Flyer.pdf [Garanzia di efficienza^].

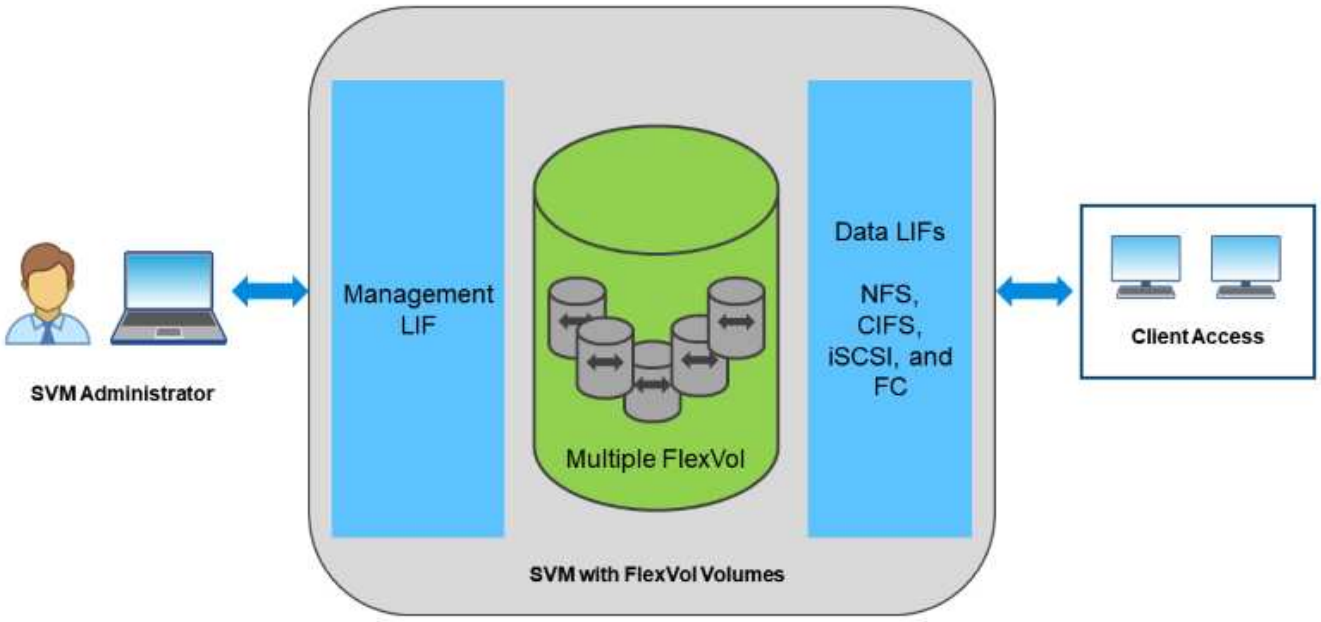
- **Multi-tenancy.** ONTAP è da tempo leader nel multitenancy, consentendo di creare più macchine virtuali di storage (SVM) su un singolo cluster. Questo approccio consente di isolare i carichi di lavoro e di fornire diversi livelli di servizio a diversi tenant, rendendolo ideale per i fornitori di servizi e le grandi aziende. L'ultima generazione di sistemi ONTAP include anche il supporto per la gestione della capacità degli inquilini. Questa funzionalità consente di impostare limiti di capacità per ciascun tenant, assicurando che nessun tenant possa consumare tutte le risorse disponibili. Questo approccio contribuisce a garantire che tutti gli inquilini ricevano il livello di servizio che si aspettano, garantendo al contempo un elevato livello di sicurezza e isolamento tra gli inquilini. Inoltre, le funzionalità multi-tenancy di ONTAP sono integrate con la piattaforma vSphere di VMware, consentendo di gestire e monitorare facilmente l'ambiente virtualizzato tramite "Strumenti ONTAP per VMware vSphere" E "Informazioni sull'infrastruttura dati" .
- **Cloud ibrido.** Che vengano utilizzate per un cloud privato on-premise, un'infrastruttura cloud pubblica o un cloud ibrido che combina il meglio di entrambi, le soluzioni ONTAP ti aiutano a creare il tuo data fabric per semplificare e ottimizzare la gestione dei dati. Inizia con sistemi all-flash ad alte prestazioni, quindi abbinali a sistemi di archiviazione su disco o cloud per la protezione dei dati e il cloud computing. Scegli tra Azure, AWS, IBM o Google Cloud per ottimizzare i costi ed evitare vincoli. Sfrutta il supporto avanzato per OpenStack e le tecnologie container in base alle tue esigenze. NetApp offre inoltre strumenti di backup basati su cloud (SnapMirror Cloud, Cloud Backup Service e Cloud Sync) e di archiviazione e suddivisione in livelli di storage (FabricPool) per ONTAP , per contribuire a ridurre le spese operative e sfruttare l'ampia portata del cloud.
- **E altro ancora.** sfrutta le performance estreme degli array NetApp AFF Serie A per accelerare l'infrastruttura virtualizzata e gestire i costi. Operazioni senza interruzioni, dalla manutenzione agli aggiornamenti fino alla sostituzione completa del sistema storage, utilizzando cluster ONTAP scale-out. Proteggi i dati inattivi con le funzionalità di crittografia NetApp senza costi aggiuntivi. Assicurati che le performance soddisfino i livelli di servizio di business grazie a funzionalità di qualità dei servizi. Fanno tutti parte dell'ampia gamma di funzionalità offerte da ONTAP, il software di Enterprise data management leader del settore.

Storage unificato

ONTAP unifica lo storage tramite un approccio software-defined semplificato per una gestione sicura ed efficiente, performance migliorate e una perfetta scalabilità. Questo approccio migliora la protezione dei dati e consente un uso efficace delle risorse cloud.

In origine, questo approccio unificato ha indicato il supporto dei protocolli NAS e SAN su un unico sistema di storage e ONTAP continua a essere una piattaforma leader per SAN e la sua forza originale nel campo delle NAS. ONTAP ora fornisce anche il supporto del protocollo a oggetti S3. Sebbene S3 non sia utilizzato per i datastore, è possibile utilizzarlo per le applicazioni in-guest. Per ulteriori informazioni sul supporto del protocollo S3 in ONTAP, consultare la "Panoramica della configurazione S3". Il termine storage unificato si è evoluto per indicare un approccio unificato alla gestione dello storage, inclusa la capacità di gestire tutte le risorse di storage da una singola interfaccia. Tra cui la possibilità di gestire le risorse di storage sia on-premise che nel cloud, i più recenti sistemi All SAN Array (ASA) e la possibilità di gestire più sistemi storage da una singola interfaccia.

Una Storage Virtual Machine (SVM) è l'unità di multi-tenancy sicura in ONTAP. Si tratta di un costrutto logico che consente l'accesso client ai sistemi che eseguono ONTAP. Le SVM possono servire i dati contemporaneamente attraverso più protocolli di accesso ai dati tramite le interfacce logiche (LIF). Le SVM forniscono l'accesso ai dati a livello di file attraverso protocolli NAS, come CIFS e NFS, e l'accesso ai dati a livello di blocco attraverso protocolli SAN, come iSCSI, FC/FCoE e NVMe. Le SVM possono fornire dati ai client SAN e NAS in modo indipendente e con S3.



Nel mondo vSphere, questo approccio potrebbe anche significare un sistema unificato per l'infrastruttura di desktop virtuale (VDI) insieme all'infrastruttura di server virtuale (VSI). I sistemi che eseguono ONTAP sono di solito meno costosi per VSI rispetto agli array aziendali tradizionali e allo stesso tempo dispongono di funzionalità avanzate per l'efficienza dello storage per gestire l'infrastruttura di desktop virtuale nello stesso sistema. ONTAP unifica inoltre una vasta gamma di supporti storage, da SSD a SATA, e può estenderli facilmente nel cloud. Non è necessario acquistare un flash array per le performance, un array SATA per gli archivi e sistemi separati per il cloud. ONTAP li lega tutti insieme.

NOTA: per ulteriori informazioni sulle SVM, sullo storage unificato e sull'accesso dei client, vedere "Virtualizzazione dello storage" Nel centro di documentazione di ONTAP 9.

Strumenti di virtualizzazione per ONTAP

NetApp fornisce diversi strumenti software standalone compatibili con i sistemi tradizionali ONTAP e ASA, integrando vSphere per gestire in modo efficace il tuo ambiente virtualizzato.

I seguenti strumenti sono inclusi nella licenza ONTAP One senza costi aggiuntivi. Vedere la Figura 1 per un'illustrazione del funzionamento di questi strumenti nell'ambiente vSphere.

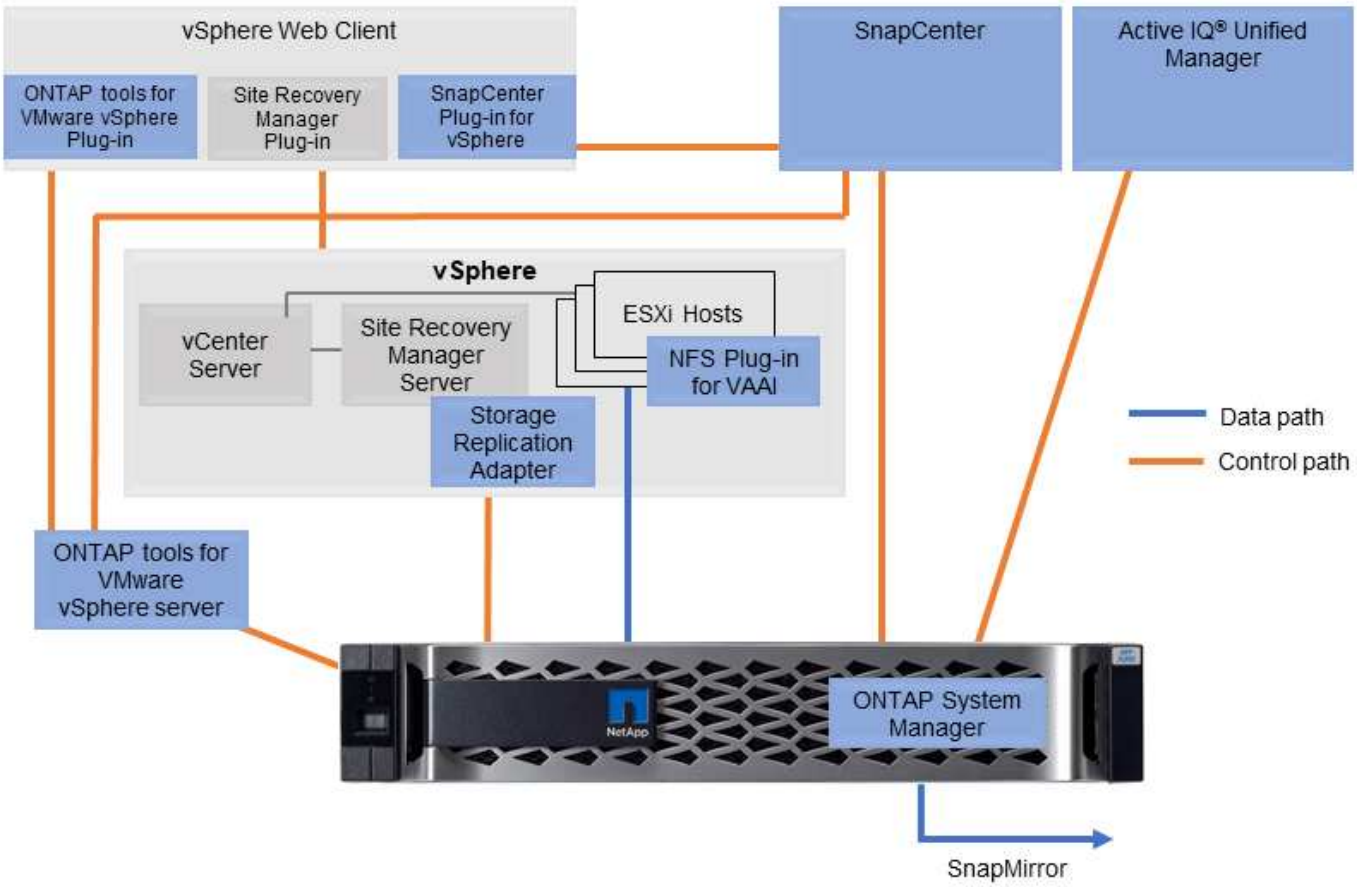
Strumenti ONTAP per VMware vSphere

"Strumenti ONTAP per VMware vSphere" È un set di tool per l'utilizzo dello storage ONTAP insieme a

vSphere. Il plug-in vCenter, precedentemente noto come Virtual Storage Console (VSC), semplifica le funzionalità di gestione ed efficienza dello storage, migliora la disponibilità e riduce i costi di storage e l'overhead operativo, sia che si utilizzi SAN che NAS. Utilizza le Best practice per il provisioning degli archivi dati e ottimizza le impostazioni degli host ESXi per gli ambienti di storage a blocchi e NFS. Per tutti questi vantaggi, NetApp consiglia di utilizzare questi strumenti ONTAP come Best practice quando si utilizza vSphere con sistemi che eseguono ONTAP. Include un'appliance server, estensioni dell'interfaccia utente per vCenter, VASA Provider e Storage Replication Adapter. Quasi tutto ciò che è contenuto negli strumenti ONTAP può essere automatizzato utilizzando semplici API REST, utilizzabili dalla maggior parte dei moderni strumenti di automazione.

- **Estensioni dell'interfaccia utente vCenter.** Le estensioni dell'interfaccia utente dei tool di ONTAP semplificano il lavoro dei team operativi e degli amministratori vCenter incorporando menu sensibili al contesto e di facile utilizzo per la gestione di host e storage, portlet informativi e funzionalità di alerting native direttamente nell'interfaccia utente di vCenter per workflow ottimizzati.
- **Provider VASA per ONTAP.** il provider VASA per ONTAP supporta il framework VMware vStorage API for Storage Awareness (VASA). Viene fornito come parte dei tool ONTAP per VMware vSphere come singola appliance virtuale per una maggiore facilità di implementazione. IL provider VASA connette vCenter Server a ONTAP per facilitare il provisioning e il monitoraggio dello storage delle macchine virtuali. Consente il supporto di VMware Virtual Volumes (vVol), la gestione dei profili di capacità dello storage e delle performance di VM vVol individuali e gli allarmi per il monitoraggio della capacità e della conformità con i profili.
- **Scheda replica archiviazione.** L'SRA viene utilizzato insieme a VMware Live Site Recovery (VLSR)/Site Recovery Manager (SRM) per gestire la replica dei dati tra i siti di produzione e di disaster recovery utilizzando SnapMirror per la replica basata su array. Può automatizzare l'attività di failover in caso di disastro e può aiutare a testare le repliche di DR senza interruzioni, per garantire un'fiducia nella tua soluzione di DR.

La figura seguente mostra gli strumenti ONTAP per vSphere.



Plug-in SnapCenter per VMware vSphere

https://mysupport.netapp.com/site/products/all/details/scv/docs-tab["Plug-in SnapCenter per VMware vSphere"]È un plug-in per vCenter Server che ti consente di gestire backup e ripristini di macchine virtuali (VM) e datastore. Fornisce una singola interfaccia per gestire backup, ripristini e cloni di VM e datastore in più sistemi ONTAP. SnapCenter supporta la replica e il recovery da siti secondari utilizzando SnapMirror. Le versioni più recenti supportano anche SnapMirror to cloud (S3), snapshot antimanomissione, SnapLock e sincronizzazione attiva SnapMirror. Il plug-in SnapCenter per VMware vSphere può essere integrato con i plug-in delle applicazioni SnapCenter per garantire backup coerenti con l'applicazione.

Plug-in NFS per VMware VAAI

```
https://mysupport.netapp.com/site/products/all/details/nfsplugin-vmware-
vaai/about-tab["Plug-in NFS NetApp per VMware VAAI"]È un plug-in per gli
host ESXi che consente loro di utilizzare le funzionalità VAAI con i
datastore NFS su ONTAP. Supporta l'offload delle copie per le operazioni
di cloning, lo space reservation per i file di dischi virtuali con thick
provisioning e l'offload delle snapshot. L'offload delle operazioni di
copia sullo storage non è necessariamente più veloce da completare, ma
riduce i requisiti di larghezza di banda della rete e scarica le risorse
host come cicli CPU, buffer e code. È possibile utilizzare i tool ONTAP
per VMware vSphere per installare il plug-in sugli host ESXi o, se
supportato, vSphere Lifecycle Manager (vLCM).
```

Opzioni software premium

I seguenti prodotti software premium sono disponibili presso NetApp. Non sono inclusi nella licenza di ONTAP One e devono essere acquistati separatamente. * "Disaster recovery BlueXP " Per VMware vSphere. Servizio basato sul cloud che offre disaster recovery e backup per gli ambienti VMware. Può essere utilizzato con o senza SnapCenter e supporta il DR on-premise o on-premise utilizzando SAN o NAS, e on-premise verso/da cloud utilizzando NFS laddove supportato. * "Informazioni sull'infrastruttura dati (DII)". Si tratta di un servizio basato su cloud che offre monitoring e analytics per gli ambienti VMware. Supporta altri vendor di soluzioni storage in ambienti di storage eterogenei, così come più vendor di switch e altri hypervisor. DII fornisce informazioni complete end-to-end su prestazioni, capacità e stato dell'ambiente VMware.

Virtual Volumes (vVol) e Storage Policy Based Management (SPBM)

Annunciato per la prima volta nel 2012, NetApp è stato un primo partner di progettazione di VMware nello sviluppo di VMware vSphere APIs for Storage Awareness (VASA), le fondamenta della gestione basata su criteri storage (SPBM, Storage Policy Based Management) con array storage Enterprise. Questo approccio offriva una gestione granulare dello storage delle macchine virtuali limitata allo storage VMFS e NFS.

In qualità di partner di progettazione tecnologica, NetApp ha fornito un input architetturale e nel 2015 ha annunciato il supporto per vVol. Questa nuova tecnologia ha ora consentito l'automazione del provisioning dello storage granulare delle macchine virtuali e realmente nativo degli array tramite SPBM.

Volumi virtuali (vVol)

I vVol sono una rivoluzionaria architettura di storage che consente la gestione granulare dello storage delle macchine virtuali, consentendo la gestione dello storage non solo in base alle macchine virtuali (compresi i metadati delle macchine virtuali) ma anche in base ai VMDK. I vVol sono un componente chiave della strategia SDDC (Software Defined Data Center) che costituisce la base di VMware Cloud Foundation (VCF), fornendo un'architettura di storage più efficiente e scalabile per gli ambienti virtualizzati.

I vVol consentono alle macchine virtuali di utilizzare lo storage per ogni macchina virtuale, perché ogni oggetto storage delle macchine virtuali è un'entità univoca in NetApp ONTAP. Con i sistemi ASA R2, che non richiedono più la gestione dei volumi, questo significa che ogni oggetto storage delle macchine virtuali è un'unica unità di storage (su) sull'array e può essere controllato in modo indipendente. Ciò consente la creazione di policy di storage che possono essere applicate a singole macchine virtuali o VMDK (e quindi include un SUS duale), fornendo un controllo granulare sui servizi storage quali performance, disponibilità e

protezione dei dati.

Gestione basata su criteri storage (SPBM)

SPBM fornisce un framework che funge da layer di astrazione tra i servizi di storage disponibili per l'ambiente di virtualizzazione e gli elementi di storage sottoposti a provisioning tramite policy. Questo approccio consente agli storage architect di progettare pool di storage con funzionalità differenti. Questi pool possono essere facilmente utilizzati dagli amministratori VM. Gli amministratori possono quindi abbinare i requisiti dei carichi di lavoro delle macchine virtuali ai pool di storage di cui è stato eseguito il provisioning. Questo approccio semplifica la gestione dello storage e permette un utilizzo più efficiente delle risorse di storage.

SPBM è un componente chiave di vVol, che fornisce un framework basato su criteri per la gestione dei servizi storage. Le policy vengono create dagli amministratori di vSphere utilizzando regole e funzionalità esposte dal provider VASA (VP) del vendor. È possibile creare policy per diversi servizi di storage, quali performance, disponibilità e protezione dei dati. È possibile assegnare le policy a singole macchine virtuali o VMDK per un controllo granulare sui servizi storage.

NetApp ONTAP e vVol

NetApp ONTAP è leader nel settore dello storage nella scalabilità dei vVol, supportando centinaia di migliaia di vVol in un singolo cluster*. Al contrario, gli array Enterprise e i vendor di flash array più piccoli supportano fino a diverse migliaia di vVol per array. ONTAP offre una soluzione storage scalabile ed efficiente per ambienti VMware vSphere, supportando i vVol con un ricco set di servizi storage, tra cui deduplica dei dati, compressione, thin provisioning e protezione dei dati. SPBM consente un'integrazione perfetta con gli ambienti VMware vSphere.

In precedenza abbiamo indicato agli amministratori delle macchine virtuali la possibilità di consumare capacità come pool di storage. Ciò avviene mediante l'utilizzo di container di storage rappresentati in vSphere come datastore logici.

I container storage vengono creati dagli amministratori dello storage e utilizzati per raggruppare le risorse storage che possono essere consumate dagli amministratori delle macchine virtuali. I container storage possono essere creati in maniera differente a seconda del tipo di sistema ONTAP che stai utilizzando. Con i cluster tradizionali di ONTAP 9, ai container viene assegnato uno o più volumi FlexVol di supporto che formano insieme il pool di storage. Con i sistemi ASA R2, l'intero cluster è il pool di storage.

Per ulteriori informazioni su VMware vSphere Virtual Volumes, SPBM e ONTAP, vedere "TR-4400: Volumi virtuali VMware vSphere con ONTAP".

*A seconda della piattaforma e del protocollo

Datastore e protocolli

Panoramica delle funzionalità del datastore e del protocollo di vSphere

Per collegare VMware vSphere ai datastore su un sistema che esegue ONTAP sono utilizzati sei protocolli:

- FCP
- NVMe/FC
- NVMe/TCP
- ISCSI

- NFS v3
- NFS v4,1

FCP, NVMe/FC, NVMe/TCP e iSCSI sono protocolli a blocchi che utilizzano il VMFS (vSphere Virtual Machine file System) per memorizzare le VM nei LUN di ONTAP o negli namespace NVMe contenuti in un ONTAP FlexVol volume. NFS è un protocollo di file che inserisce le macchine virtuali in datastore (che sono semplicemente volumi ONTAP) senza la necessità di VMFS. SMB (CIFS), iSCSI, NVMe/TCP o NFS possono essere utilizzati anche direttamente da un sistema operativo guest a ONTAP.

Le tabelle seguenti presentano le funzionalità tradizionali del datastore supportate da vSphere con ONTAP. Queste informazioni non si applicano agli archivi dati vVol, ma in genere si applicano a vSphere 6.x e alle versioni successive che utilizzano le versioni supportate di ONTAP. È inoltre possibile consultare il "Strumento VMware Configuration Maximums" per le release specifiche di vSphere per confermare i limiti specifici.

Funzionalità	FC	ISCSI	NVMe-of	NFS
Formato	VMFS o RDM (raw device mapping)	VMFS o RDM	VMFS	n/a.
Numero massimo di datastore o LUN	1024 LUN per host	1024 LUN per server	256 namespeces per server	256 connessioni NFS per host (interessate da nconnect e trunking di sessione) NFS predefinito. MaxVolumes è 8. Utilizza i tool ONTAP per VMware vSphere per aumentare fino a 256.
Dimensione massima datastore	64 TB	64 TB	64 TB	300 TB di volume FlexVol o superiore con volume FlexGroup
Dimensione massima del file del datastore	62 TB	62 TB	62 TB	62TB con ONTAP 9.12.1P2 e versioni successive
Profondità ottimale della coda per LUN o file system	64-256	64-256	Negoziazione automatica	Fare riferimento a NFS.MaxQueueDefelse in "Host ESXi consigliato e altre impostazioni ONTAP".

La seguente tabella elenca le funzionalità supportate relative allo storage VMware.

Capacità/funzionalità	FC	ISCSI	NVMe-of	NFS
VMotion	Sì	Sì	Sì	Sì
Storage vMotion	Sì	Sì	Sì	Sì

Capacità/funzionalità	FC	ISCSI	NVMe-of	NFS
VMware ha	Sì	Sì	Sì	Sì
SDR (Storage Distributed Resource Scheduler)	Sì	Sì	Sì	Sì
Software di backup abilitato per VADP (VMware vStorage APIs for Data Protection)	Sì	Sì	Sì	Sì
Microsoft Cluster Service (MSCS) o clustering di failover all'interno di una macchina virtuale	Sì	Sì [1]	Sì [1]	Non supportato
Tolleranza agli errori	Sì	Sì	Sì	Sì
Gestione ripristino sito live/ripristino sito	Sì	Sì	No [2]	V3 solo [2]
Macchine virtuali con thin provisioning (dischi virtuali)	Sì	Sì	Sì	Sì Si tratta dell'impostazione predefinita per tutte le macchine virtuali su NFS quando non si utilizza VAAI.
Multipathing nativo di VMware	Sì	Sì	Sì	Il trunking di sessione NFS v4,1 richiede ONTAP 9.14.1 e versioni successive

La tabella seguente elenca le funzionalità di gestione dello storage ONTAP supportate.

Funzionalità	FC	ISCSI	NVMe-of	NFS
Deduplica dei dati	Risparmi nell'array	Risparmi nell'array	Risparmi nell'array	Risparmi nel datastore
Thin provisioning	Datastore o RDM	Datastore o RDM	Datastore	Datastore
Ridimensiona datastore	Crescere solo	Crescere solo	Crescere solo	Crescita, crescita automatica e riduzione
Plug-in SnapCenter per applicazioni Windows e Linux (in guest)	Sì	Sì	Sì	Sì

Funzionalità	FC	ISCSI	NVMe-of	NFS
Monitoraggio e configurazione dell'host con gli strumenti ONTAP per VMware vSphere	Sì	Sì	Sì	Sì
Provisioning con gli strumenti ONTAP per VMware vSphere	Sì	Sì	Sì	Sì

La tabella seguente elenca le funzionalità di backup supportate.

Funzionalità	FC	ISCSI	NVMe-of	NFS
Snapshot ONTAP	Sì	Sì	Sì	Sì
SRM supportato da backup replicati	Sì	Sì	No [2]	V3 solo [2]
Volume SnapMirror	Sì	Sì	Sì	Sì
Accesso all'immagine VMDK	Software di backup compatibile con SnapCenter e VADP	Software di backup compatibile con SnapCenter e VADP	Software di backup compatibile con SnapCenter e VADP	Software di backup abilitato per SnapCenter e VADP, client vSphere e browser del datastore del client web vSphere
Accesso a livello di file VMDK	Software di backup compatibile con SnapCenter e VADP, solo Windows	Software di backup compatibile con SnapCenter e VADP, solo Windows	Software di backup compatibile con SnapCenter e VADP, solo Windows	Software di backup SnapCenter e VADP e applicazioni di terze parti
Granularità NDMP	Datastore	Datastore	Datastore	Datastore o macchina virtuale

[1] **NetApp consiglia** l'utilizzo di iSCSI in-guest per i cluster Microsoft piuttosto che VMDK abilitati per il multi-writer in un datastore VMFS. Questo approccio è pienamente supportato da Microsoft e VMware, offre una grande flessibilità con ONTAP (da SnapMirror ai sistemi ONTAP on-premise o nel cloud), è semplice da configurare e automatizzare e può essere protetto con SnapCenter. VSphere 7 aggiunge una nuova opzione VMDK in cluster. Si tratta di una configurazione diversa da VMDK abilitati per multi-writer, che richiede un datastore VMFS 6 con supporto VMDK per cluster abilitato. Sono previste altre restrizioni. Per le linee guida sulla configurazione, consultare la documentazione di VMware"Configurazione per il clustering di failover di Windows Server".

[2] i datastore che utilizzano NVMe-of e NFS v4,1 richiedono la replica vSphere. La replica basata su array per NFS v4,1 non è attualmente supportata da SRM. La replica basata su array con NVMe-of non è attualmente supportata dai tool ONTAP per VMware vSphere Storage Replication Adapter (SRA).

Selezione di un protocollo di storage

I sistemi che eseguono ONTAP supportano tutti i principali protocolli di storage, consentendo ai clienti di scegliere il miglior ambiente per il proprio ambiente, a seconda dell'infrastruttura di rete esistente e pianificata e delle competenze dello staff. In passato, i test di NetApp hanno generalmente mostrato una piccola differenza tra i protocolli in esecuzione a velocità di linea simili e il numero di connessioni. Tuttavia, la tecnologia NVMe-of (NVMe/TCP e NVMe/FC) mostra guadagni notevoli in termini di IOPS, riduzione della latenza e riduzione fino al 50% o più del consumo della CPU host da parte dell'io storage. Dall'altra parte dello spettro, NFS offre la massima flessibilità e facilità di gestione, in particolare per un gran numero di macchine virtuali. Tutti questi protocolli possono essere utilizzati e gestiti con i tool ONTAP per VMware vSphere, che offre una semplice interfaccia per creare e gestire datastore.

I seguenti fattori potrebbero essere utili per valutare una scelta di protocollo:

- **Ambiente operativo corrente.** Sebbene i team IT siano generalmente esperti nella gestione dell'infrastruttura IP Ethernet, non tutti sono esperti nella gestione di un fabric FC SAN. Tuttavia, l'utilizzo di una rete IP generica non progettata per il traffico di storage potrebbe non funzionare bene. Prendi in considerazione l'infrastruttura di rete in uso, gli eventuali miglioramenti pianificati e le competenze e la disponibilità del personale per gestirli.
- **Facilità di configurazione.** oltre alla configurazione iniziale del fabric FC (switch e cablaggio aggiuntivi, zoning e verifica dell'interoperabilità di HBA e firmware), i protocolli a blocchi richiedono anche la creazione e la mappatura di LUN e il rilevamento e la formattazione da parte del sistema operativo guest. Una volta creati ed esportati, i volumi NFS vengono montati dall'host ESXi e pronti all'uso. NFS non dispone di specifiche qualifiche hardware o firmware da gestire.
- **Facilità di gestione.** Con i protocolli SAN, se è necessario più spazio, è necessario eseguire diverse operazioni, tra cui la crescita di un LUN, la ripetizione della scansione per rilevare le nuove dimensioni e l'espansione del file system. Sebbene sia possibile espandere un LUN, non lo è la dimensione di un LUN. NFS consente un facile dimensionamento in alto o in basso e questo ridimensionamento può essere automatizzato dal sistema storage. La SAN offre il recupero dello spazio tramite i comandi guest OS UNSPAKE/TRIM/UNMAP, consentendo di restituire all'array spazio dai file eliminati. Questo tipo di recupero di spazio non è possibile difficile con gli archivi dati NFS.
- **Trasparenza dello spazio di storage.** l'utilizzo dello storage è in genere più semplice da visualizzare negli ambienti NFS perché il thin provisioning restituisce immediatamente risparmi. Allo stesso modo, i risparmi di deduplica e clonazione sono immediatamente disponibili per altre macchine virtuali nello stesso datastore o per altri volumi di sistemi storage. La densità delle macchine virtuali è in genere maggiore anche in un datastore NFS, che può migliorare i risparmi della deduplica e ridurre i costi di gestione grazie a un numero inferiore di datastore da gestire.

Layout del datastore

I sistemi storage ONTAP offrono una grande flessibilità nella creazione di datastore per macchine virtuali e dischi virtuali. Sebbene vengano applicate molte Best practice ONTAP quando si utilizzano gli strumenti ONTAP per il provisioning dei datastore per vSphere (elencati nella sezione "Host ESXi consigliato e altre impostazioni ONTAP"), di seguito sono riportate alcune linee guida aggiuntive da prendere in considerazione:

- L'implementazione di vSphere con datastore NFS di ONTAP offre un'implementazione facile da gestire e dalle performance elevate che offre rapporti VM-datastore che non possono essere ottenuti con protocolli di storage basati su blocchi. Questa architettura può comportare un aumento di dieci volte della densità degli archivi dati con una conseguente riduzione del numero di archivi dati. Anche se un datastore più ampio può trarre vantaggio dall'efficienza dello storage e offrire vantaggi operativi, prendi in considerazione l'utilizzo di almeno quattro datastore (volumi FlexVol) per nodo per memorizzare le macchine virtuali su un singolo controller del ONTAP, in modo da ottenere le massime performance dalle risorse hardware. Questo approccio consente inoltre di stabilire datastore con policy di recovery diverse. Alcuni possono essere sottoposti a backup o replicati più frequentemente rispetto ad altri in base alle esigenze aziendali. I volumi

FlexGroup non richiedono più datastore per le performance, in quanto sono scalabili in base alla progettazione.

- **NetApp consiglia** l'utilizzo di volumi FlexVol per la maggior parte dei datastore NFS. A partire da ONTAP 9,8, l'utilizzo dei volumi FlexGroup è supportato anche come datastore e generalmente è consigliato per alcuni casi d'utilizzo. Gli altri container di storage ONTAP, come i qtree, non sono generalmente consigliati, in quanto al momento non sono supportati dai tool ONTAP per VMware vSphere o dal plug-in NetApp SnapCenter per VMware vSphere.
- Una buona dimensione per un datastore di volumi FlexVol è di circa 4TB - 8TB. Queste dimensioni rappresentano un buon punto di equilibrio per le performance, la facilità di gestione e la protezione dei dati. Inizia in piccolo (ad esempio, 4 TB) e fai crescere il datastore in base alle necessità (fino a un massimo di 300 TB). I datastore più piccoli sono più veloci da ripristinare dal backup o dopo un disastro e possono essere spostati rapidamente nel cluster. Prendere in considerazione l'utilizzo della funzione di dimensionamento automatico di ONTAP per aumentare e ridurre automaticamente il volume in base alle modifiche dello spazio utilizzato. I tool ONTAP per la procedura guidata di provisioning del datastore di VMware vSphere utilizzano il dimensionamento automatico per impostazione predefinita per i nuovi datastore. È possibile personalizzare ulteriormente le soglie di aumento e riduzione e le dimensioni massime e minime con System Manager o la riga di comando.
- In alternativa, i datastore VMFS possono essere configurati con LUN o namespace NVMe (chiamati anche unità storage nei nuovi sistemi ASA) accessibili tramite FC, iSCSI, NVMe/FC o NVMe/TCP. VMFS consente l'accesso simultaneo ai datastore da parte di ogni server ESX in un cluster. Gli archivi di dati VMFS possono avere dimensioni fino a 64 TB e sono costituiti da un massimo di 32 LUN da 2 TB (VMFS 3) o un singolo LUN da 64 TB (VMFS 5). Le dimensioni massime del LUN del ONTAP sono di 128TB GB su sistemi AFF, ASA e FAS. NetApp consiglia sempre di utilizzare una singola LUN di grandi dimensioni per ciascun datastore, invece di provare a utilizzare estensioni. Come per NFS, prendere in considerazione l'utilizzo di datastore multipli (volumi o unità storage) per massimizzare le performance su un singolo controller del ONTAP.
- I sistemi operativi guest precedenti necessitavano di un allineamento con il sistema storage per ottenere le migliori performance ed efficienza dello storage. Tuttavia, i moderni sistemi operativi supportati dai vendor dei distributori Microsoft e Linux come Red Hat non richiedono più modifiche per allineare la partizione del file system con i blocchi del sistema storage sottostante in un ambiente virtuale. Se stai utilizzando un vecchio sistema operativo che potrebbe richiedere un allineamento, cerca nella Knowledge base di supporto NetApp gli articoli che utilizzano "allineamento VM" o richiedi una copia del documento TR-3747 a un contatto commerciale o di un partner NetApp.
- Evitare l'uso di utilità di deframmentazione all'interno del sistema operativo guest, poiché ciò non offre vantaggi in termini di prestazioni e influisce sull'efficienza dello storage e sull'utilizzo dello spazio snapshot. È inoltre consigliabile disattivare l'indicizzazione della ricerca nel sistema operativo guest per i desktop virtuali.
- ONTAP ha guidato il settore con innovative funzionalità di efficienza dello storage, che ti consentono di sfruttare al massimo lo spazio su disco utilizzabile. I sistemi AFF aumentano ulteriormente questa efficienza con la deduplica e la compressione inline predefinite. I dati vengono deduplicati in tutti i volumi in un aggregato, quindi non è più necessario raggruppare sistemi operativi simili e applicazioni simili in un singolo datastore per massimizzare i risparmi.
- In alcuni casi, potrebbe non essere necessario un datastore. Considerare i file system guest-owned come NFS, SMB, NVMe/TCP o iSCSI gestiti dal guest. Per indicazioni specifiche sulle applicazioni, consulta i report tecnici NetApp relativi alla tua applicazione. Ad esempio, "Database Oracle su ONTAP" ha una sezione sulla virtualizzazione con informazioni utili.
- I dischi di prima classe (o dischi virtuali migliorati) consentono dischi gestiti da vCenter indipendenti da una macchina virtuale con vSphere 6.5 e versioni successive. Anche se gestiti principalmente da API, possono essere utili con vVol, soprattutto se gestiti da OpenStack o Kubernetes tools. Sono supportati da ONTAP e dai tool ONTAP per VMware vSphere.

Migrazione di datastore e macchine virtuali

Quando si esegue la migrazione delle macchine virtuali da un datastore esistente su un altro sistema storage a ONTAP, è necessario tenere presente alcune procedure:

- Utilizzare Storage vMotion per spostare la maggior parte delle macchine virtuali su ONTAP. Questo approccio non solo non è disgregativo per l'esecuzione di macchine virtuali, ma consente anche funzionalità di efficienza dello storage ONTAP come la deduplica inline e la compressione per elaborare i dati durante la migrazione. Prendere in considerazione l'utilizzo delle funzionalità di vCenter per selezionare più macchine virtuali dall'elenco di inventario e quindi pianificare la migrazione (utilizzare il tasto Ctrl mentre si fa clic su azioni) in un momento appropriato.
- Sebbene sia possibile pianificare con attenzione una migrazione verso datastore di destinazione appropriati, spesso è più semplice eseguire la migrazione in blocco e poi organizzarla in un secondo momento. Potresti voler utilizzare questo approccio per guidare la migrazione verso datastore diversi, se hai esigenze specifiche di data Protection, come ad esempio diverse pianificazioni Snapshot. Inoltre, una volta che le VM sono sul cluster NetApp, storage vMotion può sfruttare gli offload VAAI per spostare le VM tra datastore nel cluster, senza richiedere una copia basata su host. Tuttavia, NFS non consente di scaricare lo storage vMotion di macchine virtuali alimentate, ma VMFS.
- Le macchine virtuali che richiedono una migrazione più accurata includono database e applicazioni che utilizzano lo storage collegato. In generale, considerare l'utilizzo degli strumenti dell'applicazione per gestire la migrazione. Per Oracle, prendere in considerazione l'utilizzo di strumenti Oracle come RMAN o ASM per migrare i file di database. Per ulteriori informazioni, vedere "Migrazione dei database Oracle sui sistemi di storage ONTAP" . Allo stesso modo, per SQL Server, prendere in considerazione l'utilizzo di SQL Server Management Studio o di strumenti NetApp come SnapManager per SQL Server o SnapCenter.

Strumenti ONTAP per VMware vSphere

La Best practice più importante quando si utilizza vSphere con sistemi che eseguono ONTAP è quella di installare e utilizzare i tool ONTAP per il plug-in VMware vSphere (precedentemente noto come Virtual Storage Console). Questo plug-in vCenter semplifica la gestione dello storage, aumenta la disponibilità e riduce i costi dello storage e l'overhead operativo, sia che si utilizzino SAN o NAS, su ASA, AFF, FAS o persino ONTAP Select (una versione software-defined ONTAP che viene eseguita in una macchina virtuale VMware o KVM). Utilizza le Best practice per il provisioning degli archivi di dati e ottimizza le impostazioni degli host ESXi per i timeout multipath e HBA (descritti nell'Appendice B). Poiché si tratta di un plug-in vCenter, è disponibile per tutti i client web vSphere che si connettono al server vCenter.

Il plug-in consente inoltre di utilizzare altri strumenti ONTAP in ambienti vSphere. Il prodotto consente di installare il plug-in NFS per VMware VAAI, che consente l'offload delle copie in ONTAP per le operazioni di cloning delle macchine virtuali, lo space reservation per i file di dischi virtuali con thick provisioning e l'offload delle snapshot ONTAP.

> Nei cluster vSphere basati su immagini, sarà comunque necessario aggiungere il plug-in NFS all'immagine in modo che non si discostino dalla conformità quando viene installato con gli strumenti ONTAP.

I tool ONTAP sono anche l'interfaccia di gestione per molte funzioni del provider VASA per ONTAP, supportando una gestione basata su policy di storage con vVol.

In generale, **NetApp consiglia** di utilizzare gli strumenti ONTAP per l'interfaccia di VMware vSphere all'interno di vCenter per effettuare il provisioning dei datastore tradizionali e vVol per assicurarsi che vengano seguite le Best practice.

Rete generale

La configurazione delle impostazioni di rete quando si utilizza vSphere con sistemi che eseguono ONTAP è semplice e simile ad altre configurazioni di rete. Ecco alcuni aspetti da considerare:

- Separare il traffico di rete dello storage dalle altre reti. È possibile ottenere una rete separata utilizzando una VLAN dedicata o switch separati per lo storage. Se la rete di storage condivide percorsi fisici come gli uplink, potrebbe essere necessario QoS o porte di uplink aggiuntive per garantire una larghezza di banda sufficiente. Non connettere gli host direttamente allo storage; utilizzare gli switch per disporre di percorsi ridondanti e consentire a VMware ha di funzionare senza alcun intervento. Vedere "Connessione di rete diretta" per ulteriori informazioni.
- I frame jumbo possono essere utilizzati se lo si desidera e supportati dalla rete, in particolare quando si utilizza iSCSI. Se vengono utilizzati, assicurarsi che siano configurati in modo identico su tutti i dispositivi di rete, VLAN e così via nel percorso tra lo storage e l'host ESXi. In caso contrario, potrebbero verificarsi problemi di connessione o di prestazioni. La MTU deve essere impostata in modo identico anche sullo switch virtuale ESXi, sulla porta VMkernel e anche sulle porte fisiche o sui gruppi di interfacce di ciascun nodo ONTAP.
- NetApp consiglia di disattivare solo il controllo di flusso di rete sulle porte di cluster Interconnect in un cluster ONTAP. NetApp non fornisce altri consigli sulle Best practice per le restanti porte di rete utilizzate per il traffico dati. Attivare o disattivare secondo necessità. Vedere "TR-4182" per ulteriori informazioni sul controllo di flusso.
- Quando gli array di storage ESXi e ONTAP sono connessi a reti di storage Ethernet, **NetApp consiglia** di configurare le porte Ethernet a cui questi sistemi si connettono come porte edge Rapid Spanning Tree Protocol (RSTP) o utilizzando la funzione Cisco PortFast. **NetApp consiglia** di abilitare la funzione di trunk PortFast Spanning-Tree in ambienti che utilizzano la funzione PortFast Cisco e che dispongono di trunking VLAN 802,1Q abilitato al server ESXi o agli array di storage ONTAP.
- **NetApp consiglia** le seguenti procedure consigliate per l'aggregazione dei collegamenti:
  - Utilizzare switch che supportano l'aggregazione di collegamenti di porte su due chassis switch separati utilizzando un approccio a gruppi di aggregazione di collegamenti multi-chassis, ad esempio Virtual PortChannel (VPC) di Cisco.
  - Disattivare LACP per le porte dello switch connesse a ESXi, a meno che non si utilizzi dvSwitch 5.1 o versioni successive con LACP configurato.
  - Utilizza LACP per creare aggregati di link per sistemi di storage ONTAP con gruppi di interfacce dinamiche multimode con hash porta o IP. Fare riferimento a. "Gestione della rete" per ulteriori indicazioni.
  - Utilizzare un criterio di raggruppamento hash IP su ESXi quando si utilizza l'aggregazione di collegamenti statici (ad esempio, EtherChannel) e vSwitch standard o l'aggregazione di collegamenti basata su LACP con gli switch distribuiti vSphere. Se non si utilizza l'aggregazione dei collegamenti, utilizzare invece "Route based on the origining virtual port ID" (percorso basato sull'ID della porta virtuale di origine).

SAN (FC, FCoE, NVMe/FC, iSCSI), RDM

In vSphere, esistono quattro modi per utilizzare i dispositivi di storage a blocchi:

- Con datastore VMFS
- Con RDM (raw device mapping)
- Come LUN connessa a iSCSI o namespace connesso a NVMe/TCP, accessibile e controllato da un iniziatore software da un sistema operativo guest di macchine virtuali

- Come datastore vVols

VMFS è un file system in cluster dalle performance elevate che fornisce datastore che sono pool di storage condivisi. Gli archivi dati VMFS possono essere configurati con LUN accessibili tramite FC, iSCSI, FCoE o namespace NVMe accessibili tramite i protocolli NVMe/FC o NVMe/TCP. VMFS consente l'accesso simultaneo allo storage da parte di ogni server ESX in un cluster. Le dimensioni massime del LUN sono generalmente di 128TB GB a partire da ONTAP 9.12.1P2 (e versioni precedenti con i sistemi ASA); pertanto, è possibile creare un datastore VMFS 5 o 6 di 64TB GB di dimensioni massime utilizzando un singolo LUN.

Le estensioni sono un concetto di storage vSphere tramite cui è possibile "unire" diverse LUN per creare un singolo datastore più grande. Non utilizzare mai estensioni per raggiungere le dimensioni desiderate del datastore. Una singola LUN è la Best practice per un datastore VMFS.

VSphere include il supporto integrato per diversi percorsi verso i dispositivi storage. VSphere è in grado di rilevare il tipo di dispositivo storage per i sistemi storage supportati e di configurare automaticamente lo stack multipath per supportare le funzionalità del sistema storage in uso, la sovranità del protocollo utilizzato o se si utilizza ASA, AFF, FAS o Software Defined ONTAP.

Sia vSphere che ONTAP supportano l'Asymmetric Logical Unit Access (ALUA) per stabilire percorsi Active/ottimizzati e Active/non ottimizzati per Fibre Channel e iSCSI e l'Asymmetric Namespace (ANA) per gli spazi dei nomi NVMe utilizzando NVMe/FC e NVMe/TCP. In ONTAP, un percorso ALUA o ANA ottimizzato segue un percorso diretto dei dati, utilizzando una porta di destinazione sul nodo che ospita la LUN o il namespace a cui si accede. ALUA/ANA è attivato per impostazione predefinita sia in vSphere che in ONTAP. Il software multipathing di vSphere riconosce il cluster ONTAP come ALUA o ANA e utilizza il plug-in nativo appropriato con la policy di bilanciamento del carico round robin.

Con i sistemi ASA di NetApp, i LUN e gli spazi dei nomi vengono presentati agli host ESXi con percorso simmetrico. Ciò significa che tutti i percorsi sono attivi e ottimizzati. Il software multipathing di vSphere riconosce il sistema ASA come simmetrico e utilizza il plug-in nativo appropriato con la policy di bilanciamento del carico round robin.

Per le impostazioni di multipathing ottimizzate, consultare la sezione"Host ESXi consigliato e altre impostazioni ONTAP".

ESXi non vede LUN, namespace o percorsi oltre i propri limiti. In un cluster ONTAP più grande, è possibile raggiungere il limite di percorso prima del limite di LUN. Per risolvere questo limite, ONTAP supporta la mappa LUN selettiva (SLM) nella versione 8.3 e successive.

Fare riferimento alla "Strumento VMware Configuration Maximums" per i limiti supportati più aggiornati in ESXi.

SLM limita i nodi che pubblicizzano i percorsi a una determinata LUN. È consigliabile utilizzare NetApp per almeno due LIF per nodo per SVM e SLM per limitare i percorsi pubblicizzati al nodo che ospita la LUN e il partner ha. Sebbene esistano altri percorsi, essi non vengono pubblicizzati per impostazione predefinita. È possibile modificare i percorsi pubblicizzati con gli argomenti del nodo di reporting add e remove all'interno di SLM. Si noti che i LUN creati nelle release precedenti alla 8,3 pubblicizzano tutti i percorsi e devono essere modificati solo per pubblicizzare i percorsi alla coppia ha di hosting. Per ulteriori informazioni su SLM, vedere la sezione 5,9 di "TR-4080". Il precedente metodo di portset può essere utilizzato anche per ridurre ulteriormente i percorsi disponibili per un LUN. I portset aiutano a ridurre il numero di percorsi visibili attraverso i quali gli iniziatori in un igroup possono vedere le LUN.

- SLM è attivato per impostazione predefinita. A meno che non si utilizzino portset, non è necessaria alcuna

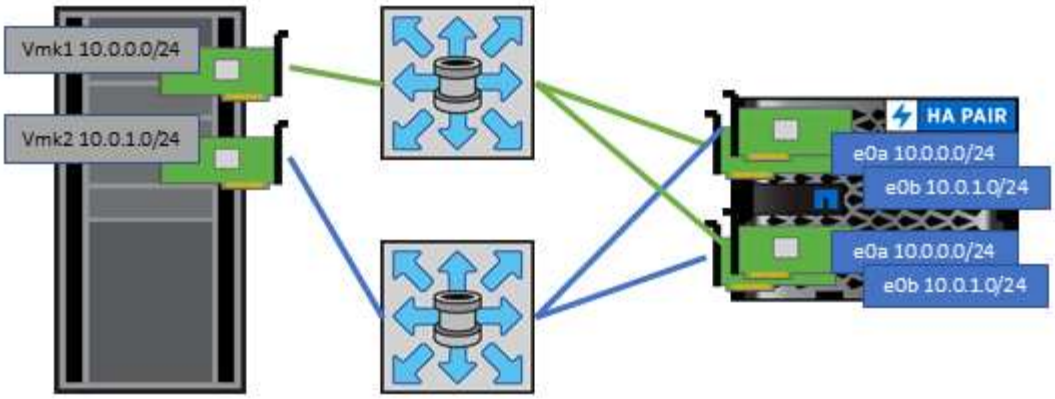
configurazione aggiuntiva.

- Per le LUN create prima di Data ONTAP 8,3, applicare manualmente SLM eseguendo il comando per rimuovere i nodi di reporting LUN e limitare l'accesso LUN `lun mapping remove-reporting-nodes` al nodo proprietario LUN e al partner ha.

I protocolli a blocchi basati su SCSI (iSCSI, FC e FCoE) accedono ai LUN usando ID LUN, numeri di serie e nomi univoci. FC e FCoE utilizzano nomi mondiali (WWN e WWPN) e iSCSI utilizza nomi qualificati iSCSI (IQN) per stabilire percorsi basati su LUN per le mappature igroup filtrate da portset e SLM. I protocolli a blocchi basati su NVMe vengono gestiti assegnando un namespace con un ID namespace generato automaticamente a un sottosistema NVMe e mappando tale sottosistema al NVMe Qualified Name (NQN) degli host. Indipendentemente da FC o TCP, i namespace NVMe vengono mappati utilizzando l'NQN e non il WWPN o il WWNN. L'host crea quindi un controller definito dal software per il sottosistema mappato per accedere ai propri spazi dei nomi. Il percorso verso LUN e namespace all'interno di ONTAP non ha alcun senso per i protocolli a blocchi e non viene presentato in alcun modo nel protocollo. Pertanto, un volume che contiene solo LUN non deve essere montato internamente e non è necessario un percorso di giunzione per i volumi che contengono LUN utilizzati negli archivi dati.

Altre Best practice da prendere in considerazione:

- Verificare "Host ESXi consigliato e altre impostazioni ONTAP"le impostazioni consigliate da NetApp in collaborazione con VMware.
- Assicurarsi che venga creata un'interfaccia logica (LIF) per ogni SVM su ciascun nodo del cluster ONTAP per garantire la massima disponibilità e mobilità. La Best practice PER LE SAN ONTAP consiste nell'utilizzare due porte fisiche e LIF per nodo, una per ciascun fabric. ALUA viene utilizzato per analizzare i percorsi e identificare i percorsi attivi ottimizzati (diretti) rispetto ai percorsi attivi non ottimizzati. ALUA viene utilizzato per FC, FCoE e iSCSI.
- Per le reti iSCSI, utilizzare più interfacce di rete VMkernel su diverse subnet di rete con raggruppamento NIC quando sono presenti più switch virtuali. È inoltre possibile utilizzare più NIC fisiche collegate a più switch fisici per fornire ha e un throughput maggiore. La figura seguente mostra un esempio di connettività multipath. In ONTAP, configurare un gruppo di interfacce single-mode per il failover con due o più collegamenti connessi a due o più switch oppure utilizzare LACP o un'altra tecnologia di aggregazione dei collegamenti con gruppi di interfacce multimodali per fornire ha e i vantaggi dell'aggregazione dei collegamenti.
- Se il protocollo CHAP (Challenge-Handshake Authentication Protocol) viene utilizzato in ESXi per l'autenticazione di destinazione, deve essere configurato anche in ONTAP utilizzando la CLI (`vserver iscsi security create`) O con System Manager (modificare Initiator Security in Storage > SVM > SVM Settings > Protocols > iSCSI).
- Utilizza i tool ONTAP per VMware vSphere per creare e gestire LUN e igroups. Il plug-in determina automaticamente le WWPN dei server e crea gli igroups appropriati. Inoltre, configura i LUN in base alle Best practice e li associa agli igroups corretti.
- Utilizzare con cautela gli RDM poiché possono essere più difficili da gestire e utilizzano anche percorsi limitati come descritto in precedenza. I LUN ONTAP supportano entrambi "modalità di compatibilità fisica e virtuale" RDM.
- Per ulteriori informazioni sull'utilizzo di NVMe/FC con vSphere 7.0, consulta questo articolo "Guida alla configurazione degli host NVMe/FC di ONTAP" e. "TR-4684"La figura seguente mostra la connettività multipath da un host vSphere a un LUN ONTAP.



NFS

ONTAP è, tra l'altro, un array NAS scale-out di livello Enterprise. ONTAP consente a VMware vSphere di accedere contemporaneamente agli archivi dati connessi a NFS da numerosi host ESXi, superando di gran lunga i limiti imposti ai file system VMFS. L'utilizzo di NFS con vSphere offre alcuni vantaggi in termini di facilità di utilizzo e di visibilità dell'efficienza dello storage, come menzionato nella "datastore"sezione.

Quando si utilizza ONTAP NFS con vSphere, si consiglia di seguire le seguenti Best practice:

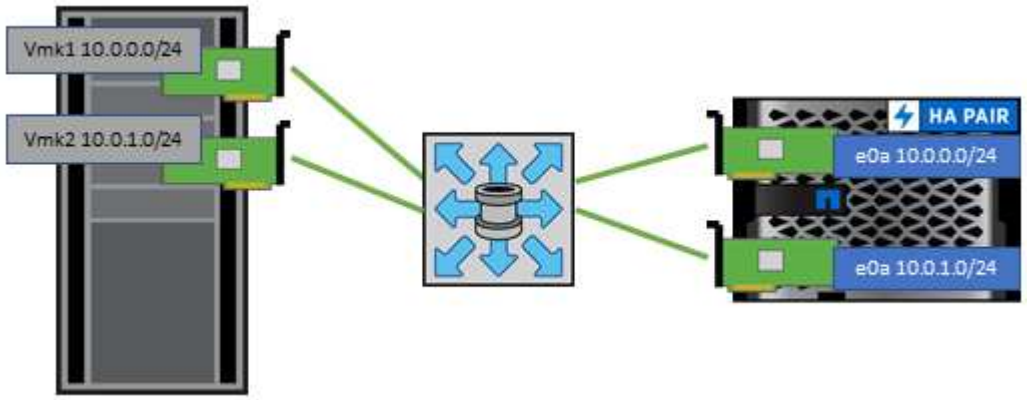
- Utilizza i tool ONTAP per VMware vSphere (la Best practice più importante):
  - Utilizza i tool di ONTAP per VMware vSphere per il provisioning dei datastore in quanto semplifica automaticamente la gestione delle policy di esportazione.
  - Quando si creano datastore per cluster VMware con il plug-in, selezionare il cluster anziché un singolo server ESX. Questa opzione attiva il montaggio automatico del datastore su tutti gli host del cluster.
  - Utilizzare la funzione di montaggio del plug-in per applicare i datastore esistenti ai nuovi server.
  - Quando non si utilizzano gli strumenti ONTAP per VMware vSphere, utilizzare una singola policy di esportazione per tutti i server o per ciascun cluster di server in cui è necessario un controllo aggiuntivo degli accessi.
- Utilizzare una singola interfaccia logica (LIF) per ogni SVM su ciascun nodo del cluster ONTAP. Le raccomandazioni precedenti di un LIF per datastore non sono più necessarie. Benché l'accesso diretto (LIF e datastore nello stesso nodo) sia migliore, non preoccuparti dell'accesso indiretto perché l'effetto sulle performance è generalmente minimo (microsecondi).
- Se si utilizza fpolicy, assicurarsi di escludere i file .lck poiché vengono utilizzati da vSphere per il blocco ogni volta che una VM viene accesa.
- Tutte le versioni di VMware vSphere attualmente supportate possono utilizzare sia NFS v3 che v4,1. Il supporto ufficiale per nconnect è stato aggiunto a vSphere 8,0 update 2 per NFS v3 e all'update 3 per NFS v4,1. Per NFS v4,1, vSphere continua a supportare il trunking della sessione, l'autenticazione Kerberos e l'autenticazione Kerberos con integrità. È importante notare che il trunking della sessione richiede ONTAP 9.14.1 o una versione successiva. È possibile ottenere ulteriori informazioni sulla funzione nconnect e sul modo in cui migliora le prestazioni a "Funzione NFSv3 nconnect con NetApp e VMware".

- Il valore massimo per nconnect in vSphere 8 è 4 e il valore predefinito è 1. Il limite massimo di valore in vSphere può essere aumentato in base all'host tramite impostazioni avanzate, tuttavia in genere non è necessario.
- Per gli ambienti che richiedono prestazioni superiori a quelle consentite da una singola connessione TCP, si consiglia di utilizzare il valore 4.
- Tenere presente che ESXi ha un limite di 256 connessioni NFS e ogni connessione nconnect conta per quel totale. Ad esempio, due datastore con nconnect=4 contano come otto connessioni totali.
- È importante verificare l'impatto delle prestazioni di nconnect sull'ambiente prima di implementare cambiamenti su larga scala negli ambienti di produzione.

- Vale la pena notare che NFSv3 e NFSv4,1 utilizzano meccanismi di bloccaggio diversi. NFSv3 utilizza il blocco lato client, mentre NFSv4,1 utilizza il blocco lato server. Anche se un volume ONTAP può essere esportato tramite entrambi i protocolli, ESXi può montare un datastore solo attraverso un protocollo. Tuttavia, ciò non significa che altri host ESXi non possano montare lo stesso datastore attraverso una versione diversa. Per evitare qualsiasi problema, è essenziale specificare la versione del protocollo da utilizzare durante il montaggio, assicurandosi che tutti gli host utilizzino la stessa versione e, quindi, lo stesso stile di blocco. È fondamentale evitare di mischiare versioni NFS tra gli host. Se possibile, utilizzare i profili host per verificare la conformità.
  - Poiché non esiste alcuna conversione automatica del datastore tra NFSv3 e NFSv4.1, creare un nuovo datastore NFSv4.1 e utilizzare Storage vMotion per migrare le macchine virtuali nel nuovo datastore.
  - Fare riferimento alle note della tabella di interoperabilità NFS v4,1 nella "Tool di matrice di interoperabilità NetApp" per i livelli di patch ESXi specifici richiesti per il supporto.
- Come menzionato in "impostazioni", se non si utilizza vSphere CSI per Kubernetes, è necessario impostare il nuovo SyncInterval per "VMware KB 386364"
- Regole delle policy di esportazione NFS vengono utilizzate per controllare l'accesso dagli host vSphere. È possibile utilizzare un criterio con più volumi (datastore). Con NFS, ESXi utilizza lo stile di sicurezza sys (UNIX) e richiede l'opzione di montaggio root per eseguire le macchine virtuali. In ONTAP, questa opzione viene definita superutente e, quando viene utilizzata l'opzione superutente, non è necessario specificare l'ID utente anonimo. Tenere presente che le regole delle policy di esportazione con valori diversi per `-anon` e `-allow-suid` possono causare problemi di rilevamento SVM con gli strumenti ONTAP. Gli indirizzi IP devono essere un elenco separato da virgole senza spazi degli indirizzi della porta vmkernel che montano gli archivi dati. Ecco un esempio di regola dei criteri:
  - Access Protocol: nfs (che include sia nfs3 che nfs4)
  - Elenco di nomi host, indirizzi IP, netgroup o domini corrispondenti ai client: 192.168.42.21,192.168.42.22
  - Regola di accesso RO: Any
  - Regola di accesso RW: Qualsiasi
  - ID utente a cui sono mappati gli utenti anonimi: 65534
  - Tipi di protezione superutente: Qualsiasi
  - Honor setuid bits in SETATTR: True
  - Consenti la creazione di dispositivi: True
- Se si utilizza il plug-in NFS NetApp per VMware VAAI, è necessario impostare il protocollo come `nfs` al momento della creazione o della modifica della regola dei criteri di esportazione. Per il funzionamento dell'offload delle copie VAAI è necessario il protocollo NFSv4, specificando che il protocollo `nfs` include automaticamente le versioni NFSv3 e NFSv4. Questa operazione è necessaria anche se il tipo di datastore

viene creato come NFS v3.

- I volumi del datastore NFS vengono svincoli dal volume root di SVM; pertanto, ESXi deve anche avere accesso al volume root per navigare e montare i volumi del datastore. La policy di esportazione per il volume root e per qualsiasi altro volume in cui la giunzione del volume del datastore è nidificata deve includere una regola o regole per i server ESXi che concedono loro l'accesso in sola lettura. Ecco un esempio di policy per il volume root, utilizzando anche il plug-in VAAI:
  - Protocollo di accesso: nfs
  - Spec. Corrispondenza client: 192.168.42.21,192.168.42.22
  - Regola di accesso RO: SIS
  - RW Access Rule: Never (miglior sicurezza per il volume root)
  - UID anonimo
  - Superutente: SYS (richiesto anche per il volume root con VAAI)
- Sebbene ONTAP offra una struttura flessibile dello spazio dei nomi dei volumi per organizzare i volumi in un albero utilizzando le giunzioni, questo approccio non ha alcun valore per vSphere. Crea una directory per ogni VM nella directory principale dell'archivio dati, indipendentemente dalla gerarchia dello spazio dei nomi dello storage. Pertanto, la Best practice consiste nel montare semplicemente il percorso di giunzione per i volumi per vSphere nel volume root della SVM, che è il modo in cui i tool ONTAP per VMware vSphere prevedono il provisioning dei datastore. La mancanza di percorsi di giunzione nidificati significa anche che nessun volume dipende da un volume diverso dal volume root e che la sua eliminazione o la sua eliminazione, anche intenzionalmente, non influisce sul percorso verso altri volumi.
- Una dimensione del blocco di 4K è adatta per le partizioni NTFS negli archivi dati NFS. La figura seguente mostra la connettività da un host vSphere a un datastore NFS ONTAP.



La seguente tabella elenca le versioni di NFS e le funzionalità supportate.

Funzionalità di vSphere	NFSv3	NFSv4,1
VMotion e Storage vMotion	Sì	Sì
Alta disponibilità	Sì	Sì
Tolleranza agli errori	Sì	Sì
DRS	Sì	Sì
Profili host	Sì	Sì
DRS dello storage	Sì	No

Funzionalità di vSphere	NFSv3	NFSv4,1
Controllo i/o dello storage	Sì	No
SRM	Sì	No
Volumi virtuali	Sì	No
Accelerazione hardware (VAAI)	Sì	Sì
Autenticazione Kerberos	No	Sì (ottimizzato con vSphere 6.5 e versioni successive per supportare AES, krb5i)
Supporto multipathing	No	Sì (ONTAP 9.14.1)

Volumi FlexGroup

Utilizza volumi ONTAP e FlexGroup con VMware vSphere per datastore semplici e scalabili che sfruttano tutta la potenza di un intero cluster ONTAP.

ONTAP 9,8, insieme ai tool ONTAP per VMware vSphere 9,8-9,13 e al plug-in SnapCenter per VMware 4,4 e versioni successive, ha aggiunto il supporto per datastore basati su volumi FlexGroup in vSphere. I volumi FlexGroup semplificano la creazione di datastore di grandi dimensioni e creano automaticamente i volumi costituenti distribuiti necessari nel cluster ONTAP, per ottenere le massime performance da un sistema ONTAP.

Utilizza i volumi FlexGroup con vSphere se desideri un singolo datastore vSphere scalabile con la potenza di un cluster ONTAP completo o se disponi di carichi di lavoro di cloning molto grandi che possono sfruttare il meccanismo di cloning FlexGroup mantenendo costantemente al caldo la cache dei cloni.

Offload delle copie

Oltre agli estesi test di sistema con i carichi di lavoro vSphere, ONTAP 9,8 ha aggiunto un nuovo meccanismo di offload delle copie per i datastore FlexGroup. Questo nuovo sistema utilizza un motore di copia migliorato per replicare i file tra i componenti in background consentendo l'accesso sia all'origine che alla destinazione. La cache locale costituente viene quindi utilizzata per creare rapidamente un'istanza dei cloni delle macchine virtuali on-demand.

Per attivare l'offload delle copie ottimizzato per FlexGroup, fare riferimento alla sezione "Come configurare i volumi ONTAP FlexGroup per consentire l'offload delle copie VAAI"

Potresti accorgerti che se utilizzi il cloning VAAI, ma non quello per mantenere calda la cache, i cloni potrebbero non essere più veloci di una copia basata su host. In questo caso, è possibile regolare il timeout della cache per soddisfare meglio le proprie esigenze.

Considerare il seguente scenario:

- Hai creato un nuovo FlexGroup con 8 componenti
- Il timeout della cache per il nuovo FlexGroup è impostato su 160 minuti

In questo scenario, i primi 8 cloni da completare saranno copie complete, non cloni di file locali. Qualsiasi clonazione aggiuntiva di tale macchina virtuale prima della scadenza del timeout di 160 secondi utilizzerà il motore di clonazione file all'interno di ciascun componente in modo round-robin per creare copie quasi immediate distribuite uniformemente tra i volumi costituenti.

Ogni nuovo processo di clonazione che un volume riceve ripristina il timeout. Se un volume costituente nel

FlexGroup di esempio non riceve una richiesta di clone prima del timeout, la cache di quella particolare VM verrà cancellata e il volume dovrà essere popolato di nuovo. Inoltre, se l'origine del clone originale cambia (ad esempio, è stato aggiornato il modello), la cache locale di ciascun componente verrà invalidata per evitare conflitti. Come indicato in precedenza, la cache può essere regolata in base alle esigenze dell'ambiente.

Per ulteriori informazioni sull'utilizzo di FlexGroup Volumes con VAAI, fai riferimento a questo articolo della KB: "VAAI: Come funziona il caching con i volumi FlexGroup?"

In ambienti in cui non è possibile sfruttare al meglio la cache FlexGroup, ma è comunque necessario un rapido cloning cross-volume, prendere in considerazione l'utilizzo di vVol. Il cloning tra volumi con vVol è molto più rapido rispetto ai datastore tradizionali, senza fare affidamento su una cache.

Impostazioni QoS

È supportata la configurazione della qualità del servizio a livello di FlexGroup utilizzando ONTAP System Manager o la shell del cluster, ma non fornisce consapevolezza delle macchine virtuali o integrazione di vCenter.

La qualità del servizio (IOPS max/min) può essere impostata su singole macchine virtuali o su tutte le macchine virtuali di un datastore in quel momento nell'interfaccia utente di vCenter o tramite API REST utilizzando i tool ONTAP. L'impostazione della QoS su tutte le macchine virtuali sostituisce le impostazioni separate per ogni macchina virtuale. Le impostazioni non si estendono alle macchine virtuali nuove o migrate in futuro; impostare la QoS sulle nuove macchine virtuali o riapplicare la QoS a tutte le macchine virtuali nel datastore.

Si noti che VMware vSphere considera tutti i/o di un datastore NFS come una singola coda per host e la limitazione della QoS su una VM può influire sulle performance delle altre VM dello stesso datastore per quell'host. Questo contrasta con i vVol, che possono mantenere le proprie impostazioni di policy di QoS se migrano in un altro datastore e non influiscono sull'io di altre macchine virtuali quando rallentano.

Metriche

ONTAP 9,8 ha inoltre aggiunto nuove metriche di performance basate su file (IOPS, throughput e latenza) per i file FlexGroup, che possono essere visualizzate nei tool ONTAP per la dashboard e i report delle macchine virtuali di VMware vSphere. Il plug-in ONTAP Tools per VMware vSphere consente inoltre di impostare le regole di qualità del servizio (QoS) utilizzando una combinazione di IOPS massimo e/o minimo. Questi possono essere impostati su tutte le macchine virtuali in un datastore o singolarmente per macchine virtuali specifiche.

Best practice

- Utilizza i tool ONTAP per creare datastore FlexGroup, per assicurarti che FlexGroup venga creato in modo ottimale e che le policy di esportazione siano configurate in modo da corrispondere al tuo ambiente vSphere. Tuttavia, dopo aver creato il volume FlexGroup con i tool ONTAP, tutti i nodi del cluster vSphere utilizzano un singolo indirizzo IP per montare il datastore. Ciò potrebbe causare un collo di bottiglia sulla porta di rete. Per evitare questo problema, smontare il datastore, quindi rimontarlo utilizzando la procedura guidata standard del datastore vSphere utilizzando un nome DNS round-robin che offre bilanciamento del carico tra le LIF della SVM. Dopo il rimontaggio, gli strumenti ONTAP saranno nuovamente in grado di gestire il datastore. Se gli strumenti ONTAP non sono disponibili, utilizzare i valori predefiniti di FlexGroup e creare il criterio di esportazione seguendo le linee guida riportate in "Datastore e protocolli: NFS".
- Quando si ridimensiona un datastore FlexGroup, tenere presente che FlexGroup è costituito da più volumi FlexVol più piccoli che creano uno spazio dei nomi più grande. Pertanto, dimensionare il datastore in modo che sia almeno 8x MB (si suppongano i 8 componenti predefiniti) delle dimensioni del file VMDK più il 10-20% di spazio inutilizzato, per garantire flessibilità nel ribilanciamento. Ad esempio, se nell'ambiente è presente un VMDK di 6TB GB, dimensionare il datastore FlexGroup non inferiore a 52,8TB GB (6x8+10%).

- VMware e NetApp supportano il trunking di sessione NFSv4,1 a partire da ONTAP 9.14.1. Per informazioni dettagliate sulle versioni specifiche, fare riferimento alle note dell'Interoperability Matrix Tool (IMT) NFS 4,1 di NetApp. NFSv3 non supporta percorsi fisici multipli a un volume ma supporta nconnect beginning in vSphere 8.0U2. Ulteriori informazioni su nconnect sono disponibili sul "Funzione NFSv3 nConnect con NetApp e VMware".
- Utilizzare il plug-in NFS per VMware VAAI per l'offload delle copie. Si noti che mentre il cloning è migliorato all'interno di un datastore FlexGroup, come menzionato in precedenza, ONTAP non offre significativi vantaggi in termini di performance rispetto alla copia dell'host ESXi quando si copiano le macchine virtuali tra volumi FlexVol e/o FlexGroup. Prendi in considerazione pertanto i carichi di lavoro di cloning al momento di decidere di utilizzare volumi VAAI o FlexGroup. La modifica del numero di volumi costituenti è un modo per ottimizzare il cloning basato su FlexGroup. Come per l'ottimizzazione del timeout della cache menzionato in precedenza.
- Utilizza i tool ONTAP per VMware vSphere 9,8-9,13 per monitorare le performance delle macchine virtuali FlexGroup utilizzando le metriche ONTAP (dashboard e report VM) e gestire la QoS sulle singole macchine virtuali. Queste metriche non sono attualmente disponibili tramite i comandi o le API ONTAP.
- Il plug-in SnapCenter per VMware vSphere versione 4,4 e successive supporta il backup e recovery delle macchine virtuali in un datastore FlexGroup nel sistema storage primario. SCV 4,6 aggiunge il supporto di SnapMirror per datastore basati su FlexGroup. L'utilizzo di snapshot e replica basate su array è il modo più efficiente per proteggere i dati.

Configurazione di rete

La configurazione delle impostazioni di rete quando si utilizza vSphere con sistemi che eseguono ONTAP è semplice e simile ad altre configurazioni di rete.

Ecco alcuni aspetti da considerare:

- Separare il traffico di rete dello storage dalle altre reti. È possibile ottenere una rete separata utilizzando una VLAN dedicata o switch separati per lo storage. Se la rete di storage condivide percorsi fisici come gli uplink, potrebbe essere necessario QoS o porte di uplink aggiuntive per garantire una larghezza di banda sufficiente. Non connettere gli host direttamente allo storage a meno che la guida alla soluzione non lo richieda specificamente; utilizzare gli switch per disporre di percorsi ridondanti e consentire a VMware ha di funzionare senza alcun intervento.
- I frame jumbo devono essere utilizzati se supportati dalla rete. Se vengono utilizzati, assicurarsi che siano configurati in modo identico su tutti i dispositivi di rete, VLAN e così via nel percorso tra lo storage e l'host ESXi. In caso contrario, potrebbero verificarsi problemi di connessione o di prestazioni. La MTU deve essere impostata in modo identico anche sullo switch virtuale ESXi, sulla porta VMkernel e anche sulle porte fisiche o sui gruppi di interfacce di ciascun nodo ONTAP.
- NetApp consiglia di disattivare solo il controllo di flusso di rete sulle porte di cluster Interconnect in un cluster ONTAP. NetApp non fornisce altre raccomandazioni per le Best practice relative al controllo di flusso per le restanti porte di rete utilizzate per il traffico dati. Se necessario, è necessario attivarlo o disattivarlo. Vedere "TR-4182" per ulteriori informazioni sul controllo di flusso.
- Quando gli array di storage ESXi e ONTAP sono collegati a reti di storage Ethernet, NetApp consiglia di configurare le porte Ethernet a cui questi sistemi si connettono come porte edge RSTP (Rapid Spanning Tree Protocol) o utilizzando la funzione PortFast di Cisco. NetApp consiglia di abilitare la funzione di trunk PortFast Spanning-Tree in ambienti che utilizzano la funzionalità Cisco PortFast e che dispongono di un trunking VLAN 802.1Q abilitato per il server ESXi o gli array di storage ONTAP.
- NetApp consiglia le seguenti Best practice per l'aggregazione dei collegamenti:
  - Utilizzare switch che supportano l'aggregazione di collegamenti di porte su due chassis switch separati utilizzando un approccio a gruppi di aggregazione di collegamenti multi-chassis, ad esempio Virtual

PortChannel (VPC) di Cisco.

- Disattivare LACP per le porte dello switch connesse a ESXi, a meno che non si utilizzi dvSwitch 5.1 o versioni successive con LACP configurato.
- Utilizzare LACP per creare aggregati di link per sistemi storage ONTAP con gruppi di interfacce multimodali dinamiche con hash IP.
- Utilizzare un criterio di raggruppamento hash IP su ESXi.

La seguente tabella fornisce un riepilogo degli elementi di configurazione di rete e indica la posizione in cui vengono applicate le impostazioni.

Elemento	ESXi	Switch	Nodo	SVM
Indirizzo IP	VMkernel	No**	No**	Sì
Aggregazione dei collegamenti	Switch virtuale	Sì	Sì	No*
VLAN	Gruppi di porte VMkernel e VM	Sì	Sì	No*
Controllo di flusso	NIC	Sì	Sì	No*
Spanning tree	No	Sì	No	No
MTU (per frame jumbo)	Switch virtuale e porta VMkernel (9000)	Sì (impostato su max)	Sì (9000)	No*
Gruppi di failover	No	No	Sì (creare)	Sì (selezionare)

*Le LIF SVM si connettono a porte, gruppi di interfacce o interfacce VLAN con VLAN, MTU e altre impostazioni. Tuttavia, le impostazioni non vengono gestite a livello di SVM.

**Questi dispositivi dispongono di indirizzi IP propri per la gestione, ma non vengono utilizzati nel contesto dello storage di rete ESXi.

SAN (FC, NVMe/FC, iSCSI, NVMe/TCP), RDM

ONTAP offre storage a blocchi di livello Enterprise per VMware vSphere utilizzando i tradizionali iSCSI e Fibre Channel Protocol (FCP) oltre al protocollo a blocchi di nuova generazione, NVMe over Fabrics (NVMe-of), ad alta efficienza e performance, con supporto per NVMe/FC e NVMe/TCP.

Per le Best practice dettagliate per l'implementazione dei protocolli a blocchi per lo storage delle macchine virtuali con vSphere e ONTAP, fare riferimento a. "Datastore e protocolli: SAN"

NFS

VSphere consente ai clienti di utilizzare array NFS di livello Enterprise per fornire l'accesso simultaneo agli archivi dati a tutti i nodi di un cluster ESXi. Come menzionato nella "datastore"sezione, quando si utilizza NFS con vSphere, esistono alcuni benefici di facilità d'uso ed efficienza dello storage.

Per le Best practice consigliate fare riferimento a. "Datastore e protocolli: NFS"

Connessione di rete diretta

Gli amministratori dello storage a volte preferiscono semplificare le loro infrastrutture rimuovendo gli switch di rete dalla configurazione. Questo può essere supportato in alcuni scenari. Tuttavia, ci sono alcune limitazioni e avvertimenti da essere informati di.

ISCSI e NVMe/TCP

Un host che utilizza iSCSI o NVMe/TCP può essere collegato direttamente a un sistema storage e funzionare normalmente. La ragione è la pedata. Le connessioni dirette a due storage controller differenti offrono due percorsi indipendenti per il flusso di dati. La perdita di percorso, porta o controller non impedisce l'utilizzo dell'altro percorso.

NFS

È possibile utilizzare lo storage NFS con connessione diretta, ma con una limitazione significativa: Il failover non funzionerà senza una significativa attività di scripting, che sarà responsabilità del cliente.

Il motivo per cui il failover senza interruzioni è complicato con lo storage NFS connesso direttamente è il routing che si verifica sul sistema operativo locale. Ad esempio, si supponga che un host abbia un indirizzo IP 192.168.1.1/24 e che sia collegato direttamente a un controller ONTAP con un indirizzo IP 192.168.1.50/24. Durante il failover, l'indirizzo 192.168.1.50 può eseguire il failover sull'altro controller e sarà disponibile per l'host, ma in che modo l'host rileva la sua presenza? L'indirizzo 192.168.1.1 originale esiste ancora sulla scheda di rete host che non si connette più a un sistema operativo. Il traffico destinato a 192.168.1.50 continuerebbe ad essere inviato a una porta di rete inutilizzabile.

La seconda scheda NIC del sistema operativo potrebbe essere configurata come 19 2.168.1.2 e sarebbe in grado di comunicare con l'indirizzo 192.168.1.50 non riuscito, ma le tabelle di routing locali avrebbero un valore predefinito di utilizzo di un solo indirizzo **e di un solo indirizzo** per comunicare con la subnet 192.168.1.0/24. Un amministratore di sistema potrebbe creare un framework di script che rilevi una connessione di rete non riuscita e alteri le tabelle di routing locali o che porti le interfacce verso l'alto e verso il basso. La procedura esatta dipende dal sistema operativo in uso.

In pratica, i clienti NetApp dispongono di NFS con connessione diretta, ma in genere solo per i workload in cui le pause io durante i failover sono accettabili. Quando si utilizzano i supporti rigidi, non devono verificarsi errori di i/o durante tali pause. L'io dovrebbe bloccarsi fino a quando i servizi non vengono ripristinati, mediante failback o intervento manuale, per spostare gli indirizzi IP tra le schede NIC dell'host.

Connessione diretta FC

Non è possibile connettere direttamente un host a un sistema storage ONTAP utilizzando il protocollo FC. Il motivo è l'uso di NPIV. Il WWN che identifica una porta FC ONTAP per la rete FC utilizza un tipo di virtualizzazione chiamato NPIV. Qualsiasi dispositivo collegato a un sistema ONTAP deve essere in grado di riconoscere un WWN NPIV. Attualmente non vi sono fornitori di HBA che offrono un HBA che può essere installato in un host in grado di supportare un target NPIV.

Clonazione di VM e datastore

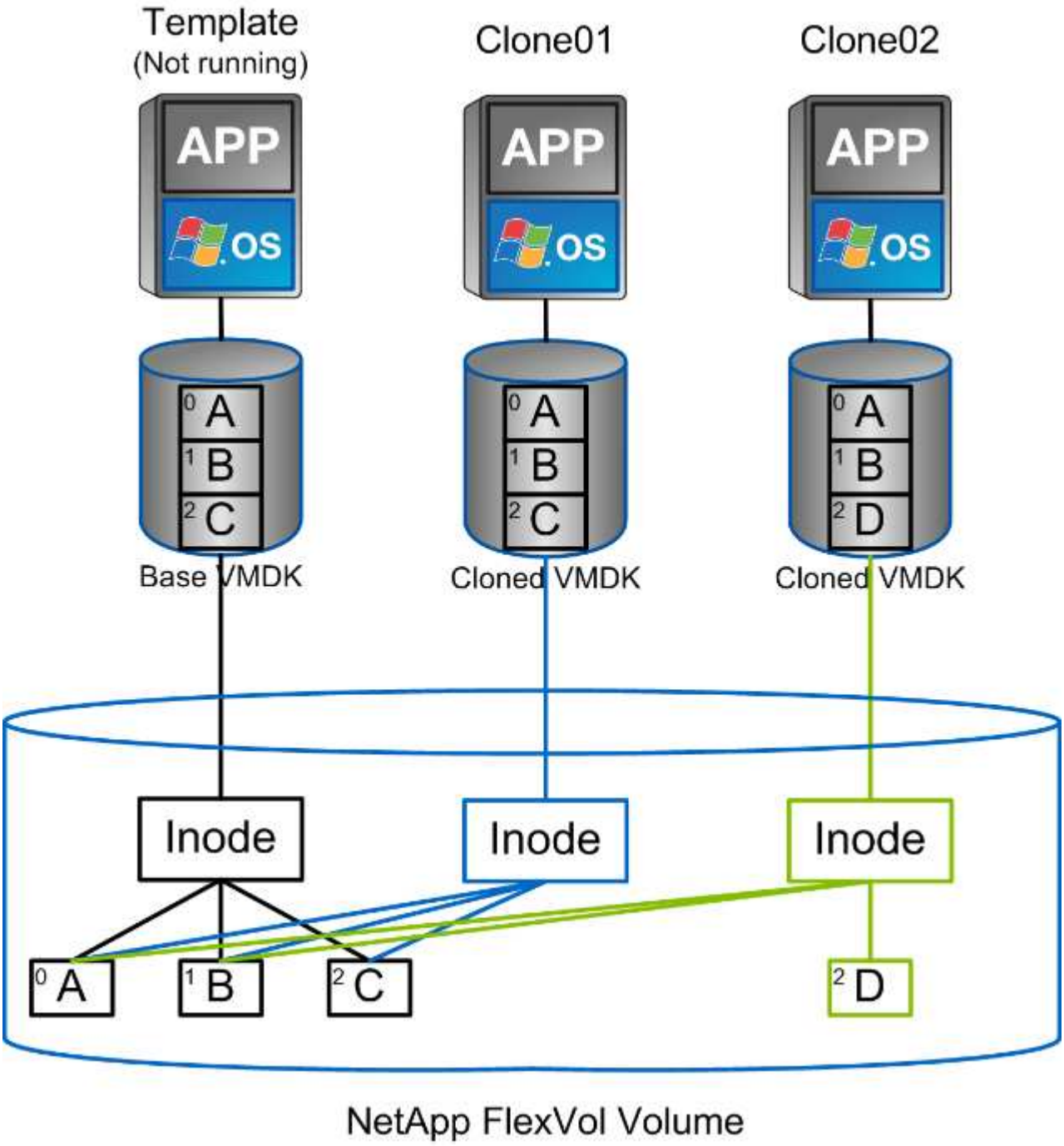
La clonazione di un oggetto storage consente di creare rapidamente copie da utilizzare ulteriormente, ad esempio il provisioning di macchine virtuali aggiuntive, operazioni di backup/recovery e così via.

In vSphere, è possibile clonare una macchina virtuale, un disco virtuale, un vVol o un datastore. Dopo essere stato clonato, l'oggetto può essere ulteriormente personalizzato, spesso attraverso un processo automatizzato.

VSphere supporta entrambi i cloni di copia completa e i cloni collegati, in cui tiene traccia delle modifiche separatamente dall'oggetto originale.

I cloni collegati sono ideali per risparmiare spazio, ma aumentano la quantità di i/o che vSphere gestisce per la macchina virtuale, influenzando le performance di quella macchina virtuale e forse dell'host in generale. Ecco perché i clienti di NetApp spesso utilizzano cloni basati su sistemi storage per ottenere il meglio di entrambi i mondi: Un utilizzo efficiente dello storage e maggiori performance.

La seguente figura illustra la clonazione ONTAP.



Il cloning può essere scaricato sui sistemi che eseguono ONTAP attraverso diversi meccanismi, generalmente a livello di VM, vVol o datastore. Questi includono quanto segue:

- VVol che utilizzano le API di NetApp vSphere per il provider di consapevolezza dello storage (VASA). I cloni ONTAP sono utilizzati per supportare le snapshot vVol gestite da vCenter, che sono efficienti in termini di spazio con effetto i/o minimo per crearle ed eliminarle. Le VM possono anche essere clonate utilizzando vCenter e vengono anche trasferite in ONTAP, sia all'interno di un singolo datastore/volume che tra datastore/volumi.
- Clonazione e migrazione di vSphere con API vSphere – integrazione array (VAAI). Le operazioni di cloning delle macchine virtuali possono essere trasferite in ONTAP negli ambienti SAN e NAS (NetApp fornisce un plug-in ESXi per consentire VAAI per NFS), mentre vSphere alleggerisce il carico delle operazioni delle macchine virtuali cold (spente) in un datastore NAS, mentre le operazioni delle macchine virtuali hot (cloning e storage vMotion) vengono anch'esse trasferite nella SAN. ONTAP utilizza l'approccio più efficiente in base all'origine e alla destinazione. Questa funzionalità viene utilizzata anche da "Omnissa Vista orizzonte".
- SRA (utilizzato con VMware Live Site Recovery/Site Recovery Manager). In questo caso, i cloni vengono utilizzati per testare il ripristino della replica DR senza interruzioni.
- Backup e recovery con strumenti NetApp come SnapCenter. I cloni delle macchine virtuali vengono utilizzati per verificare le operazioni di backup e per montare un backup di una macchina virtuale in modo che sia possibile ripristinare i singoli file.

La clonazione offload di ONTAP può essere invocata da VMware, NetApp e da strumenti di terze parti. I cloni che vengono scaricati su ONTAP presentano diversi vantaggi. Nella maggior parte dei casi, sono efficienti in termini di spazio e richiedono storage solo per le modifiche all'oggetto; non vi sono effetti aggiuntivi sulle performance per la lettura e la scrittura e in alcuni casi le performance sono migliorate grazie alla condivisione dei blocchi nelle cache ad alta velocità. Inoltre, consentono di trasferire cicli CPU e i/o di rete dal server ESXi. L'offload delle copie in un datastore tradizionale mediante un FlexVol volume può essere rapido ed efficiente con la licenza FlexClone (inclusa nella licenza ONTAP One), ma le copie tra volumi FlexVol potrebbero essere più lente. Se si mantengono i modelli di macchine virtuali come origine dei cloni, è consigliabile posizionarli all'interno del volume datastore (utilizzare cartelle o librerie di contenuti per organizzarli) per cloni veloci ed efficienti in termini di spazio.

È inoltre possibile clonare un volume o un LUN direttamente in ONTAP per clonare un datastore. Con gli archivi di dati NFS, la tecnologia FlexClone può clonare un intero volume e il clone può essere esportato da ONTAP e montato da ESXi come altro archivio di dati. Per gli archivi di dati VMFS, ONTAP può clonare un LUN all'interno di un volume o di un intero volume, inclusi uno o più LUN. Un LUN contenente un VMFS deve essere mappato a un gruppo di iniziatori ESXi (igroup) e quindi rassegnato da ESXi per essere montato e utilizzato come datastore regolare. Per alcuni casi di utilizzo temporaneo, è possibile montare un VMFS clonato senza disdire. Dopo aver clonato un datastore, è possibile registrare, riconfigurare e personalizzare le macchine virtuali all'interno dell'IT come se fossero macchine virtuali clonate singolarmente.

In alcuni casi, è possibile utilizzare funzionalità aggiuntive con licenza per migliorare la clonazione, ad esempio SnapRestore per il backup o FlexClone. Queste licenze sono spesso incluse nei bundle di licenze senza costi aggiuntivi. È necessaria una licenza FlexClone per le operazioni di cloning di vVol e per supportare le snapshot gestite di un vVol (offload dall'hypervisor a ONTAP). Una licenza FlexClone può anche migliorare alcuni cloni basati su VAAI se utilizzati all'interno di un datastore/volume (crea copie istantanee ed efficienti in termini di spazio invece di copie a blocchi). Viene inoltre utilizzato dall'SRA per il test del ripristino di una replica DR e da SnapCenter per le operazioni di clonazione e per sfogliare le copie di backup per ripristinare singoli file.

Protezione dei dati

Il backup e il ripristino rapido delle macchine virtuali (VM) sono vantaggi chiave dell'utilizzo di ONTAP per vSphere. Questa funzionalità può essere facilmente gestita all'interno di vCenter tramite il plug-in SnapCenter per VMware vSphere. Molti clienti migliorano le loro soluzioni di backup di terze parti con SnapCenter per sfruttare la

tecnologia Snapshot di ONTAP, poiché offre il modo più veloce e semplice di ripristinare una macchina virtuale tramite ONTAP. SnapCenter è disponibile gratuitamente per i clienti che dispongono della licenza ONTAP ONE; potrebbero essere disponibili anche altri bundle di licenze.

Inoltre, il plug-in SnapCenter per VMware può integrarsi con "BlueXP per il backup e ripristino di macchine virtuali", consentendo efficaci soluzioni di backup 3-2-1 per la maggior parte dei sistemi ONTAP. Si noti che alcune tariffe potrebbero essere applicate se si utilizza il backup e recovery di BlueXP per macchine virtuali con servizi premium, ad esempio gli archivi di oggetti per storage di backup aggiuntivo. In questa sezione vengono illustrate le varie opzioni disponibili per la protezione di macchine virtuali e datastore.

Snapshot dei volumi NetApp ONTAP

Utilizza le snapshot per creare copie rapide della tua macchina virtuale o del datastore senza influire sulle performance, quindi inviale a un sistema secondario utilizzando SnapMirror per la data Protection off-site a lungo termine. Questo approccio riduce al minimo lo spazio di storage e la larghezza di banda della rete memorizzando solo le informazioni modificate.

Le snapshot sono una funzionalità chiave di ONTAP, consentendoti di creare copie point-in-time dei tuoi dati. Sono efficienti in termini di spazio e possono essere create rapidamente, rendendole ideali per proteggere macchine virtuali e datastore. Gli snapshot possono essere utilizzati per vari scopi, incluso backup, recovery e test. Questi Snapshot sono diversi dalle Snapshot VMware (di coerenza) e sono adatti per una protezione a lungo termine. Gli snapshot gestiti da vCenter di VMware sono consigliati solo per un utilizzo a breve termine, a causa delle prestazioni e di altri effetti. Per "Limitazioni delle snapshot"ulteriori dettagli, fare riferimento a.

Vengono create a livello di volume copie Snapshot che possono essere utilizzate per proteggere tutte le macchine virtuali e i datastore all'interno di tale volume. Ciò significa che puoi creare una snapshot di un intero datastore, che include tutte le macchine virtuali all'interno di tale datastore.

Per gli archivi dati NFS, è possibile visualizzare facilmente i file VM nelle istantanee navigando nella directory .istantanee. Ciò consente di accedere e ripristinare rapidamente i file da uno snapshot senza dover utilizzare una soluzione di backup specifica.

Per gli archivi dati VMFS, è possibile creare un FlexClone dell'archivio dati in base allo snapshot desiderato. Ciò consente di creare un nuovo datastore basato sullo snapshot, che può essere utilizzato a scopo di test o sviluppo. La FlexClone occupa spazio solo per le modifiche apportate dopo la creazione della snapshot, creando un modo efficiente in termini di spazio per creare una copia dell'archivio dati. Una volta creato FlexClone, è possibile mappare la LUN o il namespace a un host ESXi come un normale datastore. Ciò consente non solo di ripristinare file VM specifici, ma anche di creare rapidamente ambienti di test o sviluppo in base ai dati di produzione, senza influire sulle performance dell'ambiente di produzione.

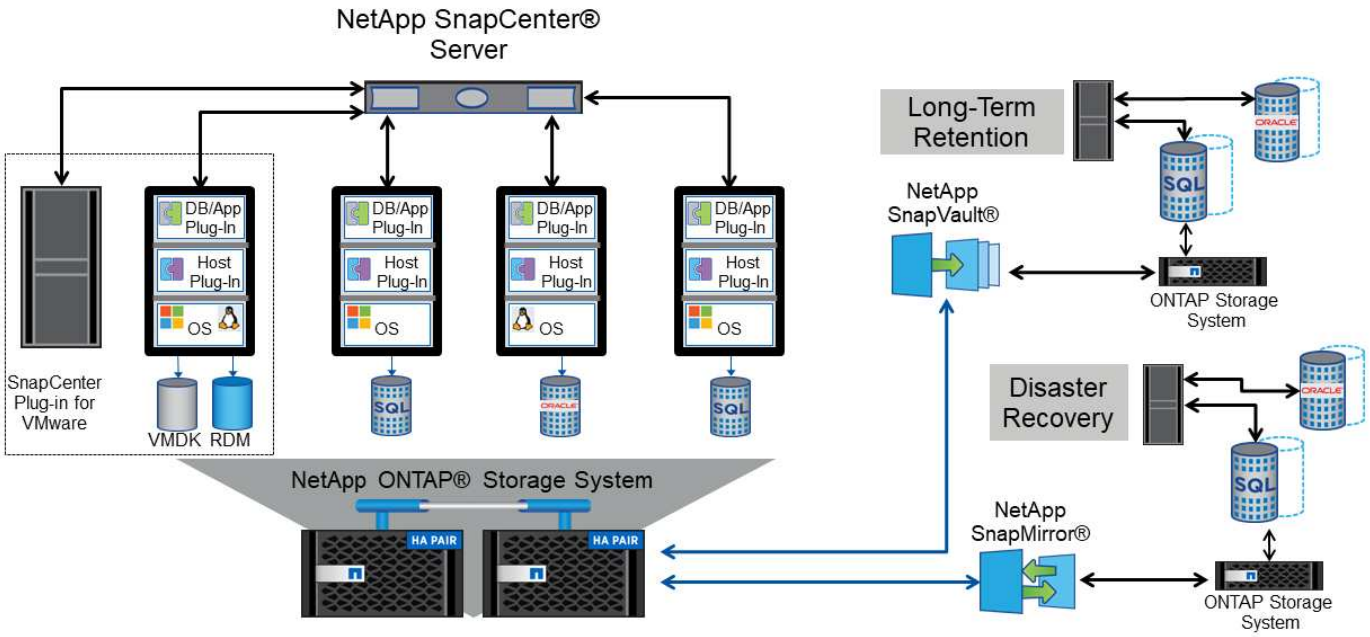
Per ulteriori informazioni sulle istantanee, consultare la documentazione di ONTAP. I seguenti link forniscono ulteriori dettagli: "Copie snapshot locali ONTAP" "Replica ONTAP SnapMirror"

Plug-in SnapCenter per VMware vSphere

SnapCenter consente di creare policy di backup che possono essere applicate a più processi. Questi criteri possono definire pianificazione, conservazione, replica e altre funzionalità. Essi continuano a consentire una selezione opzionale di snapshot coerenti con le macchine virtuali, che sfrutta la capacità dell'hypervisor di mettere in pausa l'i/o prima di scattare una snapshot VMware. Tuttavia, a causa dell'effetto delle performance delle snapshot VMware, in genere non sono consigliate, a meno che non sia necessario interrompere il file system guest. Utilizza invece le snapshot per la protezione generale e utilizza strumenti applicativi come i plug-in delle applicazioni SnapCenter per proteggere i dati transazionali come SQL Server o Oracle.

Questi plug-in offrono funzionalità estese per proteggere i database in ambienti fisici e virtuali. Grazie a vSphere, puoi utilizzarli per proteggere i database SQL Server o Oracle, in cui i dati vengono memorizzati su LUN RDM, vVol o namespace NVMe/TCP e LUN iSCSI direttamente connessi al sistema operativo guest, oppure file VMDK su datastore VMFS o NFS. I plug-in consentono di specificare diversi tipi di backup del database, supportando il backup online o offline e proteggendo i file di database insieme ai file di registro. Oltre al backup e alla recovery, i plug-in supportano anche la clonazione dei database a scopo di sviluppo o test.

La figura seguente mostra un esempio di implementazione di SnapCenter.



Per informazioni sul dimensionamento, fare riferimento a. "Guida al dimensionamento per il plugin SnapCenter per VMware vSphere"

Tool ONTAP per VMware vSphere con VMware Live Site Recovery

I tool di ONTAP per VMware vSphere (OT4VS) sono un plug-in gratuito che offre una perfetta integrazione tra VMware vSphere e NetApp ONTAP. Consente di gestire lo storage ONTAP direttamente dal client web vSphere, semplificando l'esecuzione di attività come il provisioning dello storage, la gestione della replica e il monitoraggio delle performance.

Per funzionalità di disaster recovery migliorate, considerare l'utilizzo di NetApp SRA for ONTAP, che fa parte degli strumenti ONTAP per VMware vSphere, insieme a VMware Live Site Recovery (precedentemente noto come Site Recovery Manager). Questo tool non solo supporta la replica di datastore in un sito di disaster recovery mediante SnapMirror, ma consente anche di eseguire test senza interruzioni nell'ambiente di disaster recovery mediante il cloning dei datastore replicati. Inoltre, il recovery da un disastro e la reprotezione della produzione dopo la risoluzione di un black-out sono ottimizzati grazie alle funzionalità di automazione integrate.

Disaster recovery BlueXP

Il Disaster Recovery (DR) BlueXP è un servizio basato sul cloud che fornisce una soluzione completa per proteggere i vostri dati e le applicazioni in caso di disastro. Offre una gamma di funzionalità, tra cui failover e failback automatizzati, recovery point-in-time multipli, disaster recovery coerente con l'applicazione e supporto per sistemi ONTAP on-premise e basati sul cloud. BlueXP DR è progettato per funzionare senza problemi con ONTAP e con il tuo ambiente VMware vSphere, offrendo una soluzione unificata per il disaster recovery.

VSphere Metro Storage Cluster (vMSC) con sincronizzazione attiva NetApp MetroCluster e SnapMirror

Infine, per il massimo livello di protezione dei dati, prendere in considerazione la configurazione di VMware vSphere Metro Storage Cluster (vMSC) che utilizza NetApp MetroCluster. VMSC è una soluzione supportata da NetApp con certificazione VMware che utilizza la replica sincrona, offrendo gli stessi vantaggi di un cluster ad alta disponibilità ma distribuita in siti separati per la protezione contro il disastro del sito. La sincronizzazione attiva NetApp SnapMirror, con ASA e AFF e MetroCluster con AFF, offre configurazioni convenienti per una replica sincrona con recovery trasparente da qualsiasi guasto di un componente storage singolo, nonché recovery trasparente nel caso di SnapMirror Active Sync o recovery con singolo comando in caso di disastro del sito con MetroCluster. VMSC è descritto in maggior dettaglio in "TR-4128".

Qualità del servizio (QoS)

I limiti di throughput sono utili per controllare i livelli di servizio, gestire carichi di lavoro sconosciuti o testare le applicazioni prima della distribuzione per assicurarsi che non influiscano su altri carichi di lavoro in produzione. Possono anche essere utilizzati per limitare un carico di lavoro ingombrante dopo l'identificazione.

Supporto della policy QoS di ONTAP

I sistemi che eseguono ONTAP possono utilizzare la funzionalità di qualità del servizio di storage per limitare il throughput in Mbps e/o i/o al secondo (IOPS) per diversi oggetti di storage come file, LUN, volumi o intere SVM.

Sono supportati anche i livelli minimi di servizio basati sugli IOPS per fornire performance costanti per gli oggetti SAN in ONTAP 9.2 e per gli oggetti NAS in ONTAP 9.3.

Il limite massimo di throughput QoS su un oggetto può essere impostato in Mbps e/o IOPS. Se vengono utilizzati entrambi, il primo limite raggiunto viene applicato da ONTAP. Un carico di lavoro può contenere più oggetti e una policy QoS può essere applicata a uno o più carichi di lavoro. Quando una policy viene applicata a più carichi di lavoro, i carichi di lavoro condividono il limite totale della policy. Gli oggetti nidificati non sono supportati (ad esempio, i file all'interno di un volume non possono avere una propria policy). I valori minimi di QoS possono essere impostati solo in IOPS.

I seguenti strumenti sono attualmente disponibili per la gestione delle policy di qualità del servizio ONTAP e per applicarle agli oggetti:

- CLI ONTAP
- Gestore di sistema di ONTAP
- OnCommand Workflow Automation
- Active IQ Unified Manager
- Kit di strumenti NetApp PowerShell per ONTAP
- Strumenti ONTAP per il provider VMware vSphere VASA

Per assegnare una policy di QoS a un LUN, inclusi VMFS e RDM, è possibile ottenere la SVM di ONTAP (visualizzata come Vserver), il percorso del LUN e il numero di serie dal menu dei sistemi storage nella home page degli strumenti ONTAP per VMware vSphere. Seleziona il sistema storage (SVM), quindi gli oggetti correlati > SAN. Utilizzare questo approccio quando si specifica la qualità del servizio utilizzando uno degli strumenti ONTAP.

Fare riferimento a. "Panoramica sulla gestione e sul monitoraggio delle performance" per ulteriori informazioni.

Datastore NFS non vVol

È possibile applicare una policy di QoS ONTAP all'intero datastore o ai singoli file VMDK al suo interno. Tuttavia, è importante comprendere che tutte le macchine virtuali di un datastore NFS tradizionale (non vVol) condividono una coda i/o comune da un determinato host. Se una macchina virtuale viene rallentata da una policy di QoS ONTAP, in pratica tutto l'i/o del datastore sembrerà rallentato per quell'host.

Esempio:
* È stato configurato un limite QoS su VM1.vmdk per un volume montato come datastore NFS tradizionale dall'host esxi-01.
* Lo stesso host (esxi-01) utilizza VM2.vmdk e si trova sullo stesso volume.
* Se VM1.vmdk viene rallentato, allora anche VM2.vmdk sembrerà essere rallentato poiché condivide la stessa coda io con VM1.vmdk.

Questo non si applica ai vVol.

A partire da vSphere 6,5 è possibile gestire limiti granulari dei file sui datastore non vVol sfruttando la gestione basata su criteri dello storage (SPBM, Storage Policy-Based Management) con Storage i/o Control (SIOC) v2.

Fare riferimento ai link seguenti per ulteriori informazioni sulla gestione delle prestazioni con i criteri SIOC e SPBM.

"Regole basate su host SPBM: SIOC v2"
"Gestisci le risorse i/o di storage con vSphere"

Per assegnare un criterio QoS a un VMDK su NFS, attenersi alle seguenti linee guida:

- La policy deve essere applicata a `vmname-flat.vmdk` che contiene l'immagine effettiva del disco virtuale, non il `vmname.vmdk` (file di descrizione del disco virtuale) o. `vmname.vmx` (File descrittore VM).
- Non applicare policy ad altri file di macchine virtuali, ad esempio file di swap virtuali (`vmname.vswp`).
- Quando si utilizza il client Web vSphere per trovare i percorsi di file (datastore > file), tenere presente che combina le informazioni di `- flat.vmdk` e. `. vmdk` e mostra semplicemente un file con il nome di `. vmdk` ma le dimensioni di `- flat.vmdk`. Aggiungi `-flat` nel nome del file per ottenere il percorso corretto.

Gli archivi dati FlexGroup offrono funzionalità QoS avanzate quando si utilizzano gli strumenti ONTAP per VMware vSphere 9.8 e versioni successive. È possibile impostare facilmente la QoS su tutte le macchine virtuali di un datastore o su macchine virtuali specifiche. Per ulteriori informazioni, consultare la sezione FlexGroup di questo report. Tieni presente che si applicano ancora le limitazioni di qualità del servizio menzionate in precedenza per i datastore NFS tradizionali.

Datastore VMFS

Utilizzando le LUN di ONTAP, le policy di QoS possono essere applicate al volume FlexVol che contiene le LUN o le singole LUN, ma non ai singoli file VMDK perché ONTAP non conosce il file system VMFS.

Datastore vVol

È possibile impostare facilmente una qualità del servizio minima e/o massima su singole macchine virtuali o VMDK senza impatti su altre macchine virtuali o VMDK grazie alla gestione basata su policy di storage e ai vVol.

Durante la creazione di un profilo di capacità storage per il container vVol, specifica un valore IOPS max e/o

min in termini di performance, quindi fai riferimento a questo SCP con la policy storage delle macchine virtuali. Utilizzare questo criterio quando si crea la macchina virtuale o si applica il criterio a una macchina virtuale esistente.

VVol richiede l'utilizzo dei tool ONTAP per VMware vSphere, che funziona come provider VASA per ONTAP. Fare riferimento a "Volumi virtuali di VMware vSphere (vVol) con ONTAP" per le procedure consigliate per i vVol.

QoS ONTAP e SIOC VMware

ONTAP QoS e VMware vSphere Storage i/o Control (SIOC) sono tecnologie complementari che gli amministratori di vSphere e dello storage possono utilizzare insieme per gestire le performance delle VM vSphere ospitate in sistemi che eseguono ONTAP. Ogni strumento ha i propri punti di forza, come mostrato nella tabella seguente. A causa dei diversi ambiti di VMware vCenter e ONTAP, alcuni oggetti possono essere visti e gestiti da un sistema e non dall'altro.

Proprietà	QoS ONTAP	VMware SIOC
Se attivo	La policy è sempre attiva	Attivo quando esiste un conflitto (latenza dell'archivio dati oltre la soglia)
Tipo di unità	IOPS, Mbps	IOPS, condivisioni
VCenter o ambito applicativo	Più ambienti vCenter, altri hypervisor e applicazioni	Singolo server vCenter
Impostare QoS su VM?	VMDK solo su NFS	VMDK su NFS o VMFS
Impostare QoS su LUN (RDM)?	Sì	No
Impostare la qualità del servizio su LUN (VMFS)?	Sì	Sì (il datastore può essere rallentato)
Impostare QoS sul volume (datastore NFS)?	Sì	Sì (il datastore può essere rallentato)
Impostare QoS su SVM (tenant)?	Sì	No
Approccio basato su policy?	Sì; può essere condiviso da tutti i carichi di lavoro della policy o applicato in toto a ciascun carico di lavoro della policy.	Sì, con vSphere 6.5 e versioni successive.
Licenza richiesta	Incluso con ONTAP	Enterprise Plus

VMware Storage Distributed Resource Scheduler

VMware Storage Distributed Resource Scheduler (SDR) è una funzionalità vSphere che consente di posizionare le macchine virtuali sullo storage in base alla latenza i/o corrente e all'utilizzo dello spazio. Quindi, sposta le VM o i VMDK senza interruzioni tra gli archivi dati in un cluster di datastore (noto anche come pod), selezionando il migliore datastore in cui posizionare le VM o i VMDK nel cluster di datastore. Un cluster di datastore è un insieme di datastore simili che vengono aggregati in una singola unità di consumo dal punto di vista dell'amministratore di vSphere.

Quando si utilizzano DSP con strumenti ONTAP per VMware vSphere, è necessario prima creare un datastore con il plug-in, utilizzare vCenter per creare il cluster di datastore e quindi aggiungere il datastore. Una volta creato il cluster di datastore, è possibile aggiungere ulteriori datastore al cluster di datastore direttamente dalla

procedura guidata di provisioning nella pagina Dettagli.

Altre Best practice ONTAP per I DSP includono:

- Tutti gli archivi dati del cluster devono utilizzare lo stesso tipo di storage (ad esempio SAS, SATA o SSD), tutti gli archivi dati VMFS o NFS e avere le stesse impostazioni di replica e protezione.
- Considerare l'utilizzo DEGLI SDR in modalità predefinita (manuale). Questo approccio consente di rivedere i suggerimenti e decidere se applicarli o meno. Tenere presente i seguenti effetti delle migrazioni VMDK:
  - Quando GLI SDR spostano i VMDK tra datastore, qualsiasi risparmio di spazio derivante dalla clonazione o deduplica ONTAP viene perso. È possibile rieseguire la deduplica per recuperare questi risparmi.
  - Dopo che LE SDR spostano i VMDK, NetApp consiglia di ricreare gli snapshot nel datastore di origine, poiché lo spazio è altrimenti bloccato dalla VM che è stata spostata.
  - Lo spostamento di VMDK tra datastore sullo stesso aggregato ha pochi benefici e GLI SDR non hanno visibilità su altri carichi di lavoro che potrebbero condividere l'aggregato.

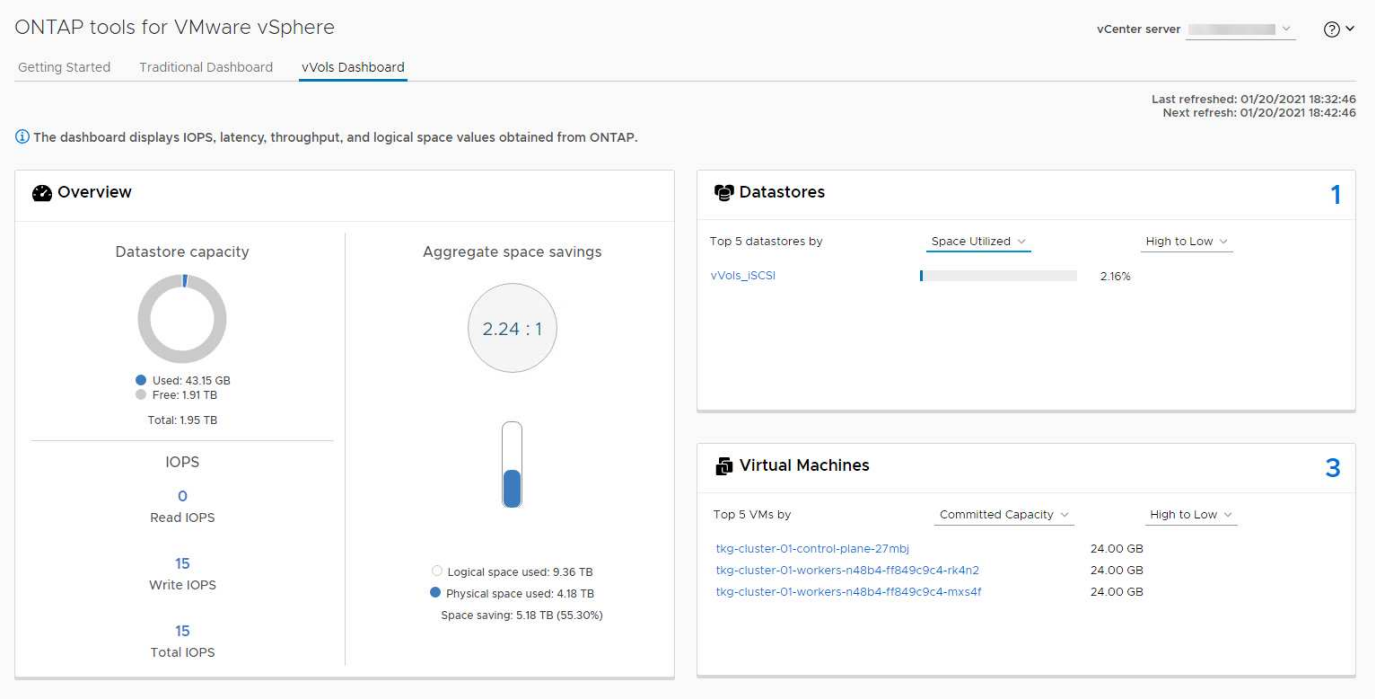
Gestione basata su criteri di archiviazione e vVol

Le API VMware vSphere per Storage Awareness (VASA) consentono a un amministratore dello storage di configurare con facilità i datastore con funzionalità ben definite, consentendo all'amministratore delle macchine virtuali di utilizzare tali dati quando necessario per eseguire il provisioning delle macchine virtuali senza dover interagire tra loro. Vale la pena di dare un'occhiata a questo approccio per scoprire in che modo può semplificare le operazioni di virtualizzazione dello storage ed evitare un lavoro molto banale.

Prima di VASA, gli amministratori delle macchine virtuali potevano definire policy di storage delle macchine virtuali, ma dovevano collaborare con l'amministratore dello storage per identificare i datastore appropriati, spesso utilizzando documentazione o convenzioni di naming. Con VASA, l'amministratore dello storage può definire una serie di funzionalità di storage, tra cui performance, tiering, crittografia e replica. Un insieme di funzionalità per un volume o un set di volumi viene definito SCP (Storage Capability Profile).

SCP supporta la qualità del servizio minima e/o massima per i vVol di dati di una VM. La QoS minima è supportata solo sui sistemi AFF. Gli strumenti ONTAP per VMware vSphere includono una dashboard che visualizza le performance granulari delle macchine virtuali e la capacità logica per i vVol sui sistemi ONTAP.

La figura seguente mostra i tool ONTAP per il dashboard di VMware vSphere 9.8 vVol.



Una volta definito il profilo di capacità dello storage, è possibile utilizzarlo per eseguire il provisioning delle macchine virtuali utilizzando la policy di storage che ne identifica i requisiti. La mappatura tra il criterio di storage delle macchine virtuali e il profilo di capacità dello storage del datastore consente a vCenter di visualizzare un elenco di datastore compatibili per la selezione. Questo approccio è noto come gestione basata su criteri di storage.

VASA offre la tecnologia per eseguire query sullo storage e restituire un set di funzionalità di storage a vCenter. I vendor provider VASA forniscono la traduzione tra le API e i costrutti del sistema storage e le API VMware comprese da vCenter. Il provider VASA di NetApp per ONTAP viene offerto come parte dei tool ONTAP per macchina virtuale dell'appliance VMware vSphere, mentre il plug-in vCenter fornisce l'interfaccia per il provisioning e la gestione dei datastore vVol, nonché la capacità di definire profili di funzionalità dello storage (SCP).

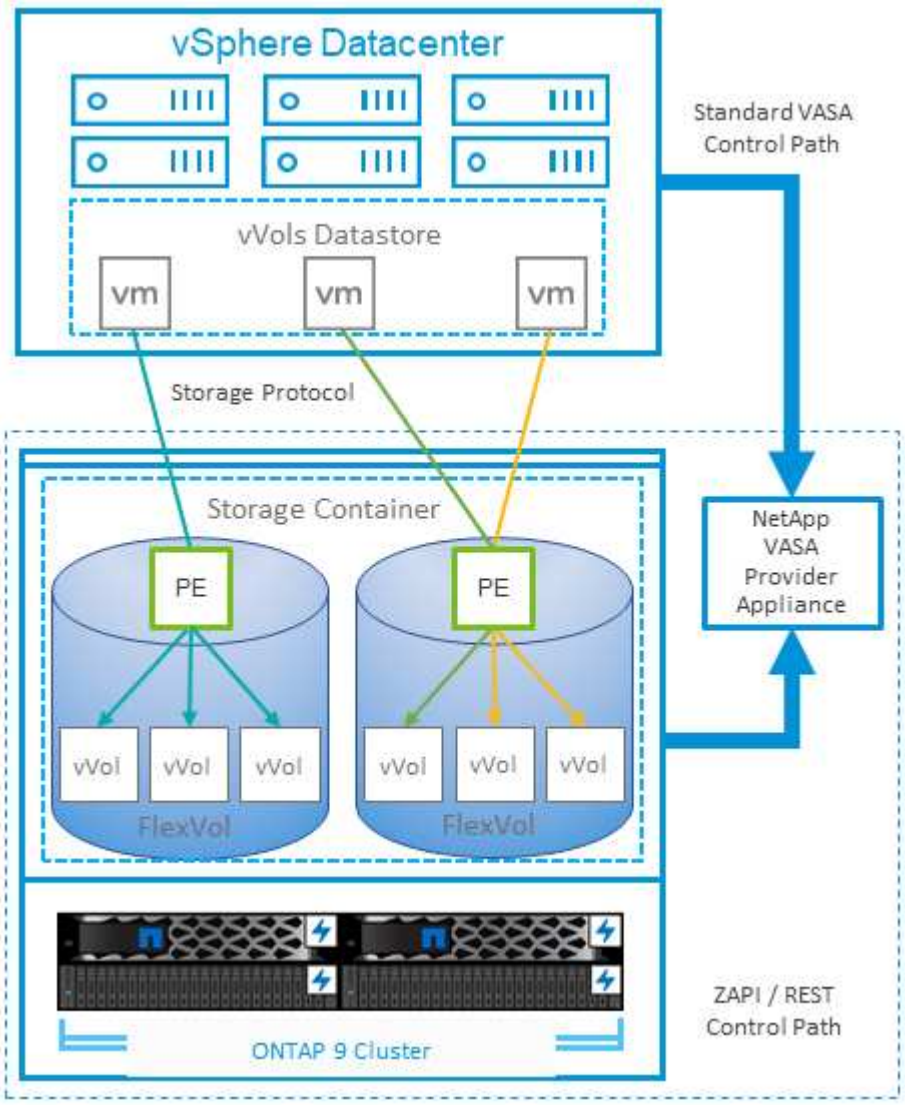
ONTAP supporta gli archivi dati VMFS e NFS vVol. L'utilizzo di vVol con datastore SAN offre alcuni dei vantaggi di NFS, come la granularità a livello di macchine virtuali. Di seguito sono riportate alcune Best practice da prendere in considerazione e ulteriori informazioni sono disponibili in "TR-4400":

- Un datastore vVol può essere costituito da più volumi FlexVol su più nodi del cluster. L'approccio più semplice è un singolo datastore, anche quando i volumi hanno funzionalità diverse. SPBM garantisce l'utilizzo di un volume compatibile per la macchina virtuale. Tuttavia, tutti i volumi devono far parte di una singola SVM ONTAP e devono essere accessibili utilizzando un singolo protocollo. È sufficiente una LIF per nodo per ogni protocollo. Evitare di utilizzare più release di ONTAP all'interno di un singolo datastore vVol, poiché le funzionalità dello storage potrebbero variare tra le varie release.
- Utilizza i tool ONTAP per il plug-in VMware vSphere per creare e gestire datastore vVol. Oltre a gestire il datastore e il relativo profilo, crea automaticamente un endpoint del protocollo per accedere ai vVol, se necessario. Se si utilizzano LUN, tenere presente che i LUN PES vengono mappati utilizzando LUN ID 300 e superiori. Verificare che l'impostazione di sistema avanzata dell'host ESXi sia corretta `Disk.MaxLUN` Consente un numero di ID LUN superiore a 300 (il valore predefinito è 1,024). Eseguire questa operazione selezionando l'host ESXi in vCenter, quindi la scheda Configura e trova `Disk.MaxLUN` Nell'elenco delle Advanced System Settings (Impostazioni di sistema avanzate).
- Non installare o migrare il provider VASA, il server vCenter (basato su appliance o Windows) o i tool ONTAP per VMware vSphere in sé su un datastore vVols, perché in tal caso sono dipendenti

reciprocamente, limitando la possibilità di gestirli in caso di interruzione dell'alimentazione o di altre interruzioni del data center.

- Eseguire regolarmente il backup della VM del provider VASA. Crea almeno snapshot orarie del datastore tradizionale che contiene il provider VASA. Per ulteriori informazioni sulla protezione e il ripristino del provider VASA, consulta questa sezione "Articolo della Knowledge base".

La figura seguente mostra i componenti di vVol.



Migrazione e backup del cloud

Un altro punto di forza di ONTAP è l'ampio supporto per il cloud ibrido, che unisce i sistemi nel tuo cloud privato on-premise con funzionalità di cloud pubblico. Ecco alcune soluzioni cloud NetApp che possono essere utilizzate insieme a vSphere:

- **Offerte di prima parte.** Amazon FSX per NetApp ONTAP, Google Cloud NetApp Volumes e Azure NetApp Files offrono servizi di storage gestito multiprotocollo dalle performance elevate negli ambienti di cloud pubblico leader. Possono essere utilizzati direttamente da VMware Cloud su AWS (VMC on AWS), Azure VMware Solution (AVS) e Google Cloud VMware Engine (GCVE) come datastore o storage per sistemi operativi guest (GOS) e istanze di calcolo.

- **Cloud Volumes ONTAP.** il software per la gestione dei dati NetApp Cloud Volumes ONTAP offre controllo, protezione, flessibilità ed efficienza ai tuoi dati sul cloud di tua scelta. Cloud Volumes ONTAP è un software di gestione dei dati nativo del cloud basato sullo storage ONTAP. Utilizzare insieme a Cloud Manager per implementare e gestire le istanze di Cloud Volumes ONTAP insieme ai sistemi ONTAP on-premise. Sfrutta le funzionalità NAS e SAN iSCSI avanzate insieme a una gestione dei dati unificata, incluse le snapshot e la replica SnapMirror.
- **Servizi cloud.** Utilizza il backup e recovery BlueXP o il cloud SnapMirror per proteggere i dati dai sistemi on-premise utilizzando il cloud storage pubblico. Cloud Sync aiuta a migrare e mantenere i dati sincronizzati tra NAS, archivi di oggetti e storage Cloud Volumes Service. Il disaster recovery di BlueXP è una soluzione conveniente ed efficiente per sfruttare le tecnologie NetApp come base per una soluzione di disaster recovery solida e capace per disaster recovery su cloud, DR on-premise e on-premise.
- **FabricPool.** FabricPool offre tiering rapido e semplice per i dati ONTAP. È possibile migrare i blocchi cold in un archivio di oggetti nei cloud pubblici o in un archivio di oggetti StorageGRID privato e vengono richiamati automaticamente quando si accede nuovamente ai dati ONTAP. Oppure utilizzare il Tier di oggetti come terzo livello di protezione per i dati già gestiti da SnapVault. Questo approccio può consentirti di farlo "Memorizzazione di più snapshot delle macchine virtuali" Sui sistemi storage ONTAP primari e/o secondari.
- **ONTAP Select.** utilizza lo storage software-defined di NetApp per estendere il tuo cloud privato attraverso Internet a sedi e uffici remoti, dove puoi utilizzare ONTAP Select per supportare i servizi di file e blocchi e le stesse funzionalità di gestione dei dati vSphere presenti nel tuo data center aziendale.

Quando si progettano le applicazioni basate su macchine virtuali, considerare la futura mobilità del cloud. Ad esempio, invece di mettere insieme file di applicazioni e dati, utilizza un'esportazione LUN o NFS separata per i dati. Ciò consente di migrare la macchina virtuale e i dati separatamente ai servizi cloud.

Per un'analisi approfondita di altri argomenti relativi alla sicurezza, fare riferimento alle seguenti risorse.

- "Documentazione Cloud Volumes ONTAP"
- "Documentazione ONTAP Select"
- "Documentazione di backup e ripristino BlueXP"
- "Documentazione di disaster recovery di BlueXP"
- "Amazon FSX per NetApp ONTAP"
- "VMware Cloud su AWS"
- "Che cos'è Azure NetApp Files?
"
- "Soluzione VMware Azure"
- "Motore VMware Google Cloud"
- "Che cos'è Google Cloud NetApp Volumes?"

Crittografia per i dati vSphere

Oggi, la necessità di proteggere i dati inattivi è in aumento grazie alla crittografia. Sebbene l'attenzione iniziale fosse concentrata sulle informazioni finanziarie e sanitarie, c'è sempre più interesse a proteggere tutte le informazioni, che siano archiviate in file, database o altri tipi di dati.

I sistemi che eseguono ONTAP semplificano la protezione di qualsiasi dato grazie alla crittografia a riposo. La crittografia dello storage NetApp (NSE) utilizza dischi con crittografia automatica (SED) con ONTAP per

proteggere i dati di SAN e NAS. NetApp offre inoltre NetApp Volume Encryption e NetApp aggregate Encryption come approccio semplice e basato su software per crittografare i volumi su qualsiasi disco. Questa crittografia software non richiede unità disco speciali o gestori di chiavi esterne ed è disponibile per i clienti ONTAP senza costi aggiuntivi. Puoi eseguire l'aggiornamento e iniziare a utilizzarlo senza interruzioni per client o applicazioni, validati secondo lo standard FIPS 140-2 livello 1, incluso Onboard Key Manager.

Esistono diversi approcci per la protezione dei dati delle applicazioni virtualizzate in esecuzione su VMware vSphere. Un approccio consiste nel proteggere i dati con il software all'interno della macchina virtuale a livello di sistema operativo guest. Gli hypervisor più recenti, come vSphere 6.5, ora supportano la crittografia a livello di VM come alternativa. Tuttavia, la crittografia del software NetApp è semplice e offre i seguenti vantaggi:

- **Nessun effetto sulla CPU del server virtuale.** alcuni ambienti di server virtuali richiedono ogni ciclo di CPU disponibile per le proprie applicazioni, tuttavia i test hanno dimostrato che sono necessarie fino a 5 risorse di CPU con crittografia a livello di hypervisor. Anche se il software di crittografia supporta il set di istruzioni AES-NI di Intel per l'offload del carico di lavoro di crittografia (come fa la crittografia del software NetApp), questo approccio potrebbe non essere fattibile a causa del requisito di nuove CPU che non sono compatibili con i server meno recenti.
- **Key Manager integrato incluso.** La crittografia software NetApp include un gestore delle chiavi integrato senza costi aggiuntivi, il che semplifica l'avvio senza server di gestione delle chiavi ad alta disponibilità che sono complessi da acquistare e utilizzare.
- **Nessun effetto sull'efficienza dello storage.** le tecniche di efficienza dello storage, come deduplica e compressione, sono ampiamente utilizzate oggi e sono fondamentali per utilizzare i supporti su disco flash in modo conveniente. Tuttavia, i dati crittografati non possono in genere essere deduplicati o compressi. La crittografia dello storage e dell'hardware NetApp opera a un livello inferiore e consente l'utilizzo completo delle funzionalità di efficienza dello storage NetApp leader del settore, a differenza di altri approcci.
- **Crittografia granulare semplice del datastore.** con NetApp Volume Encryption, ogni volume ottiene la propria chiave AES a 256 bit. Se è necessario modificarlo, è possibile farlo con un singolo comando. Questo approccio è ideale se hai più tenant o hai bisogno di dimostrare una crittografia indipendente per diversi reparti o applicazioni. Questa crittografia viene gestita a livello di datastore, il che è molto più semplice della gestione di singole macchine virtuali.

Iniziare a utilizzare la crittografia del software è semplice. Una volta installata la licenza, è sufficiente configurare Onboard Key Manager specificando una passphrase e quindi creare un nuovo volume o spostare un volume lato storage per attivare la crittografia. NetApp sta lavorando per aggiungere un supporto più integrato per le funzionalità di crittografia nelle versioni future dei suoi strumenti VMware.

Per un'analisi approfondita di altri argomenti relativi alla sicurezza, fare riferimento alle seguenti risorse.

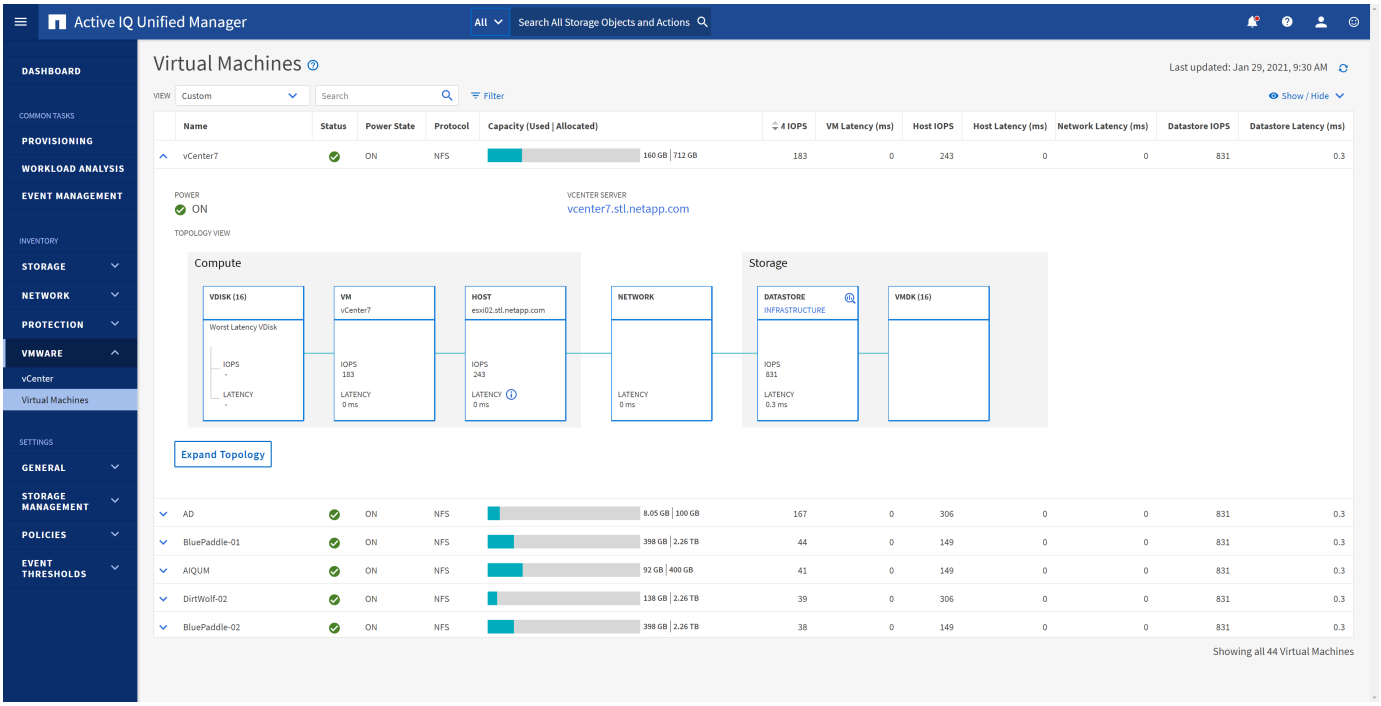
- "Report tecnici sulla sicurezza"
- "Guide per la protezione avanzata"
- "Documentazione del prodotto di sicurezza e crittografia dei dati ONTAP"

Active IQ Unified Manager

Active IQ Unified Manager offre visibilità sulle macchine virtuali dell'infrastruttura virtuale e consente il monitoraggio e la risoluzione dei problemi relativi a storage e performance nell'ambiente virtuale.

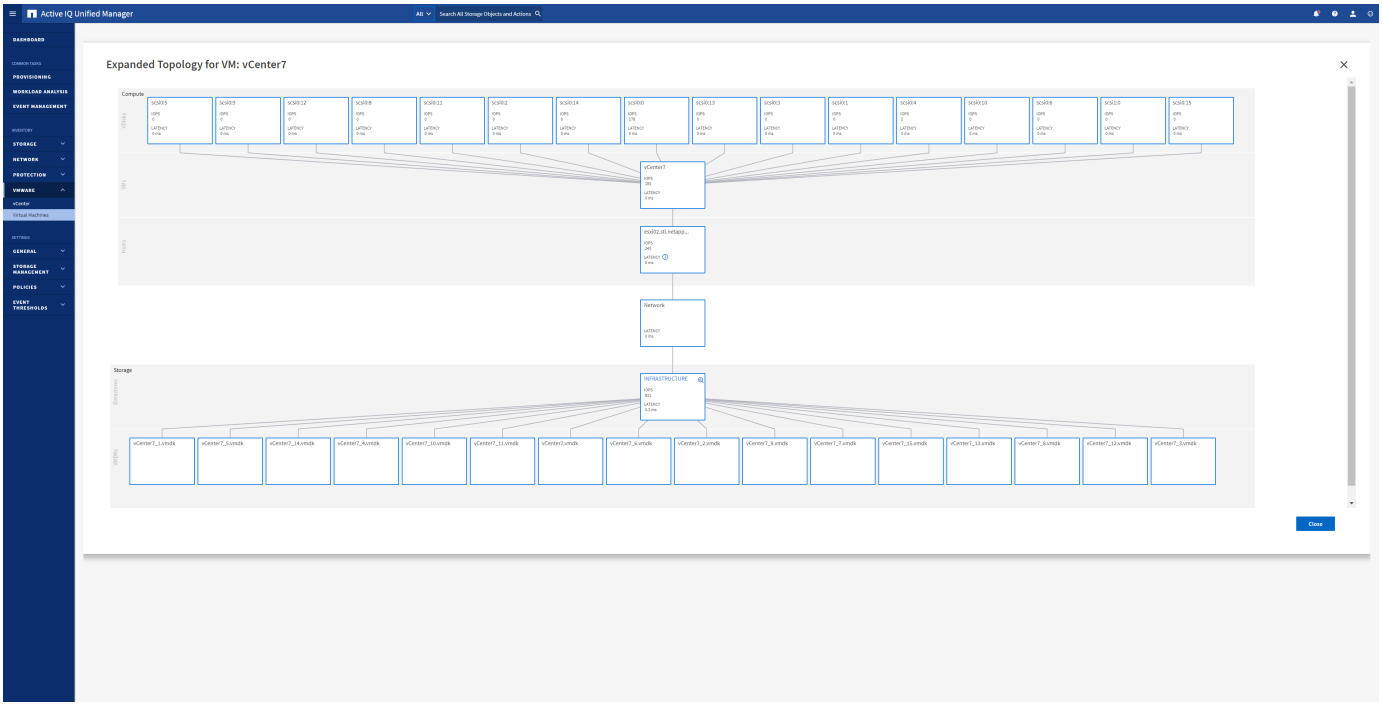
Una tipica implementazione di un'infrastruttura virtuale su ONTAP include diversi componenti distribuiti tra livelli di calcolo, rete e storage. Eventuali ritardi nelle performance in un'applicazione VM potrebbero verificarsi a causa di una combinazione di latenze affrontate dai vari componenti nei rispettivi layer.

La seguente schermata mostra la vista macchine virtuali Active IQ Unified Manager.



Unified Manager presenta il sottosistema sottostante di un ambiente virtuale in una vista topologica per determinare se si è verificato un problema di latenza nel nodo di calcolo, nella rete o nello storage. La vista evidenzia anche l'oggetto specifico che causa il ritardo delle performance per l'adozione di misure correttive e la risoluzione del problema sottostante.

La seguente schermata mostra la topologia espansa di AIQUM.



Gestione basata su criteri di archiviazione e vVol

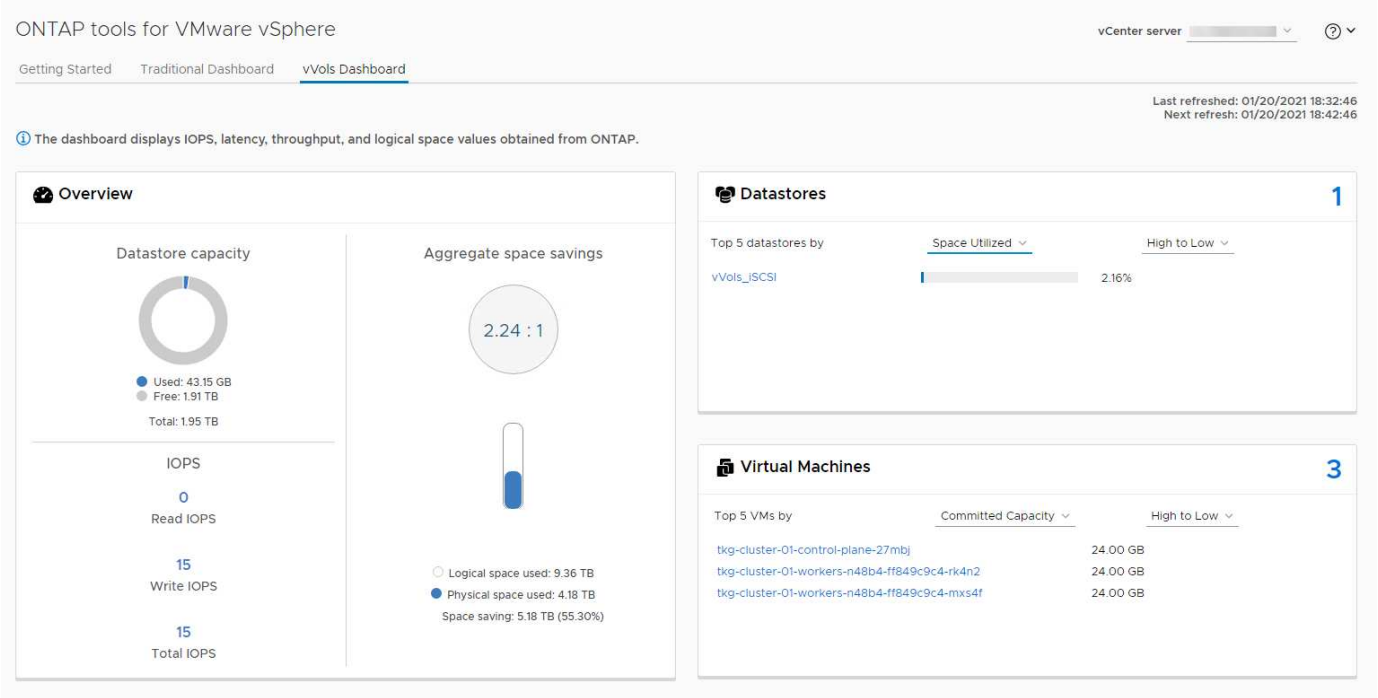
Le API VMware vSphere per Storage Awareness (VASA) consentono a un amministratore dello storage di configurare con facilità i datastore con funzionalità ben definite, consentendo all'amministratore delle macchine virtuali di utilizzare tali dati quando necessario per eseguire il provisioning delle macchine virtuali senza dover interagire tra loro.

Vale la pena di dare un'occhiata a questo approccio per scoprire in che modo può semplificare le operazioni di virtualizzazione dello storage ed evitare un lavoro molto banale.

Prima di VASA, gli amministratori delle macchine virtuali potevano definire policy di storage delle macchine virtuali, ma dovevano collaborare con l'amministratore dello storage per identificare i datastore appropriati, spesso utilizzando documentazione o convenzioni di naming. Con VASA, l'amministratore dello storage può definire una serie di funzionalità di storage, tra cui performance, tiering, crittografia e replica. Un insieme di funzionalità per un volume o un set di volumi viene definito SCP (Storage Capability Profile).

SCP supporta la qualità del servizio minima e/o massima per i vVol di dati di una VM. La QoS minima è supportata solo sui sistemi AFF. Gli strumenti ONTAP per VMware vSphere includono una dashboard che visualizza le performance granulari delle macchine virtuali e la capacità logica per i vVol sui sistemi ONTAP.

La figura seguente mostra i tool ONTAP per il dashboard di VMware vSphere 9.8 vVol.



Una volta definito il profilo di capacità dello storage, è possibile utilizzarlo per eseguire il provisioning delle macchine virtuali utilizzando la policy di storage che ne identifica i requisiti. La mappatura tra il criterio di storage delle macchine virtuali e il profilo di capacità dello storage del datastore consente a vCenter di visualizzare un elenco di datastore compatibili per la selezione. Questo approccio è noto come gestione basata su criteri di storage.

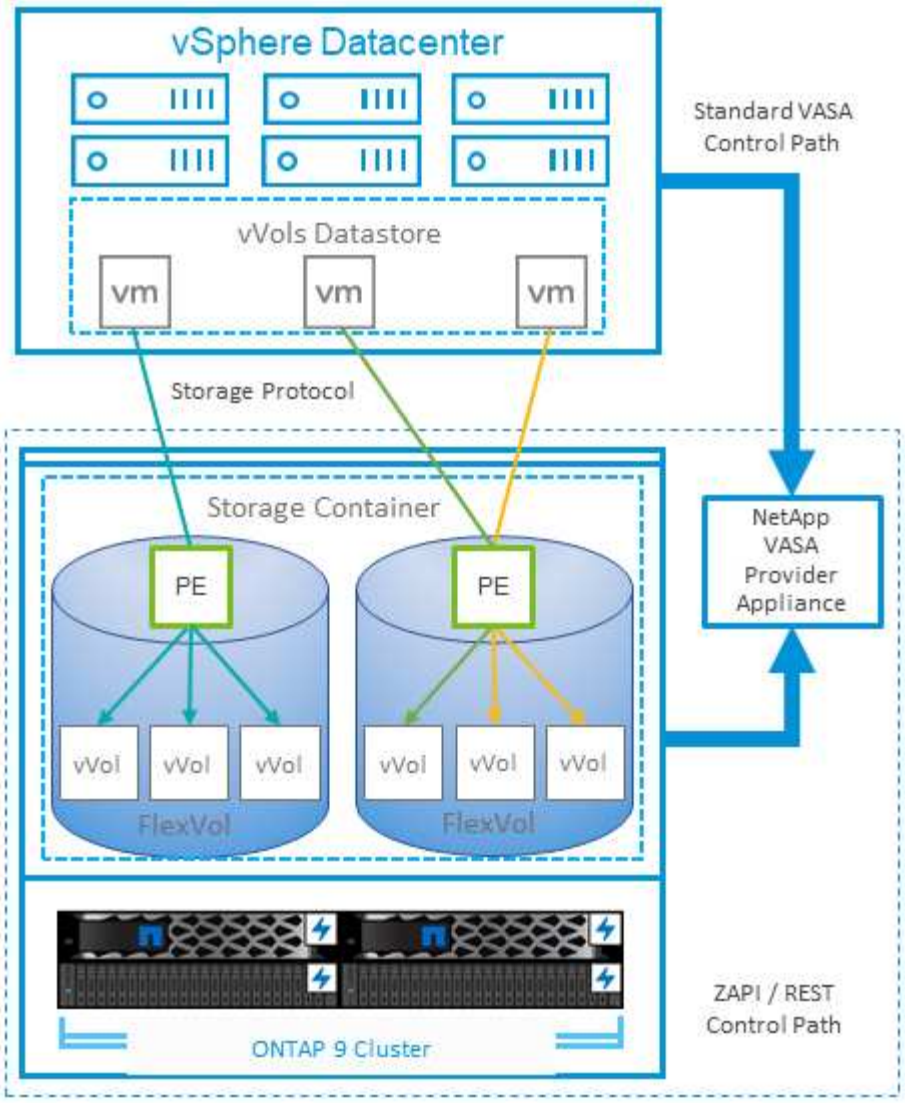
VASA offre la tecnologia per eseguire query sullo storage e restituire un set di funzionalità di storage a vCenter. I vendor provider VASA forniscono la traduzione tra le API e i costrutti del sistema storage e le API VMware comprese da vCenter. Il provider VASA di NetApp per ONTAP viene offerto come parte dei tool

ONTAP per macchina virtuale dell'appliance VMware vSphere, mentre il plug-in vCenter fornisce l'interfaccia per il provisioning e la gestione dei datastore vVol, nonché la capacità di definire profili di funzionalità dello storage (SCP).

ONTAP supporta gli archivi dati VMFS e NFS vVol. L'utilizzo di vVol con datastore SAN offre alcuni dei vantaggi di NFS, come la granularità a livello di macchine virtuali. Di seguito sono riportate alcune Best practice da prendere in considerazione e ulteriori informazioni sono disponibili in "TR-4400":

- Un datastore vVol può essere costituito da più volumi FlexVol su più nodi del cluster. L'approccio più semplice è un singolo datastore, anche quando i volumi hanno funzionalità diverse. SPBM garantisce l'utilizzo di un volume compatibile per la macchina virtuale. Tuttavia, tutti i volumi devono far parte di una singola SVM ONTAP e devono essere accessibili utilizzando un singolo protocollo. È sufficiente una LIF per nodo per ogni protocollo. Evitare di utilizzare più release di ONTAP all'interno di un singolo datastore vVol, poiché le funzionalità dello storage potrebbero variare tra le varie release.
- Utilizza i tool ONTAP per il plug-in VMware vSphere per creare e gestire datastore vVol. Oltre a gestire il datastore e il relativo profilo, crea automaticamente un endpoint del protocollo per accedere ai vVol, se necessario. Se si utilizzano LUN, tenere presente che i LUN PES vengono mappati utilizzando LUN ID 300 e superiori. Verificare che l'impostazione di sistema avanzata dell'host ESXi sia corretta `Disk.MaxLUN` Consente un numero di ID LUN superiore a 300 (il valore predefinito è 1,024). Eseguire questa operazione selezionando l'host ESXi in vCenter, quindi la scheda Configura e trova `Disk.MaxLUN` Nell'elenco delle Advanced System Settings (Impostazioni di sistema avanzate).
- Non installare o migrare il provider VASA, il server vCenter (basato su appliance o Windows) o i tool ONTAP per VMware vSphere in sé su un datastore vVols, perché in tal caso sono dipendenti reciprocamente, limitando la possibilità di gestirli in caso di interruzione dell'alimentazione o di altre interruzioni del data center.
- Eseguire regolarmente il backup della VM del provider VASA. Crea almeno snapshot orarie del datastore tradizionale che contiene il provider VASA. Per ulteriori informazioni sulla protezione e il ripristino del provider VASA, consulta questa sezione "Articolo della Knowledge base".

La figura seguente mostra i componenti di vVol.



VMware Storage Distributed Resource Scheduler

VMware Storage Distributed Resource Scheduler (SDR) è una funzione vSphere che posiziona automaticamente le macchine virtuali in un cluster di datastore in base alla latenza di i/o corrente e all'utilizzo dello spazio.

Quindi, sposta le VM o i VMDK senza interruzioni tra gli archivi dati in un cluster di datastore (noto anche come pod), selezionando il migliore datastore in cui posizionare le VM o i VMDK nel cluster di datastore. Un cluster di datastore è un insieme di datastore simili che vengono aggregati in una singola unità di consumo dal punto di vista dell'amministratore di vSphere.

Quando si utilizzano DSP con strumenti ONTAP per VMware vSphere, è necessario prima creare un datastore con il plug-in, utilizzare vCenter per creare il cluster di datastore e quindi aggiungere il datastore. Una volta creato il cluster di datastore, è possibile aggiungere ulteriori datastore al cluster di datastore direttamente dalla procedura guidata di provisioning nella pagina Dettagli.

Altre Best practice ONTAP per I DSP includono:

- Non utilizzare i DSP a meno che non si disponga di un requisito specifico per farlo.
  - I DSP non sono necessari quando si utilizza ONTAP. GLI SDR non sono consapevoli delle funzionalità

di efficienza dello storage ONTAP come la deduplica e la compressione, per cui potrebbero prendere decisioni non ottimali per l'ambiente in uso.

- GLI SDR non sono a conoscenza delle policy QoS di ONTAP, pertanto potrebbero prendere decisioni che non sono ottimali per le performance.
- GLI SDR non sono a conoscenza delle copie snapshot ONTAP, pertanto potrebbero prendere decisioni che causano una crescita esponenziale delle snapshot. Ad esempio, spostando una macchina virtuale in un altro datastore vengono creati nuovi file nel nuovo datastore, con una conseguente crescita dello snapshot. Ciò vale in particolare per le macchine virtuali con dischi di grandi dimensioni o molte istantanee. Quindi, se la macchina virtuale dovesse essere spostata di nuovo nel datastore originale, lo snapshot nel datastore originale crescerà ulteriormente.

Se si utilizzano gli SDR, prendere in considerazione le seguenti procedure consigliate:

- Tutti gli archivi dati del cluster devono utilizzare lo stesso tipo di storage (ad esempio SAS, SATA o SSD), tutti gli archivi dati VMFS o NFS e avere le stesse impostazioni di replica e protezione.
- Considerare l'utilizzo DEGLI SDR in modalità predefinita (manuale). Questo approccio consente di rivedere i suggerimenti e decidere se applicarli o meno. Tenere presente i seguenti effetti delle migrazioni VMDK:
  - Quando LE SDR spostano i VMDK tra datastore, qualsiasi risparmio di spazio derivante dal cloning o dalla deduplica ONTAP può essere ridotto a seconda della qualità di deduplica o compressione della destinazione.
  - Dopo che LE SDR spostano i VMDK, NetApp consiglia di ricreare gli snapshot nel datastore di origine, poiché lo spazio è altrimenti bloccato dalla VM che è stata spostata.
  - Lo spostamento di VMDK tra datastore sullo stesso aggregato ha pochi benefici e GLI SDR non hanno visibilità su altri carichi di lavoro che potrebbero condividere l'aggregato.

Ulteriori informazioni sugli SDR sono disponibili nella documentazione VMware all'indirizzo "Domande frequenti su Storage DRS".

Host ESXi consigliato e altre impostazioni ONTAP

NetApp ha sviluppato una serie di impostazioni ottimali per l'host ESXi sia per protocolli NFS sia per protocolli a blocchi. Sono inoltre fornite indicazioni specifiche per le impostazioni di multipathing e timeout HBA per un corretto comportamento con ONTAP in base ai test interni di NetApp e VMware.

Questi valori sono facilmente impostabili utilizzando gli strumenti ONTAP per VMware vSphere: Dalla pagina di panoramica degli strumenti ONTAP, scorrere fino alla fine e fare clic su Applica impostazioni consigliate nel portlet conformità host ESXi.

Di seguito sono riportate le impostazioni dell'host consigliate per tutte le versioni di ONTAP attualmente supportate.

Impostazione host	Valore consigliato da NetApp	Riavvio richiesto
Configurazione avanzata ESXi		
VMFS3.HardwareAcceleratedLocking	Mantieni predefinito (1)	No

Impostazione host	Valore consigliato da NetApp	Riavvio richiesto
VMFS3.EnableBlockDelete	Mantenere l'impostazione predefinita (0), ma può essere modificata se necessario. Per ulteriori informazioni, vedere "Recupero di spazio per VMFS5 macchine virtuali"	No
VMFS3.EnableVMFS6Unmap	Mantenere l'impostazione predefinita (1) per ulteriori informazioni, vedere "API VMware vSphere: Integrazione degli array (VAAI)"	No
Impostazioni NFS		
NewSyncInterval	Se non si utilizza vSphere CSI per Kubernetes, impostare come indicato in "VMware KB 386364"	No
NET.TcpipelHeapSize	VSphere 6.0 o versione successiva, impostato su 32. Tutte le altre configurazioni NFS, impostate su 30	Sì
NET.TcpipelHeapMax	Impostato su 512 MB per la maggior parte delle release di vSphere 6.X. Impostare sul valore predefinito (1024MB) per 6.5U3, 6.7U3 e 7,0 o versioni successive.	Sì
NFS.MaxVolumes	VSphere 6,0 o versioni successive, impostare su 256 Tutte le altre configurazioni NFS sono impostate su 64.	No
NFS41.MaxVolumes	VSphere 6,0 o versioni successive, impostare su 256.	No
NFS.MaxQueueDepth[1]	VSphere 6.0 o versione successiva, impostato su 128	Sì
NFS.HeartbeatMaxFailures	Impostare su 10 per tutte le configurazioni NFS	No
NFS.HeartbeatFrequency	Impostato su 12 per tutte le configurazioni NFS	No
NFS.HeartbeatTimeout	Impostare su 5 per tutte le configurazioni NFS.	No
SunRPC.MaxConnPerIP	vSphere da 7.0 a 8.0, impostato su 128. Questa impostazione viene ignorata nelle versioni ESXi successive alla 8.0.	No
Impostazioni FC/FCoE		

Impostazione host	Valore consigliato da NetApp	Riavvio richiesto
Policy di selezione del percorso	Impostare su RR (round robin) quando si utilizzano percorsi FC con ALUA. Impostare su FISSO per tutte le altre configurazioni. L'impostazione di questo valore su RR consente di fornire il bilanciamento del carico in tutti i percorsi attivi/ottimizzati. Il valore FISSO è per le configurazioni precedenti non ALUA e aiuta a prevenire i/o proxy In altre parole, consente di evitare che l'i/o venga collegato all'altro nodo di una coppia ad alta disponibilità (ha) in un ambiente con Data ONTAP in 7-Mode	No
Disk.QFullSampleSize	Impostare su 32 per tutte le configurazioni. L'impostazione di questo valore aiuta a prevenire gli errori di i/O.	No
Disk.QFullThreshold	Impostare su 8 per tutte le configurazioni. L'impostazione di questo valore aiuta a prevenire gli errori di i/O.	No
Timeout HBA FC Emulex	Utilizzare il valore predefinito.	No
Timeout HBA FC QLogic	Utilizzare il valore predefinito.	No
Impostazioni iSCSI		
Policy di selezione del percorso	Impostare su RR (round robin) per tutti i percorsi iSCSI. L'impostazione di questo valore su RR consente di fornire il bilanciamento del carico in tutti i percorsi attivi/ottimizzati.	No
Disk.QFullSampleSize	Impostare su 32 per tutte le configurazioni. L'impostazione di questo valore aiuta a prevenire gli errori di i/O.	No
Disk.QFullThreshold	Impostare su 8 per tutte le configurazioni. L'impostazione di questo valore aiuta a prevenire gli errori di i/O.	No

L'opzione di configurazione avanzata NFS MaxQueueDepth potrebbe non funzionare come previsto quando si utilizzano VMware vSphere ESXi 7.0.1 e VMware vSphere ESXi 7,0.2. Per ulteriori informazioni, fare riferimento "Tastiera VMware 86331"a.

Gli strumenti ONTAP specificano anche alcune impostazioni predefinite durante la creazione di ONTAP

FlexVol Volumes e LUN:

Strumento ONTAP	Impostazione predefinita
Riserva di Snapshot (-percento-spazio-snapshot)	0
Riserva frazionaria (-riserva frazionaria)	0
Access time update (-atime-update)	Falso
Readahead minimo (-min-readahead)	Falso
Istantanee pianificate	Nessuno
Efficienza dello storage	Attivato
Garanzia di volume	Nessuno (con thin provisioning)
Dimensionamento automatico del volume	grow_shrink
Prenotazione di spazio LUN	Disattivato
Allocazione dello spazio del LUN	Attivato

Impostazioni multipath per performance superiori

Sebbene non sia attualmente configurato dagli strumenti ONTAP disponibili, NetApp suggerisce le seguenti opzioni di configurazione:

- Quando si utilizzano sistemi non ASA in ambienti ad alte prestazioni o quando si testano le prestazioni con un singolo datastore LUN, valutare la possibilità di modificare l'impostazione del bilanciamento del carico della policy di selezione del percorso (PSP) round-robin (VMW_PSP_RR) dall'impostazione IOPS predefinita di 1000 a un valore di 1. Vedere"VMware KB 2069356" per maggiori informazioni.
- In vSphere 6.7 Update 1, VMware ha introdotto un nuovo meccanismo di bilanciamento del carico di latenza per il Round Robin PSP. L'opzione di latenza è ora disponibile anche quando si utilizza HPP (High Performance Plugin) con namespace NVMe e con vSphere 8.0u2 e versioni successive, LUN connesse tramite iSCSI e FCP. La nuova opzione considera la larghezza di banda I/O e la latenza del percorso quando seleziona il percorso ottimale per I/O. NetApp consiglia di utilizzare l'opzione di latenza in ambienti con connettività di percorso non equivalente, ad esempio nei casi con più hop di rete su un percorso rispetto a un altro, oppure quando si utilizza un sistema NetApp ASA . Vedere "Modifica dei parametri predefiniti per la latenza Round Robin" per maggiori informazioni.

Documentazione aggiuntiva

Per FCP e iSCSI con vSphere 7, è possibile trovare ulteriori dettagli all'indirizzo "Utilizzo di VMware vSphere 7.x con ONTAP" per FCP e iSCSI con vSphere 8. Per ulteriori dettagli, visitare la pagina "Utilizzo di VMware vSphere 8.x con ONTAP" relativa a NVMe-of con vSphere 7. Per ulteriori dettagli, visitare il sito "Per NVMe-of, ulteriori dettagli sono disponibili nella pagina NVMe-of host Configuration per ESXi 7.x con ONTAP" per NVMe-of con vSphere 8, ulteriori dettagli sono disponibili all'indirizzo "Per NVMe-of, ulteriori dettagli sono disponibili nella pagina NVMe-of host Configuration per ESXi 8.x con ONTAP"

Volumi virtuali (vVol) con strumenti ONTAP 10

Panoramica

ONTAP è stata una soluzione storage leader per gli ambienti VMware vSphere da oltre

vent'anni e continua ad aggiungere funzionalità innovative per semplificare la gestione e ridurre i costi.

Questo documento tratta le funzionalità di ONTAP per i volumi virtuali VMware vSphere (vVol), incluse le informazioni più recenti sui prodotti e i casi di utilizzo, oltre a Best practice e altre informazioni per semplificare l'implementazione e ridurre gli errori.

Questa documentazione sostituisce i report tecnici precedentemente pubblicati *TR-4400: Volumi virtuali VMware vSphere (vVol) con ONTAP*

Le Best practice integrano altri documenti come guide ed elenchi di compatibilità. Sono sviluppati in base a test di laboratorio e a un'ampia esperienza sul campo da parte di tecnici e clienti NetApp. Potrebbero non essere le uniche pratiche che funzionano o sono supportate, ma sono generalmente le soluzioni più semplici che soddisfano le esigenze della maggior parte dei clienti.

Questo documento è stato aggiornato per includere le nuove funzionalità di vVol presenti in vSphere 8,0 update 3, nella versione 10,4 degli strumenti ONTAP e nei nuovi sistemi NetApp ASA.

Panoramica dei volumi virtuali (vVol)

Nel 2012, NetApp ha iniziato a collaborare con VMware per supportare le API vSphere per la consapevolezza dello storage (VASA) per vSphere 5. Questo primo provider VASA consentiva la definizione delle funzionalità di storage in un profilo che poteva essere utilizzato per filtrare i datastore durante il provisioning e per verificare successivamente la conformità con la policy. Nel corso del tempo, questo si è evoluto per aggiungere nuove funzionalità per consentire una maggiore automazione nel provisioning, oltre all'aggiunta di volumi virtuali o vVol, in cui i singoli oggetti storage vengono utilizzati per i file delle macchine virtuali e i dischi virtuali. Questi oggetti potrebbero essere LUN, file e ora con vSphere 8 - namespace NVMe (utilizzati con gli strumenti ONTAP 9.13P2). NetApp ha lavorato a stretto contatto con VMware come partner di riferimento per vVol rilasciati con vSphere 6 nel 2015 e ancora come partner di progettazione per vVol che utilizzano NVMe over Fabrics in vSphere 8. NetApp continua a migliorare vVol per sfruttare le più recenti funzionalità di ONTAP.

Esistono diversi componenti di cui tenere conto:

Provider VASA

Questo è il componente software che gestisce la comunicazione tra VMware vSphere e il sistema storage. Per ONTAP, il provider VASA viene eseguito in un'appliance nota come tool ONTAP per VMware vSphere (in breve, strumenti ONTAP). Gli strumenti ONTAP includono anche un plugin vCenter, un adattatore per la replica dello storage (SRA) per VMware Site Recovery Manager e un server API REST per la creazione di automazione. Una volta configurati e registrati gli strumenti ONTAP con vCenter, non è più necessario interagire direttamente con il sistema ONTAP, poiché quasi tutte le esigenze di storage possono essere gestite direttamente dall'interfaccia utente di vCenter o tramite l'automazione delle API REST.

Punto terminale del protocollo (PE)

L'endpoint del protocollo è un proxy per i/o tra gli host ESXi e il datastore vVols. Il provider ONTAP VASA crea automaticamente questi elementi, scegliendo una LUN endpoint di protocollo (4MB GB) per volume FlexVol del datastore vVol o un punto di montaggio NFS per interfaccia NFS (LIF) sul nodo storage che ospita un volume FlexVol nel datastore. L'host ESXi monta questi endpoint di protocollo direttamente piuttosto che singoli LUN vVol e file di dischi virtuali. Non è necessario gestire gli endpoint del protocollo poiché vengono creati, montati, rimossi ed eliminati automaticamente dal provider VASA, insieme a eventuali gruppi di interfacce o policy di esportazione necessari.

Virtual Protocol Endpoint (VPE)

Novità di vSphere 8: Quando si utilizza NVMe over Fabrics (NVMe-of) con vVol, il concetto di endpoint del protocollo non è più rilevante in ONTAP. Al contrario, l'host ESXi crea automaticamente un'istanza di PE virtuale per ciascun gruppo ANA non appena viene accesa la prima macchina virtuale. ONTAP crea automaticamente gruppi ANA per ogni volume FlexVol utilizzato dall'archivio dati.

Un ulteriore vantaggio dell'utilizzo di NVMe-of per vVol è che non sono richieste di bind da parte del provider VASA. L'host ESXi gestisce invece la funzionalità di binding vVol internamente in base a VPE. In questo modo si riduce l'opportunità di un vVol bind storm di impatto sul servizio.

Per ulteriori informazioni, vedere "NVMe e volumi virtuali" acceso "vmware.com"

Datastore di volumi virtuali

| il datastore Virtual Volume è una rappresentazione logica del datastore di un contenitore vVol creato e gestito da un provider VASA. Il container rappresenta un pool di capacità di storage fornito dai sistemi storage gestiti dal provider VASA. Gli strumenti ONTAP supportano l'allocazione di più volumi FlexVol (noti come volumi di backup) a un singolo datastore vVols e questi datastore vVols possono estendersi su più nodi in un cluster ONTAP, combinando sistemi flash e ibridi con funzionalità diverse. L'amministratore può creare nuovi volumi FlexVol utilizzando la procedura guidata di provisioning o l'API REST oppure selezionare volumi FlexVol pre-creati per il backup dello storage, se disponibili.

Volumi virtuali (vVol)

I vVol sono i file e i dischi della macchina virtuale memorizzati nel datastore vVols. Il termine vVol (singolo) si riferisce a un singolo file, LUN o namespace specifico. ONTAP crea spazi dei nomi NVMe, LUN o file a seconda del protocollo utilizzato dal datastore. Esistono diversi tipi diversi di vVol; i più comuni sono Config (l'unico con VMFS su di esso, contiene file di metadati come il file VMX della VM), dati (disco virtuale o VMDK) e Swap (creato quando la VM è accesa). I vVol protetti dalla crittografia VMware VM saranno di tipo Altro. La crittografia di VMware VM non deve essere confusa con la crittografia aggregata o del volume ONTAP.

Gestione basata su criteri

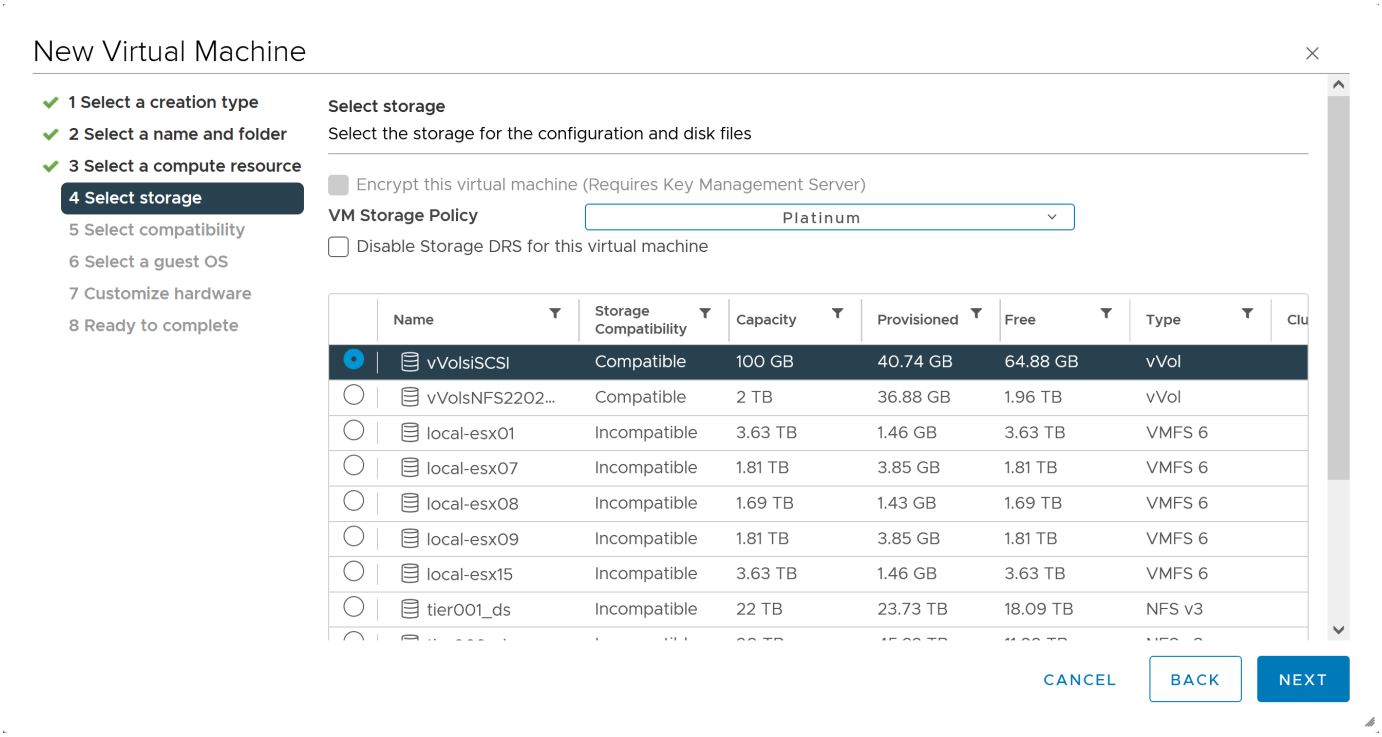
Le API VMware vSphere per la consapevolezza dello storage (VASA) semplificano l'utilizzo da parte di un amministratore delle macchine virtuali delle funzionalità di storage necessarie per il provisioning delle macchine virtuali senza dover interagire con il proprio team di storage. Prima di VASA, gli amministratori delle macchine virtuali potevano definire policy di storage delle macchine virtuali, ma dovevano collaborare con i loro amministratori dello storage per identificare i datastore appropriati, spesso utilizzando documentazione o convenzioni di naming. Con VASA, gli amministratori di vCenter con le autorizzazioni appropriate possono definire una serie di funzionalità di storage che gli utenti di vCenter possono utilizzare per eseguire il provisioning delle macchine virtuali. Il mapping tra la policy dello storage delle macchine virtuali e le funzionalità del datastore consente a vCenter di visualizzare un elenco di datastore compatibili da selezionare, oltre a consentire ad altre tecnologie come aria (precedentemente nota come vRealize) Automation o Tanzu Kubernetes Grid di selezionare automaticamente lo storage da una policy assegnata. Questo approccio è noto come gestione basata su criteri di storage. Mentre le regole del provider VASA e le policy storage delle macchine virtuali possono anche essere utilizzate con i datastore tradizionali, la nostra attenzione si concentra sui datastore vVol.

Policy di storage delle VM

I criteri di storage delle macchine virtuali vengono creati in vCenter in Criteri e profili. Per i vVol, creare un set di regole utilizzando le regole del provider del tipo di storage NetApp vVols. Gli strumenti ONTAP 10.X ora offrono un approccio più semplice rispetto ai tool ONTAP 9.X, consentendo di specificare direttamente gli attributi di storage nel criterio di storage della VM.

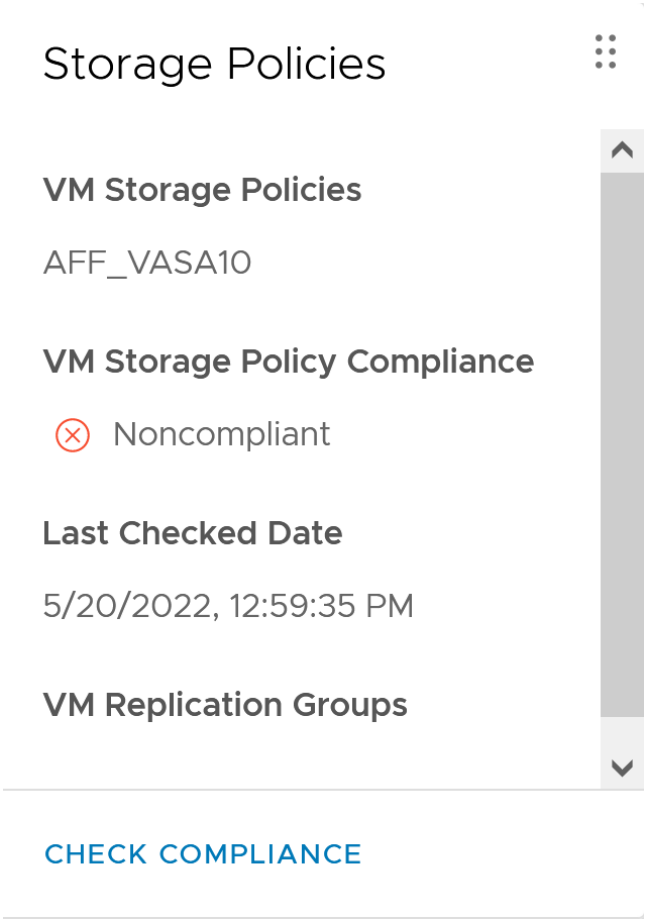
Come menzionato in precedenza, l'utilizzo delle policy può contribuire a semplificare l'attività di provisioning di una macchina virtuale o di un VMDK. Basta selezionare una policy appropriata e il provider VASA mostrerà i datastore vVol che supportano tale policy e posizioneranno il vVol in un singolo FlexVol volume conforme.

Implementare la macchina virtuale utilizzando i criteri di storage



Una volta eseguito il provisioning di una VM, il provider VASA continuerà a verificare la conformità e avviserà l'amministratore della VM tramite un allarme in vCenter quando il volume di backup non è più conforme alla policy.

Conformità delle policy di storage delle macchine virtuali



Supporto vVol NetApp

ONTAP ha supportato la specifica VASA dalla sua versione iniziale nel 2012. Sebbene altri sistemi storage NetApp possano supportare VASA, questo documento si concentra sulle versioni attualmente supportate di ONTAP 9.

ONTAP

Oltre a ONTAP 9 su sistemi AFF, ASA e FAS, NetApp supporta i carichi di lavoro VMware su ONTAP Select, Amazon FSX per NetApp con VMware Cloud su AWS, la soluzione Azure NetApp Files con Azure VMware, Cloud Volumes Service con Google Cloud VMware Engine e NetApp Private Storage in Equinix, tuttavia, le funzionalità specifiche possono variare in base al provider di servizi e alla connettività di rete disponibile. È inoltre disponibile l'accesso dai guest vSphere ai dati memorizzati in tali configurazioni e a Cloud Volumes ONTAP.

Al momento della pubblicazione, gli ambienti hyperscaler sono limitati solo agli archivi dati NFS v3 tradizionali, pertanto i vVol sono disponibili solo con sistemi ONTAP on-premise o con sistemi connessi al cloud che offrono la funzionalità completa di sistemi on-premise come quelli ospitati da partner e provider di servizi NetApp in tutto il mondo.

Per ulteriori informazioni su ONTAP, vedere "Documentazione del prodotto ONTAP"

Per ulteriori informazioni sulle Best practice di ONTAP e VMware vSphere, vedere "TR-4597"

Vantaggi dell'utilizzo di vVol con ONTAP

Quando nel 2015 VMware ha introdotto il supporto per vVol con VASA 2,0, lo ha descritto come "un framework

di integrazione e gestione che offre un nuovo modello operativo per lo storage esterno (SAN/NAS)". Questo modello operativo offre diversi vantaggi insieme allo storage ONTAP.

Gestione basata su criteri

Come descritto nella sezione 1,2, la gestione basata su criteri consente di eseguire il provisioning delle macchine virtuali e di gestirle successivamente utilizzando criteri predefiniti. Questo può aiutare le operazioni IT in diversi modi:

- **Aumentare la velocità.** i tool ONTAP eliminano il requisito per l'amministratore di vCenter di aprire i ticket con il team di storage per le attività di provisioning dello storage. Tuttavia, i ruoli RBAC dei tool ONTAP in vCenter e nel sistema ONTAP consentono ancora ai team indipendenti (come i team di storage) o alle attività indipendenti dello stesso team limitando l'accesso a funzioni specifiche, se necessario.
- **Provisioning più intelligente.** le funzionalità del sistema di storage possono essere esposte attraverso le API VASA, consentendo ai flussi di lavoro di provisioning di sfruttare funzionalità avanzate senza che l'amministratore delle macchine virtuali debba comprendere come gestire il sistema di storage.
- **Provisioning più rapido.** diverse funzionalità di storage possono essere supportate in un singolo datastore e selezionate automaticamente in base alla policy della macchina virtuale.
- **Evitare errori.** le policy di storage e macchine virtuali vengono sviluppate in anticipo e applicate in base alle necessità senza dover personalizzare lo storage ogni volta che viene eseguito il provisioning di una macchina virtuale. Gli allarmi di compliance vengono generati quando le funzionalità dello storage si scostano dalle policy definite. Come accennato in precedenza, gli SCP rendono il provisioning iniziale prevedibile e ripetibile, mentre basare le policy di storage delle macchine virtuali sugli SCP garantisce un posizionamento preciso.
- **Migliore gestione della capacità.** Gli strumenti VASA e ONTAP consentono di vedere la capacità dello storage fino al singolo livello di aggregato, se necessario, e fornire più livelli di avviso in caso di esaurimento della capacità.

Gestione granulare delle macchine virtuali nella moderna SAN

I sistemi storage SAN che utilizzano Fibre Channel e iSCSI sono stati i primi ad essere supportati da VMware per ESX, ma non hanno la capacità di gestire singoli file e dischi VM dal sistema storage. Al contrario, vengono forniti i LUN e VMFS gestisce i singoli file. Questo rende difficile per il sistema storage gestire direttamente le performance, la clonazione e la protezione dello storage delle singole macchine virtuali. VVol offre una granularità dello storage di cui già godono i clienti che utilizzano lo storage NFS, con le solide funzionalità SAN ad alte performance di ONTAP.

Ora, con gli strumenti vSphere 8 e ONTAP per VMware vSphere 9.12 e versioni successive, gli stessi controlli granulari utilizzati da vVol per i protocolli basati su SCSI legacy sono ora disponibili nella MODERNA SAN Fibre Channel che utilizza NVMe over Fabrics per ottenere performance ancora maggiori su larga scala. Con vSphere 8.0 update 1, è ora possibile implementare una soluzione NVMe end-to-end completa utilizzando vVol senza alcuna traduzione i/o nello stack di storage dell'hypervisor.

Maggiori funzionalità di offload dello storage

Mentre VAAI offre una varietà di operazioni che vengono trasferite allo storage, ci sono alcune lacune che vengono affrontate dal provider VASA. SAN VAAI non è in grado di trasferire le snapshot gestite da VMware al sistema storage. NFS VAAI è in grado di trasferire le snapshot gestite da macchine virtuali, ma esistono dei limiti per una macchina virtuale con snapshot native dello storage. Poiché i vVol utilizzano LUN, spazi dei nomi o file singoli per i dischi delle macchine virtuali, ONTAP può clonare in modo rapido ed efficiente i file o le LUN per creare snapshot granulari delle macchine virtuali che non richiedono più file delta. Inoltre, NFS VAAI non supporta operazioni di offload dei cloni per le migrazioni vMotion di storage a caldo (attivate). La macchina virtuale deve essere spenta per consentire l'offload della migrazione quando si utilizza VAAI con datastore

NFS tradizionali. Il provider VASA negli strumenti ONTAP consente cloni quasi istantanei ed efficienti in termini di storage per le migrazioni a caldo e a freddo e supporta anche copie quasi istantanee per le migrazioni tra volumi di vVol. Grazie a questi significativi vantaggi in termini di efficienza dello storage, è possibile sfruttare al meglio i carichi di lavoro vVol in base a. "Garanzia di efficienza" programma. Allo stesso modo, se i cloni cross-volume con VAAI non soddisfano i tuoi requisiti, sarai in grado di risolvere le sfide per il tuo business grazie ai miglioramenti nell'esperienza di copia con i vVol.

Casi di utilizzo comuni per i vVol

Oltre a questi vantaggi, vediamo anche questi casi di utilizzo comuni per lo storage vVol:

- **Provisioning su richiesta delle VM**
  - Cloud privato o provider di servizi IaaS.
  - Sfrutta automazione e orchestrazione tramite la suite aria (in precedenza vRealize), OpenStack e così via.
- **Dischi di prima classe (FCD)**
  - VMware Tanzu Kubernetes Grid [TKG] volumi persistenti.
  - Fornire servizi di Amazon EBS attraverso una gestione indipendente del ciclo di vita VMDK.
- **Provisioning on-demand delle macchine virtuali temporanee**
  - Laboratori di test/sviluppo
  - Ambienti di training

Vantaggi comuni con vVol

Se utilizzato a pieno vantaggio, come nei casi di utilizzo precedenti, i vVol forniscono i seguenti miglioramenti specifici:

- I cloni vengono creati rapidamente all'interno di un singolo volume o su più volumi in un cluster ONTAP, un vantaggio rispetto ai cloni abilitati VAAI tradizionali. Sono inoltre efficienti in termini di storage. I cloni all'interno di un volume utilizzano il clone del file ONTAP, simile ai volumi FlexClone, e memorizzano solo le modifiche dal file/LUN/namespace vVol di origine. In questo modo, le macchine virtuali a lungo termine per la produzione o altri scopi applicativi vengono create rapidamente, occupano poco spazio e possono beneficiare della protezione a livello di macchine virtuali (utilizzando il plug-in NetApp SnapCenter per VMware vSphere, le snapshot gestite da VMware o il backup VADP) e della gestione delle performance (con QoS ONTAP). I cloni cross volume sono molto più veloci con vVol che con VAAI becuase con VASA, possiamo creare il clone e permetterne l'accesso alla destinazione prima del completamento della copia. I blocchi di dati vengono copiati come processo in background per popolare il vVol di destinazione. In questo modo, lo spostamento delle LUN senza interruzioni di ONTAP funziona anche con le LUN tradizionali.
- I vVol sono la tecnologia di storage ideale quando si utilizza TKG con vSphere CSI, fornendo classi di storage e capacità discrete gestite dall'amministratore di vCenter.
- Amazon EBS-like Services può essere fornito attraverso FCD perché un FCD VMDK, come suggerisce il nome, è un cittadino di prima classe in vSphere e ha un ciclo di vita che può essere gestito in modo indipendente separato dalle macchine virtuali a cui potrebbe essere collegato.

Elenco di controllo

Utilizzare questo elenco di controllo per garantire una distribuzione corretta (aggiornato per 10,3 e versioni successive).

1 Pianificazione iniziale

- Prima di iniziare l'installazione, è necessario controllare il "Tool di matrice di interoperabilità (IMT)" per assicurarsi che la distribuzione sia stata certificata.
- Determina le dimensioni e il tipo di configurazione degli strumenti ONTAP richiesti dall'ambiente. Per ulteriori informazioni, fare riferimento alla "Limiti di configurazione per l'implementazione dei tool ONTAP per VMware vSphere" .
- Determina se utilizzi SVM multitenant o consenti un accesso completo al cluster. Se utilizzi SVM multitenant, dovrai avere un LIF di gestione SVM su ciascuna SVM da utilizzare. Questa LIF deve essere raggiungibile tramite la porta 443 dagli strumenti ONTAP.
- Determinare se si utilizzerà Fibre Channel (FC) per la connettività dello storage. In tal caso, occorre "configurare lo zoning" utilizzare gli switch FC per abilitare la connettività tra gli host ESXi e le LIF FC della SVM.
- Determinare se si intende utilizzare l'adattatore di replica dello storage (SRA) degli strumenti ONTAP per VMware Site Recovery Manager (SRM) o Live Site Recovery (VLSR). In tal caso, sarà necessario accedere all'interfaccia di gestione del server SRM/VLSR per installare SRA.
- Se si utilizza la replica di SnapMirror gestita dagli strumenti ONTAP (inclusa, ma non solo, la sincronizzazione attiva di SnapMirror), l'amministratore di ONTAP deve "Creare una relazione di peer cluster in ONTAP" e "Creare una relazione di peer intercluster SVM in ONTAP" prima di poter utilizzare gli strumenti ONTAP con SnapMirror.
- "Scarica" ONTAP mette a disposizione OVA e, se necessario, il file SRA tar.gz.

2 Fornire indirizzi IP e record DNS

- Richiedere le seguenti informazioni IP al proprio team di rete. Sono necessari i primi tre indirizzi IP; il nodo due e il nodo tre sono utilizzati per implementazioni di ha (high Availability) scale-out. I record host DNS sono obbligatori e tutti i nomi dei nodi e tutti gli indirizzi devono trovarsi sulla stessa VLAN e subnet.
- ONTAP tools indirizzo applicazione _____ . _____ . _____ . _____
- Indirizzo servizi interni _____ . _____ . _____ . _____
- Il nome host DNS del nodo uno _____
- L'indirizzo IP del nodo uno _____ . _____ . _____ . _____
- Maschera di sottorete _____ . _____ . _____ . _____
- Gateway predefinito _____ . _ . _ . __
- Server DNS 1 _____ . _____ . _____ . _____
- Server DNS 2 _____ . _____ . _____ . _____
- Dominio di ricerca DNS _____
- Nome host DNS del nodo due (opzionale) _____
- Indirizzo IP del nodo due (opzionale) _____ . _____ . _____ . _____
- Nome host DNS del nodo tre (opzionale) _____
- Indirizzo IP del nodo tre (opzionale) _____ . _____ . _____ . _____
- Creare record DNS per tutti gli indirizzi IP indicati sopra.

3 Configurazione del firewall di rete

- Aprire le porte richieste per gli indirizzi IP sopra indicati nel firewall di rete. Per l'aggiornamento più recente, consultare la sezione "Requisiti delle porte".

4 Flash in modo Smart

- È necessario un datastore su un dispositivo storage condiviso. In alternativa, è possibile utilizzare una libreria di contenuti nello stesso datastore del nodo uno per semplificare la clonazione rapida del modello con VAAI.
- Libreria di contenuti (richiesta solo per ha) _____
- Nodo 1 datastore _____
- Datastore nodo due (opzionale, ma consigliato per ha) ___________
- Datastore nodo tre (opzionale, ma consigliato per ha) ____

5 Distribuire l'OVA

- Si noti che questo passaggio può richiedere fino a 45 minuti
- "Distribuire l'OVA" Utilizzando il client vSphere.
- Nel passaggio 3 della distribuzione OVA, selezionare l'opzione "Customize this virtual machine's hardware" (Personalizza l'hardware di questa macchina virtuale) e impostare quanto segue nel passaggio 10:
- "Attiva aggiunta a caldo CPU"
- "Hot plug memoria"

6 Aggiungere vCenter agli strumenti ONTAP

- "Aggiungere istanze di vCenter Server" In ONTAP Tools Manager.

7 Aggiungi i backend di storage ai tool ONTAP

- "Configurare i ruoli e i privilegi degli utenti ONTAP" Utilizzo del file JSON incluso se non si utilizza admin.
- Se utilizzato se si utilizza la multi-tenancy dello storage con SVM
- "cluster integrati" In ONTAP Tools Manager e associarli a vCenter.
- "SVM integrate" Negli strumenti ONTAP, l'interfaccia utente vCenter.
- In caso contrario* utilizzando SVM multitenant
- "cluster integrati" Direttamente nell'interfaccia utente vCenter degli strumenti ONTAP. In alternativa, in questo scenario è possibile aggiungere SVM direttamente quando non si utilizzano i vVol.

8 Configurare i servizi delle appliance (opzionali)

- Per utilizzare vVol, è necessario prima "Modificare le impostazioni dell'appliance e attivare il servizio VASA". Allo stesso tempo, rivedere i due elementi seguenti.

- Se si prevede di utilizzare vVol in produzione, "abilita l'alta disponibilità" con i due indirizzi IP opzionali sopra indicati.
- Se si prevede di utilizzare l'adattatore di replica dello storage (SRA, Storage Replication Adapter) degli strumenti ONTAP per VMware Site Recovery Manager o Live Site Recovery, "Attivare i servizi SRA".

9 **Certificati (opzionali)**

- Per VMware, i certificati CA firmati sono necessari se si utilizzano vVol con più vCenter.
- Servizi VASA _____
- Servizi amministrativi _____\

10 **Altre attività successive all'implementazione**

- Crea regole di affinità per le macchine virtuali in un'implementazione ha.
- Se si utilizza l'ha, lo storage vMotion si nodi due e tre per separare i datastore (facoltativo, ma consigliato).
- "utilizzare gestisci certificati" In ONTAP Tools Manager per installare tutti i certificati CA firmati richiesti.
- Se SRA è stato abilitato per SRM/VLSR per proteggere i datastore tradizionali, "Configurare SRA sull'appliance VMware Live Site Recovery".
- Configurare i backup nativi per "RPO prossimo allo zero".
- Configurare backup regolari su altri supporti di archiviazione.

Utilizzo di vVol con ONTAP

La chiave per l'utilizzo di vVol con NetApp sono i tool ONTAP per VMware vSphere, che funge da interfaccia provider VASA (vSphere API for Storage Awareness) per i sistemi ONTAP 9 di NetApp.

Gli strumenti ONTAP includono inoltre estensioni dell'interfaccia utente vCenter, servizi API REST, adattatori di replica storage per VMware Site Recovery Manager/Live Site Recovery, strumenti di monitoraggio e configurazione host e una serie di report che consentono di gestire al meglio l'ambiente VMware.

Prodotti e documentazione

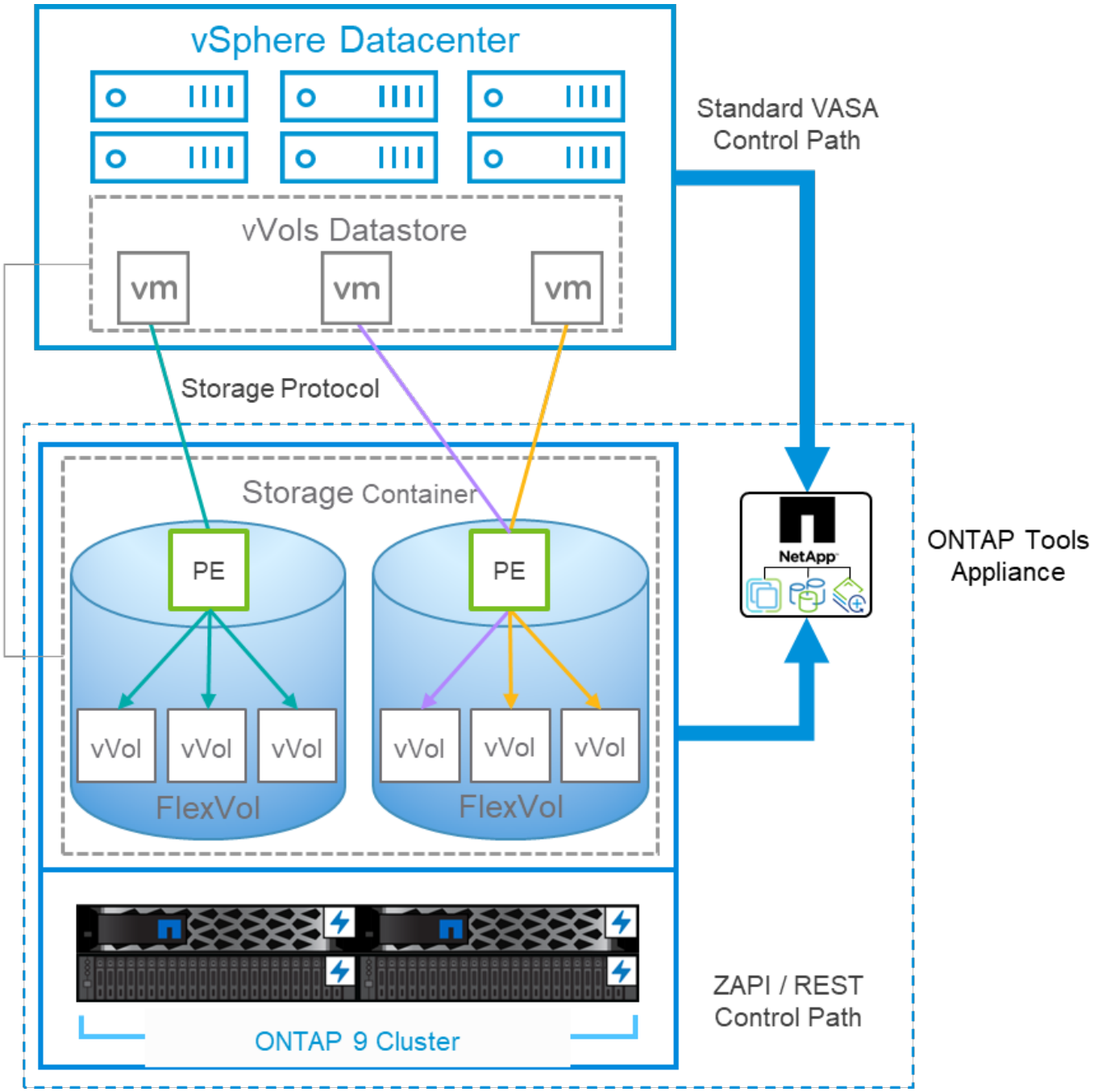
La licenza di ONTAP One include tutte le licenze necessarie per l'utilizzo di vVol con i sistemi ONTAP. L'unico requisito aggiuntivo è il ONTAP gratuito strumenti OVA, che agisce come il fornitore VASA. In un ambiente vVol, il software VASA Provider traduce le funzionalità di array in attributi basati su policy che possono essere sfruttati attraverso le API VASA senza che l'amministratore vSphere debba sapere come le funzionalità vengono gestite dietro le quinte. Ciò consente un consumo dinamico della capacità dello storage allocata in base alle policy, eliminando la necessità di creare manualmente datastore tradizionali e di gestire i rispettivi tassi di consumo dello storage individuale. In breve, i vVol eliminano tutta la complessità della gestione dello storage Enterprise e lo astraggono dall'amministratore vSphere, consentendo loro di concentrarsi sul layer di virtualizzazione.

Per i clienti che utilizzano VMware Cloud Foundation con vSAN, è possibile aggiungere vVol a qualsiasi dominio di gestione o carico di lavoro come storage supplementare. VVol si integra perfettamente con vSAN attraverso un framework di gestione comune basato su criteri di storage.

La famiglia di strumenti ONTAP di nuova generazione release 10 modernizza le funzionalità precedenti con

un'architettura scalabile, containerizzata e basata su microservizi, che può essere implementata tramite una semplice appliance in formato OVA su ESXi. Strumenti ONTAP 10 combina tutte le funzionalità di tre precedenti appliance e prodotti in un'unica implementazione. Per la gestione di vVol, userai le estensioni intuitive dell'interfaccia utente di vCenter o le API REST per il provider VASA degli strumenti ONTAP. Tenere presente che il componente SRA è destinato a datastore tradizionali; VMware Site Recovery Manager non utilizza SRA per vVol.

ONTAP mette a disposizione l'architettura provider VASA quando si utilizza iSCSI o FCP con sistemi unificati



Installazione del prodotto

Per le nuove installazioni, implementa l'appliance virtuale nel tuo ambiente vSphere. Una volta implementato, potrai accedere all'interfaccia utente del manager o utilizzare le API REST per scalare in verticale o in orizzontale l'implementazione, gli vCenter integrati (che registrano il plug-in con vCenter), i sistemi storage

integrati e associare i sistemi storage ai vCenter. L'integrazione dei sistemi storage nell'interfaccia utente del gestore dei tool ONTAP e l'associazione dei cluster ai vCenter sono necessari solo se intendi utilizzare la multi-tenancy sicura con SVM dedicate, altrimenti potrai semplicemente integrare i cluster storage desiderati nelle estensioni dell'interfaccia utente di vCenter dei tool ONTAP o utilizzando le API REST.

Fare riferimento a "Implementazione dello storage vVol" nel presente documento, o "Tool ONTAP per la documentazione di VMware vSphere".

È consigliabile archiviare gli strumenti ONTAP e le appliance vCenter sui tradizionali datastore NFS o VMFS, in modo da evitare conflitti di interdipendenza. Poiché sia i tool vCenter che ONTAP devono comunicare tra loro durante le operazioni dei vVol, non installare o spostare le appliance dei tool ONTAP o le appliance vCenter Server (VCSA) nello storage vVol che stanno gestendo. In questo caso, il riavvio delle appliance per gli strumenti vCenter o ONTAP può causare un'interruzione dell'accesso al piano di controllo e l'impossibilità di avviare l'appliance.

Gli aggiornamenti in-place degli strumenti ONTAP sono supportati utilizzando il file ISO di aggiornamento disponibile per il download all'indirizzo "Tool ONTAP per VMware vSphere 10 - Download" sul sito di supporto NetApp (è richiesto l'accesso). Seguire le "Aggiornamento dai tool ONTAP per VMware vSphere 10.x alla 10,3" istruzioni della guida per aggiornare l'apparecchio. È inoltre possibile eseguire un aggiornamento affiancato dagli strumenti ONTAP 9,13 a 10,3. Fare riferimento a "Migrazione dai tool ONTAP per VMware vSphere 9.x a 10,3" per un'analisi più approfondita dell'argomento.

Per il dimensionamento dell'appliance virtuale e la comprensione dei limiti di configurazione, fare riferimento alla sezione "Limiti di configurazione per l'implementazione dei tool ONTAP per VMware vSphere"

Documentazione del prodotto

La seguente documentazione è disponibile per facilitare l'implementazione degli strumenti ONTAP.

"Tool ONTAP per la documentazione di VMware vSphere"

Inizia subito

- "Note di rilascio"
- "Panoramica sui tool ONTAP per VMware vSphere"
- "Implementare gli strumenti ONTAP"
- "Aggiornare i tool ONTAP"

Utilizzare gli strumenti ONTAP

- "Eseguire il provisioning degli archivi dati"
- "Configurare il controllo degli accessi in base al ruolo"
- "Configurare la disponibilità elevata"
- "Modificare le impostazioni dell'host ESXi"

Proteggere e gestire i datastore

- "Configurare vSphere Metro Storage Cluster (vMSC) utilizzando gli strumenti ONTAP e la sincronizzazione attiva SnapMirror"
- "Proteggere le macchine virtuali" Con SRM

- "Monitorare cluster, datastore e macchine virtuali"

Dashboard del provider VASA

Il provider VASA include una dashboard con informazioni su performance e capacità per le singole VM vVol. Queste informazioni provengono direttamente da ONTAP per i file e le LUN di vVol, inclusi latenza, IOPS, throughput e altro ancora. È attivata per impostazione predefinita quando si utilizzano tutte le versioni attualmente supportate di ONTAP 9. Si noti che dopo la configurazione iniziale possono essere necessari fino a 30 minuti affinché i dati popolino la dashboard.

Altre Best practice

L'utilizzo di ONTAP vVol con vSphere è semplice e segue i metodi vSphere pubblicati (per la versione di ESXi in uso, vedere utilizzo dei volumi virtuali in vSphere Storage nella documentazione VMware). Di seguito sono riportate alcune procedure aggiuntive da prendere in considerazione in combinazione con ONTAP.

Limiti

In generale, ONTAP supporta i limiti vVol definiti da VMware (vedere pubblicato "Valori massimi di configurazione"). Verificare sempre la "NetApp Hardware Universe" presenza di limiti aggiornati su numero e dimensioni di LUN, namespace e file.

Utilizzare i tool ONTAP per le estensioni dell'interfaccia utente di VMware vSphere o le API REST per eseguire il provisioning degli archivi dati vVol e degli endpoint del protocollo.

Anche se è possibile creare datastore vVol con l'interfaccia generale vSphere, utilizzando gli strumenti ONTAP sarà possibile creare automaticamente gli endpoint di protocollo in base alle necessità e creare volumi FlexVol (non richiesti con ASA R2) utilizzando le Best practice ONTAP. È sufficiente fare clic con il pulsante destro del mouse sull'host/cluster/data center, quindi selezionare *ONTAP tools* e *provisioning datastore*. Da qui, è sufficiente scegliere le opzioni vVol desiderate nella procedura guidata.

Non memorizzare mai l'appliance ONTAP Tools o l'appliance vCenter Server (VCSA) su un datastore vVol gestito.

Questo può risultare in una "situazione uova e pollo" se è necessario riavviare gli apparecchi perché non sarà in grado di riassociare i loro vVol durante il riavvio. È possibile memorizzarli in un datastore vVol gestito da un diverso tool ONTAP e da una distribuzione vCenter.

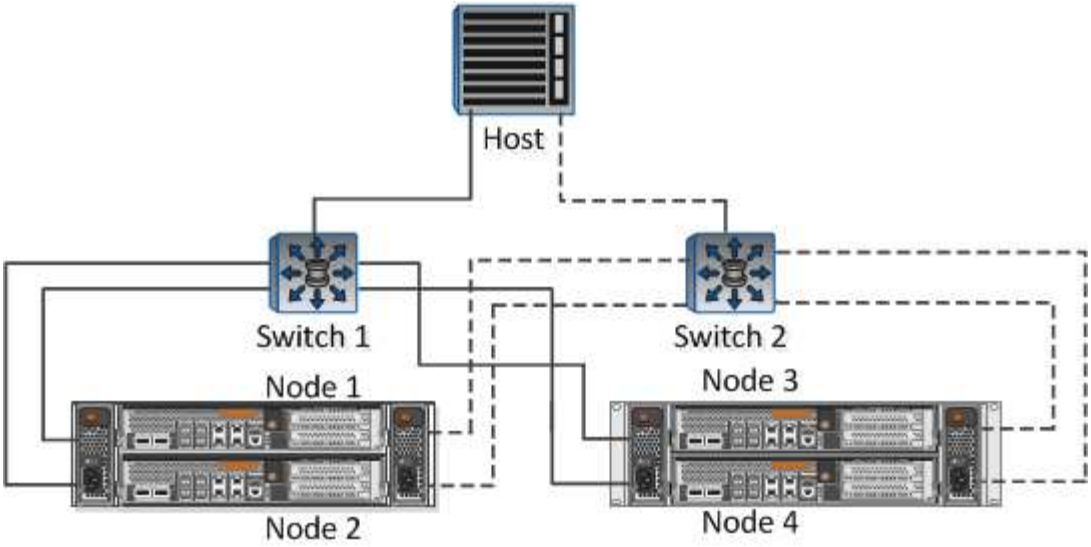
Evitare le operazioni vVol in diverse release di ONTAP.

Le funzionalità di storage supportate, come QoS, personalità e molto altro, sono cambiate in varie versioni del provider VASA e alcune dipendono dalla release di ONTAP. L'utilizzo di release diverse in un cluster ONTAP o lo spostamento di vVol tra cluster con release diverse può causare comportamenti imprevisti o allarmi di compliance.

Zona del fabric Fibre Channel prima di utilizzare FCP per vVol.

Il provider ONTAP Tools VASA si occupa della gestione degli igroup FCP e iSCSI, nonché dei sottosistemi NVMe in ONTAP in base agli iniziatori rilevati degli host ESXi gestiti. Tuttavia, non si integra con gli switch Fibre Channel per gestire lo zoning. Lo zoning deve essere eseguito in base alle Best practice prima di eseguire qualsiasi provisioning. Di seguito è riportato un esempio di zoning a initiator singolo per quattro sistemi ONTAP:

Zoning a initiator singolo:



Fare riferimento ai seguenti documenti per ulteriori Best practice:

"*TR-4080 Best practice per la MODERNA SAN ONTAP 9*"

"*TR-4684 implementazione e configurazione delle moderne SAN con NVMe-of*"

Pianificate i vostri volumi FlexVol di backup in base alle vostre esigenze.

Per i sistemi diversi da ASA R2, può essere conveniente aggiungere diversi volumi di backup al datastore vVol per distribuire il carico di lavoro nel cluster ONTAP, supportare diverse opzioni di policy o aumentare il numero di LUN o file consentiti. Tuttavia, se è richiesta la massima efficienza dello storage, posizionare tutti i volumi di backup su un singolo aggregato. In alternativa, se sono richieste le massime prestazioni di cloning, prendere in considerazione l'utilizzo di un singolo volume FlexVol e la conservazione dei modelli o della libreria di contenuti nello stesso volume. Il provider VASA trasferisce molte operazioni di storage vVol a ONTAP, tra cui migrazione, cloning e snapshot. Quando questa operazione viene eseguita all'interno di un singolo volume FlexVol, vengono utilizzati cloni di file efficienti in termini di spazio e sono quasi immediatamente disponibili. Quando questo viene eseguito su volumi FlexVol, le copie sono rapidamente disponibili e utilizzano la deduplica e la compressione inline, ma la massima efficienza dello storage potrebbe non essere ripristinata fino a quando i processi in background non vengono eseguiti su volumi che utilizzano la deduplica e la compressione in background. A seconda dell'origine e della destinazione, un certo livello di efficienza potrebbe risultare degradato.

Con i sistemi ASA R2, questa complessità viene rimossa dal momento che il concetto di un volume o aggregato viene astratto dall'utente. Il posizionamento dinamico viene gestito automaticamente e gli endpoint del protocollo vengono creati in base alle necessità. È possibile creare automaticamente al volo endpoint di protocollo aggiuntivi qualora sia necessaria una maggiore scalabilità.

Prendere in considerazione l'utilizzo di IOPS massimi per controllare macchine virtuali sconosciute o di test.

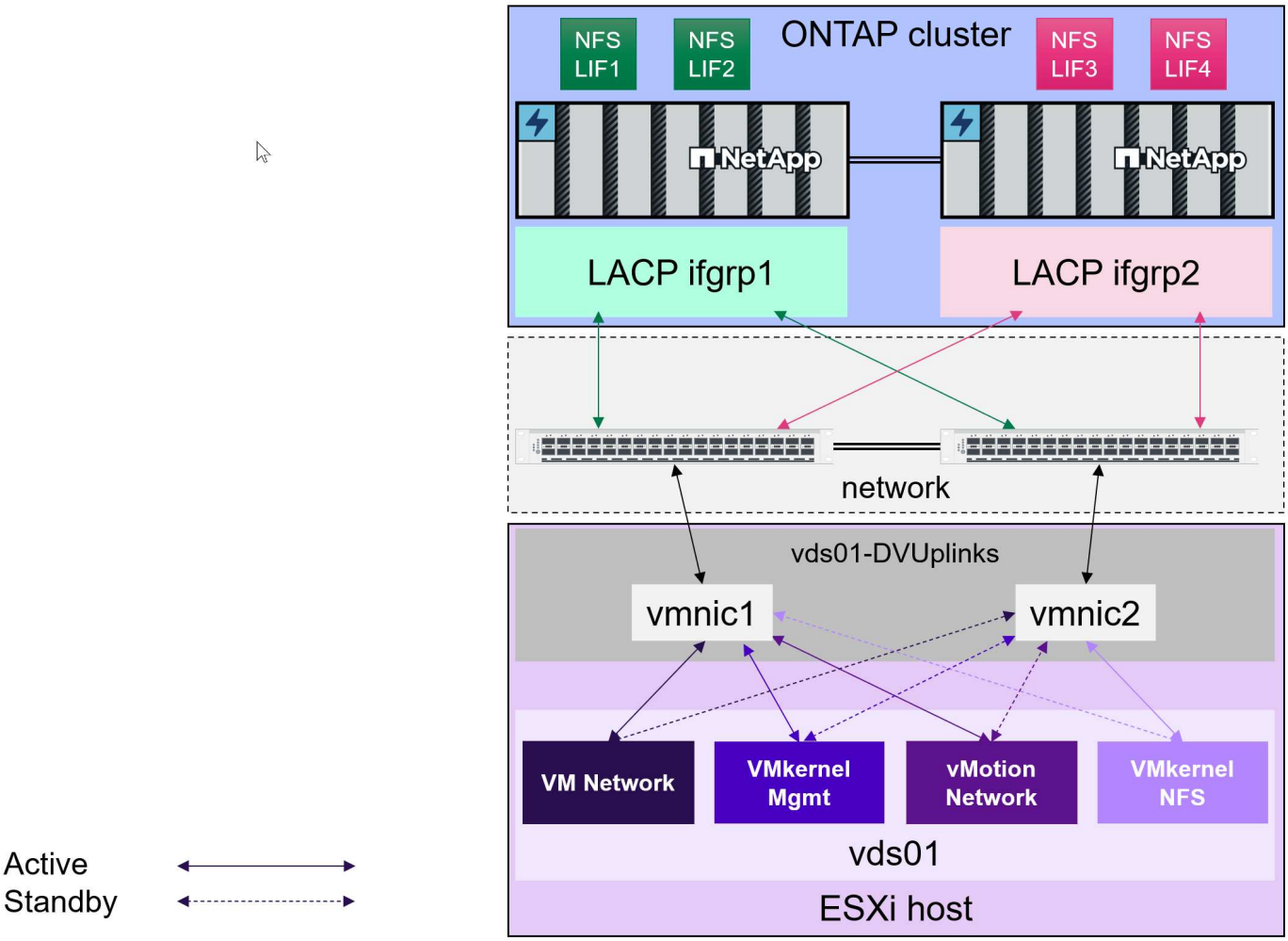
Per la prima volta disponibile nel provider VASA 7.1, è possibile utilizzare il massimo IOPS per limitare gli IOPS a un vVol specifico per un carico di lavoro sconosciuto, in modo da evitare impatti su altri carichi di lavoro più critici. Per ulteriori informazioni sulla gestione delle performance, vedere la Tabella 4.

Assicurarsi di disporre di LIF di dati sufficienti. Fare riferimento alla "Implementazione dello storage vVol".

Seguire tutte le Best practice del protocollo.

Fare riferimento alle altre guide alle Best practice di NetApp e VMware specifiche per il protocollo selezionato. In generale, non vi sono modifiche diverse da quelle già menzionate.

Esempio di configurazione di rete utilizzando vVol su NFS v3



Implementazione di vVol su sistemi AFF, ASA, ASA R2 e FAS

Seguire queste Best practice per la creazione dello storage vVol per le macchine virtuali.

Il provisioning dei datastore vVol implica diversi passaggi. I sistemi ASA R2 di NetApp sono progettati per carichi di lavoro VMware e offrono un'esperienza utente diversa dai sistemi ONTAP tradizionali. Quando si utilizzano i sistemi ASA R2, i tool ONTAP versione 10,3 o successive richiedono pochi passaggi per configurare e includere le estensioni dell'interfaccia utente e il supporto delle API REST ottimizzato per la nuova architettura storage.

Preparazione alla creazione di archivi dati vVol con gli strumenti ONTAP

È possibile saltare le prime due fasi del processo di distribuzione se si utilizzano già strumenti ONTAP per gestire, automatizzare e creare report sui sistemi di storage VMFS o tradizionali basati su NFS. Per la distribuzione e la configurazione degli strumenti ONTAP, è anche possibile fare riferimento a questa sezione completa"elenco di controllo".

1. Creare la Storage Virtual Machine (SVM) e la relativa configurazione del protocollo. Questa operazione potrebbe non essere necessaria per i sistemi ASA R2, in quanto in genere disporranno già di una singola

SVM per i servizi dati. È possibile selezionare NVMe/FC (solo tool ONTAP 9,13), NFSv3, NFSv4,1, iSCSI, FCP o un mix di queste opzioni. NVMe/TCP e NVMe/FC possono essere utilizzati anche per datastore VMFS tradizionali con tool ONTAP 10,3 e versioni successive. È possibile utilizzare le procedure guidate di ONTAP System Manager o la riga di comando della shell del cluster.

- "Assegnazione dei Tier locali (aggregati) alle SVM" Per tutti i sistemi non ASA R2.
- Almeno un LIF per nodo per ogni connessione switch/fabric. Come Best practice, creare due o più per nodo per i protocolli basati su FCP, iSCSI o NVMe. Un'unica LIF per nodo è sufficiente per i vVol basati su NFS, ma questa LIF deve essere protetta da un ifgroup LACP. Per ulteriori informazioni, fare riferimento alla "Panoramica sulla configurazione delle LIF" e "Combina le porte fisiche per creare gruppi di interfacce" alla.
- Almeno una LIF di gestione per SVM se intendi utilizzare le credenziali scoped della SVM per i vCenter del tenant.
- Se si prevede di utilizzare SnapMirror, assicurarsi che l'origine e la destinazione "Peering di cluster ONTAP e SVM"siano corrette.
- Per i sistemi non ASA R2, è possibile creare volumi in questo momento, ma è consigliabile consentire la creazione della procedura guidata *Provision DataStore* negli strumenti ONTAP. L'unica eccezione a questa regola è se si prevede di utilizzare la replica vVol con VMware Site Recovery Manager e gli strumenti ONTAP 9,13. Questa operazione è più semplice da configurare con volumi FlexVol preesistenti con relazioni SnapMirror esistenti. Prestare attenzione a non abilitare la qualità del servizio su alcun volume da utilizzare per i vVol, in quanto questa operazione deve essere gestita dai tool SPBM e ONTAP.

2. "Implementa i tool ONTAP per VMware vSphere" Utilizzando l'OVA scaricato dal sito di assistenza NetApp.
   - Gli strumenti ONTAP 10,0 e versioni successive supportano più server vCenter per appliance. Non è più necessario implementare una singola appliance di strumenti ONTAP per vCenter.
     - Se si prevede di collegare più centri virtuali a una singola istanza di strumenti ONTAP, è necessario creare e installare certificati CA firmati. Fare riferimento alla "Gestire i certificati" per i passi.
   - A partire dal 10,3, i tool ONTAP ora vengono implementati come appliance di piccole dimensioni a nodo singolo adatta alla maggior parte dei workload non vVol.

- La Best practice consigliata è "Tool ONTAP a scalabilità orizzontale" 10,3 GbE e versioni successive alla configurazione ha (High Availability, alta disponibilità) a 3 nodi per tutti i carichi di lavoro di produzione. Ai fini dei laboratori o del test, è possibile utilizzare un'implementazione a nodo singolo.
- La Best practice consigliata per l'uso dei vVol di produzione è eliminare ogni singolo punto di errore. Creare regole di anti-affinità per impedire l'esecuzione delle VM degli strumenti ONTAP sullo stesso host. Dopo l'implementazione iniziale, si consiglia anche di utilizzare storage vMotion per posizionare le macchine virtuali del tool ONTAP in datastore differenti. Ulteriori informazioni su "Utilizzo delle regole di affinità senza vSphere DRS" o "Creare una regola di affinità VM-VM". È inoltre necessario pianificare backup frequenti e/o "utilizzare l'utilità di backup della configurazione integrata".

1. Configurare gli strumenti ONTAP 10,3 per il proprio ambiente.
   - "Aggiungere istanze di vCenter Server" Nell'interfaccia utente di ONTAP Tools Manager.
   - Gli strumenti ONTAP 10,3 supportano la multi-tenancy sicura. Se non hai bisogno della multi-tenancy sicura, puoi semplicemente "Aggiungi i tuoi cluster ONTAP" accedere al menu degli strumenti di ONTAP in vCenter, fare clic su *backend di archiviazione* e fare clic sul pulsante *add*.
   - In un ambiente multitenant sicuro in cui desideri delegare alcune Storage Virtual Machine (SVM) a vCenter specifici, devi eseguire le seguenti operazioni.

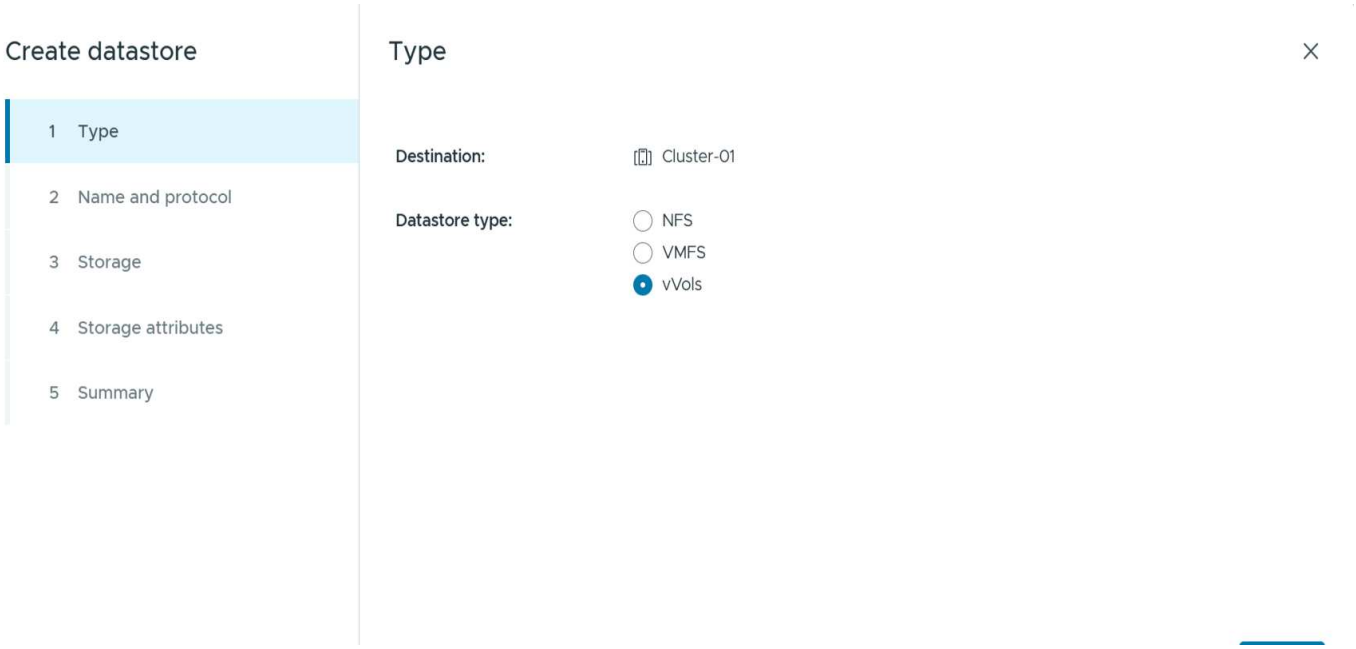
- Accedere all'interfaccia utente di ONTAP Tools Manager
- "Integrare il cluster di storage"
- "Associazione di un backend dello storage a un'istanza di vCenter Server"
- Fornisci le credenziali SVM specifiche all'amministratore vCenter, che aggiungerà la SVM come backend storage nel menu backend storage degli strumenti ONTAP in vCenter.

- È una Best practice creare ruoli RBAC per gli account storage.
- Gli strumenti ONTAP includono un file JSON contenente le autorizzazioni di ruolo necessarie per gli account di storage degli strumenti ONTAP. Puoi caricare il file JSON in ONTAP System Manager per semplificare la creazione di ruoli e utenti RBAC.
- Per ulteriori informazioni sui ruoli RBAC di ONTAP, visitare il sito Web all'indirizzo "Configurare i ruoli e i privilegi degli utenti ONTAP".

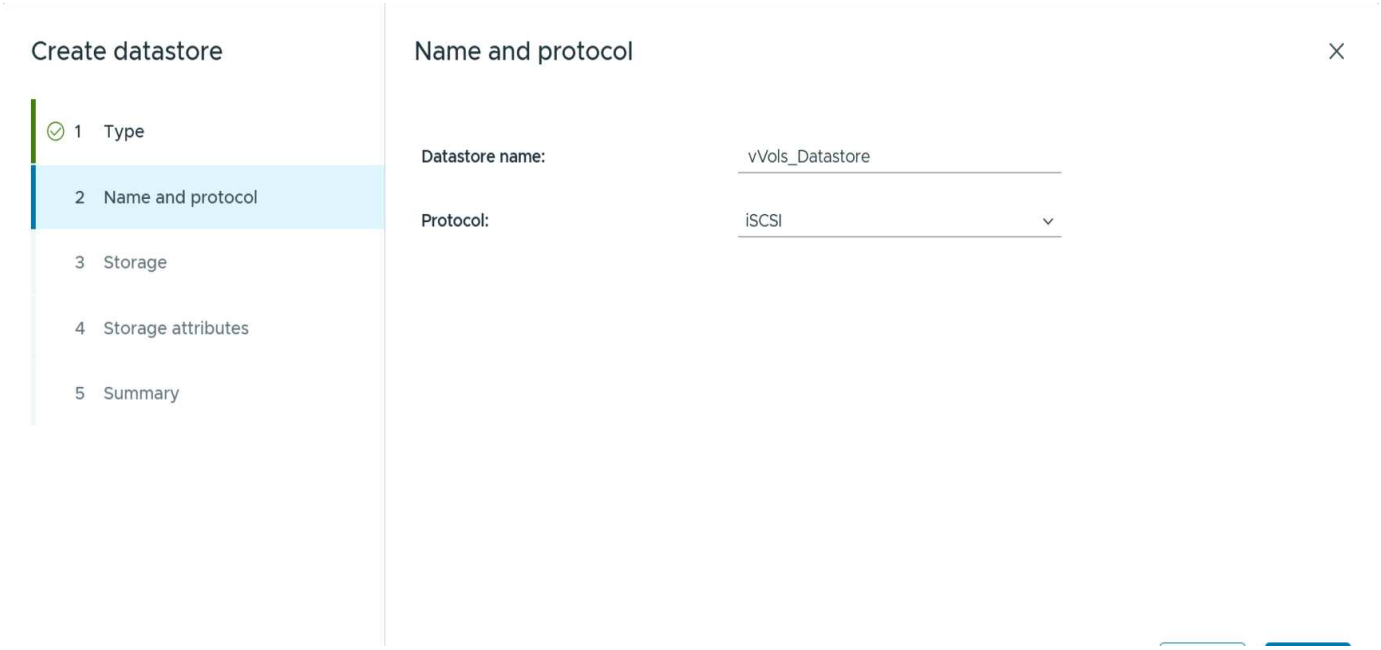
Il motivo per cui l'intero cluster deve essere integrato nell'interfaccia utente del gestore degli strumenti di ONTAP è che molte delle API utilizzate per i vVol sono disponibili solo a livello di cluster.

Creazione di archivi dati vVol con gli strumenti ONTAP

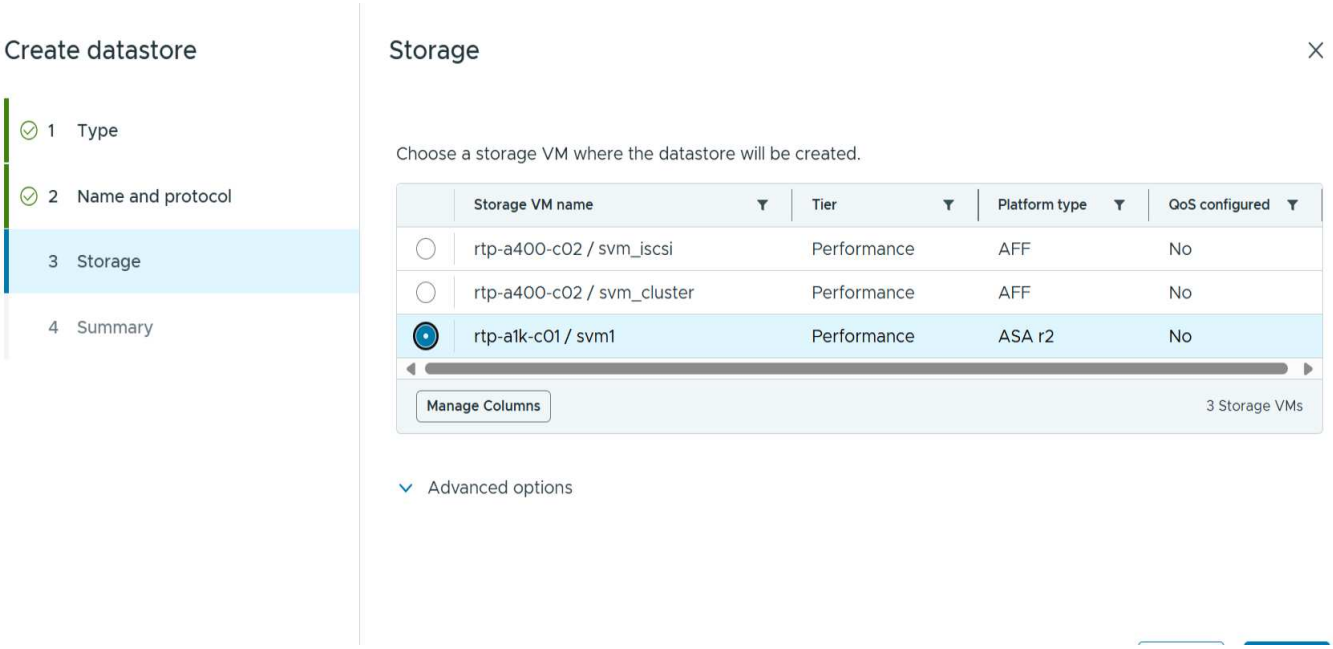
Fare clic con il pulsante destro del mouse sull'host, sul cluster o sul data center su cui si desidera creare il datastore vVols, quindi selezionare *ONTAP Tools* > *Provision Datastore*.



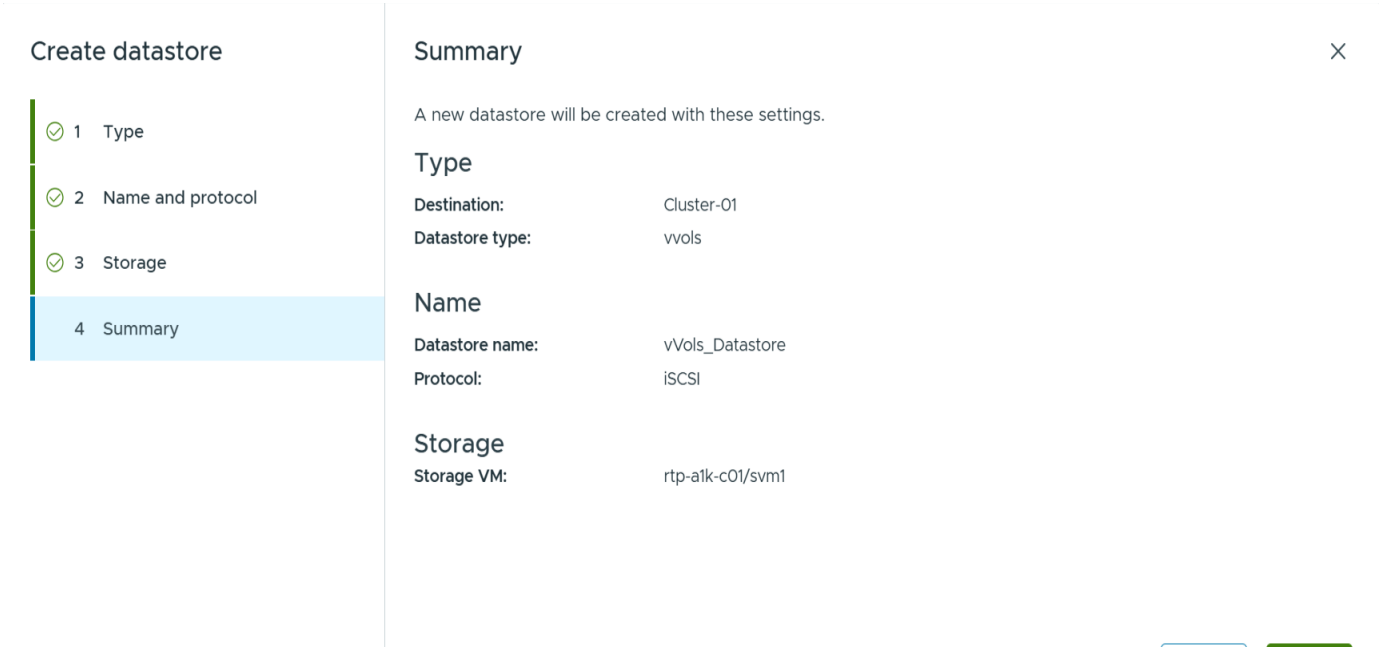
- Scegliere vVol e fornire un nome significativo e selezionare il protocollo desiderato. È anche possibile fornire una descrizione del datastore.
  - Strumenti ONTAP 10,3 con ASA R2.



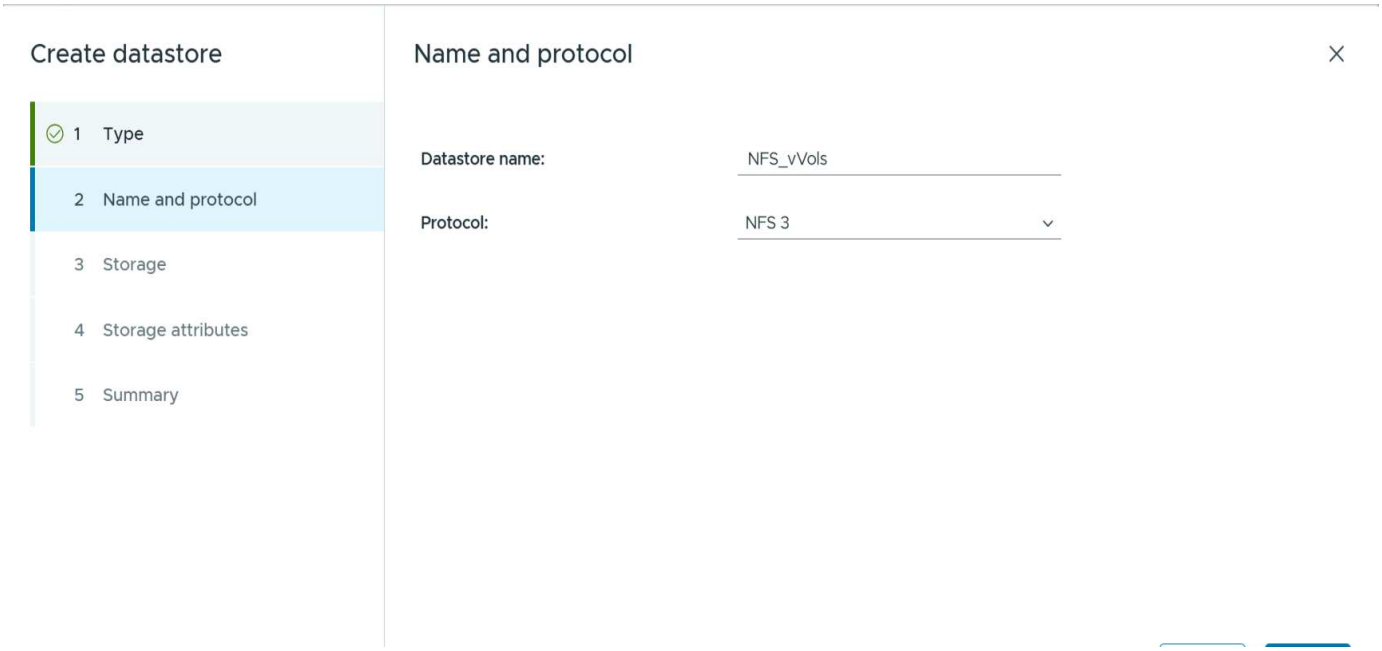
- Seleziona la SVM del sistema ASA R2 e fai clic su *next*.



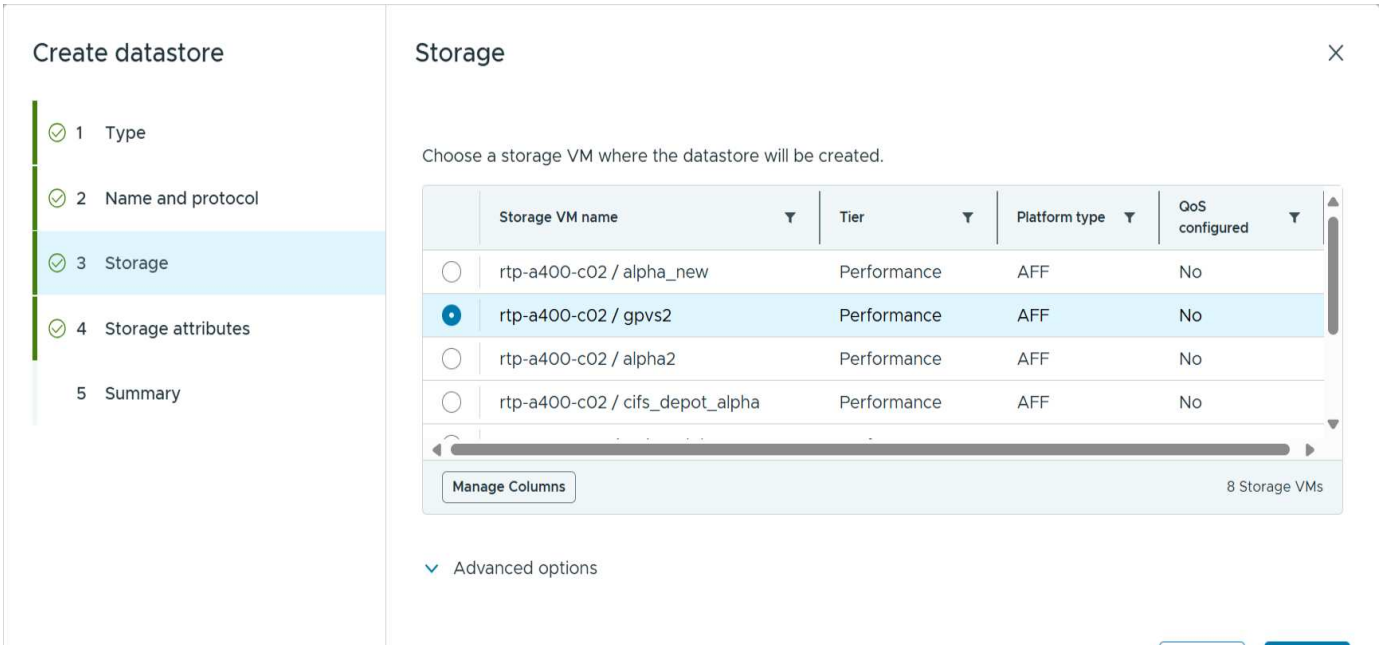
- Fare clic su *fine*



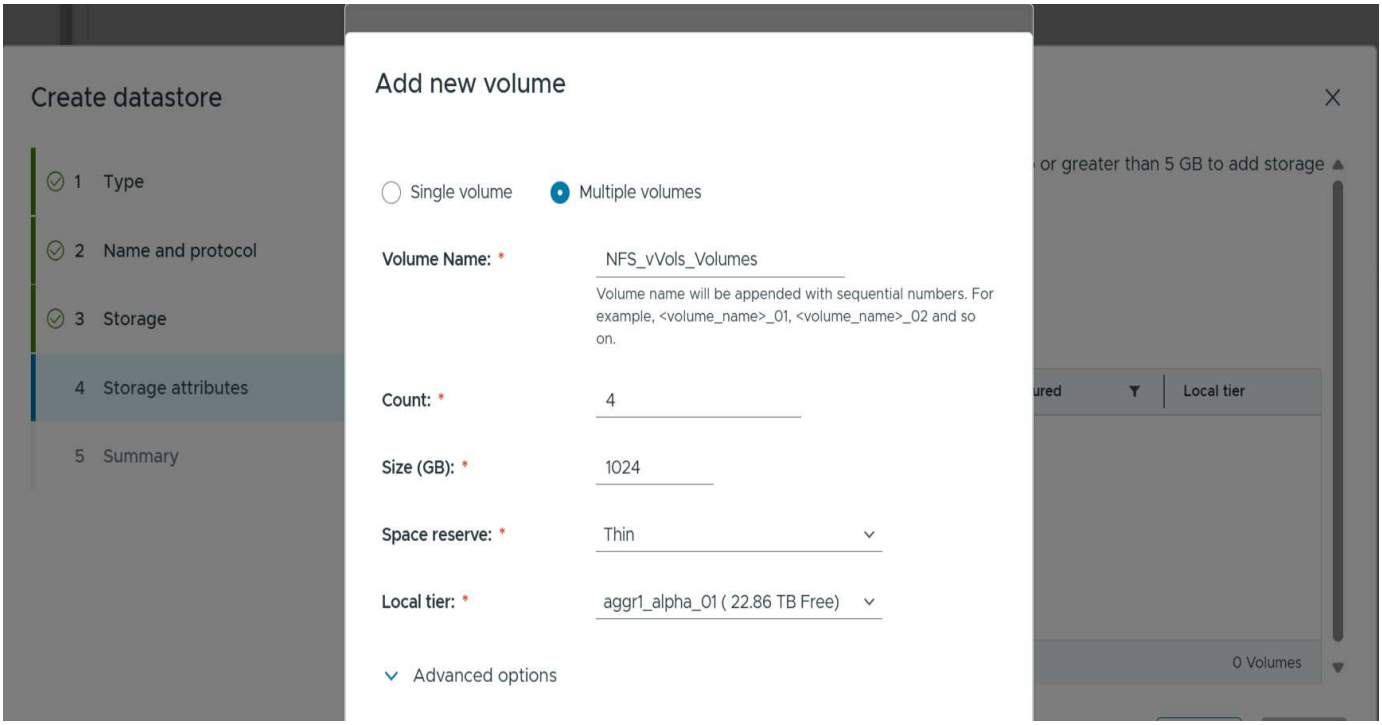
- È facile!
  - ONTAP Tools 10,3 con ONTAP FAS, AFF e ASA Prior ASA R2.
- Selezionare il protocollo

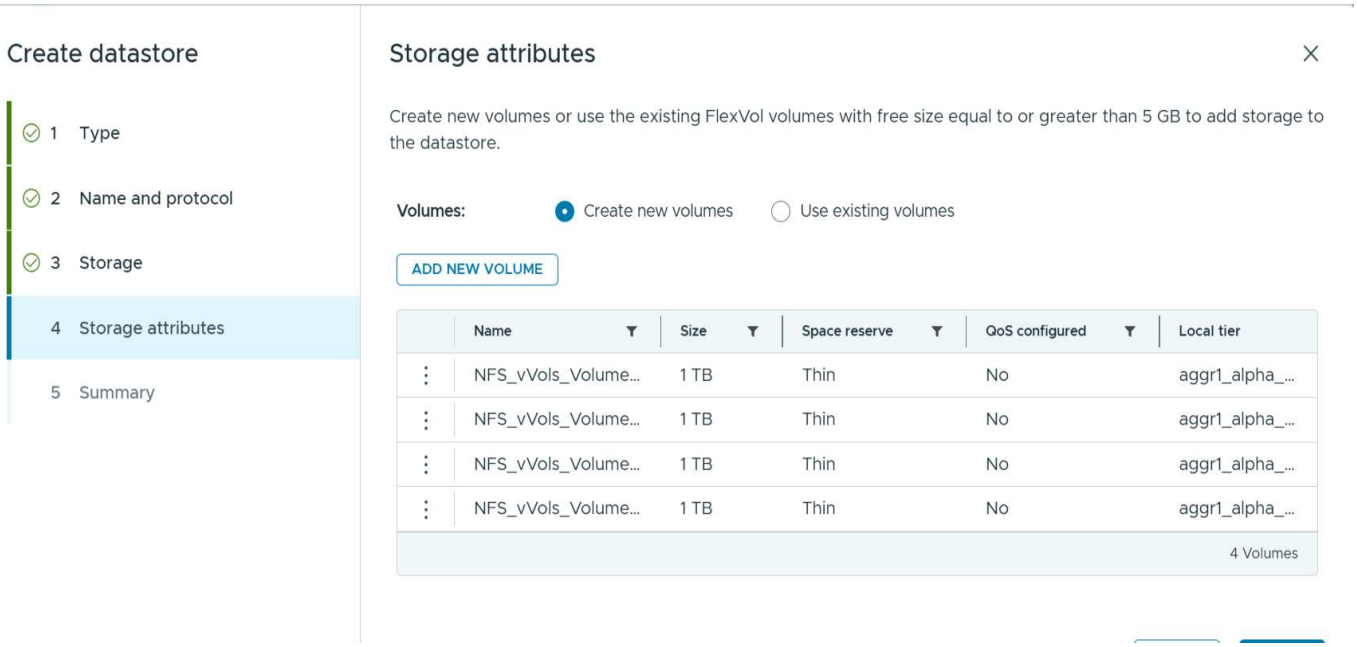


- Seleziona la SVM e fai clic su *next*.

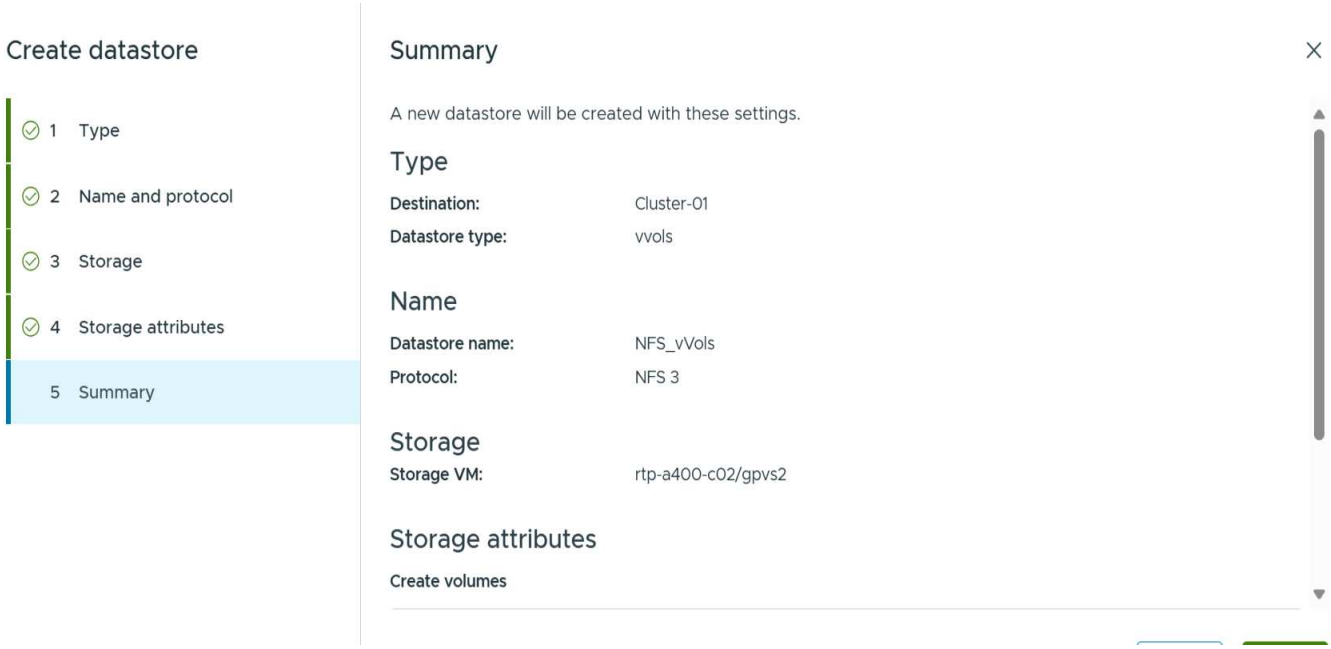


- Fare clic su *add new volumes* o *use existing volume* e specificare gli attributi. Nota: Negli strumenti di ONTAP 10,3 puoi richiedere la creazione contemporanea di più volumi. È possibile anche aggiungere manualmente più volumi per bilanciarli nel cluster ONTAP. Fare clic su *avanti*

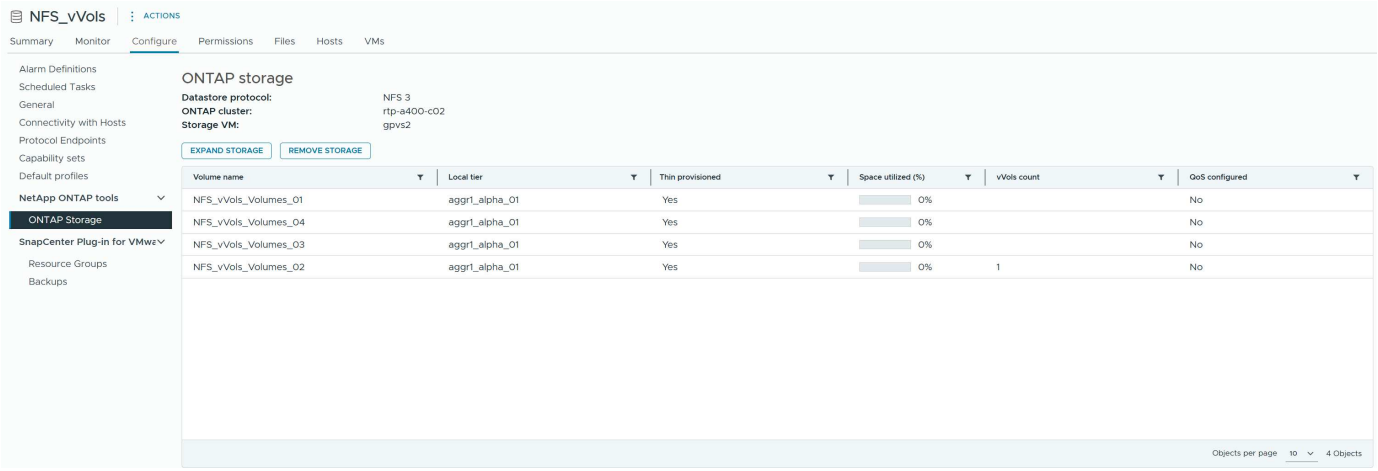




- Fare clic su *fine*



- I volumi assegnati possono essere visualizzati nel menu ONTAP tools della scheda Configure per l'archivio dati.



- Ora puoi creare policy storage delle macchine virtuali dal menu *Policies and Profiles* nell'interfaccia utente di vCenter.

Migrazione di macchine virtuali da datastore tradizionali a vVol

La migrazione delle macchine virtuali dai datastore tradizionali a un datastore vVol è semplice quanto lo spostamento delle macchine virtuali tra datastore tradizionali. È sufficiente selezionare le macchine virtuali, quindi Migrate (Migra) dall'elenco delle azioni e selezionare un tipo di migrazione di *change storage only*. Quando richiesto, seleziona una policy storage della macchina virtuale che corrisponda al datastore vVol. È possibile eseguire l'offload delle operazioni di copia della migrazione con vSphere 6,0 e versioni successive per le migrazioni da VMFS SAN a vVol, ma non da VMDK NAS a vVol.

Gestione delle VM mediante policy

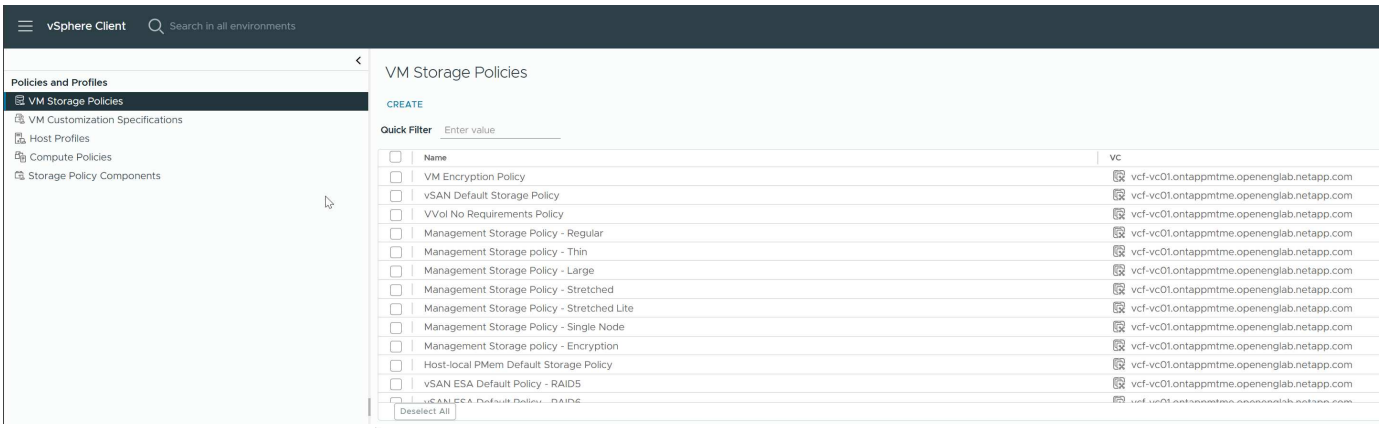
Per automatizzare il provisioning dello storage con una gestione basata su criteri, è necessario creare policy di storage delle macchine virtuali associate alle funzionalità di storage desiderate.

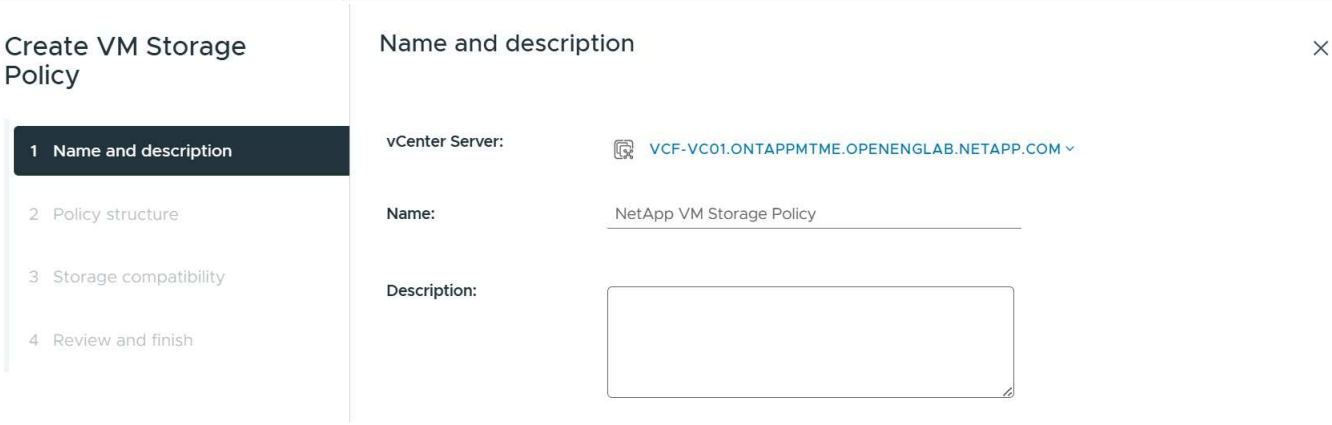
> Gli strumenti ONTAP 10,0 e versioni successive non utilizzano più i profili di funzionalità dello storage come le versioni precedenti. Le funzionalità di storage vengono invece definite direttamente nel criterio di storage delle macchine virtuali.

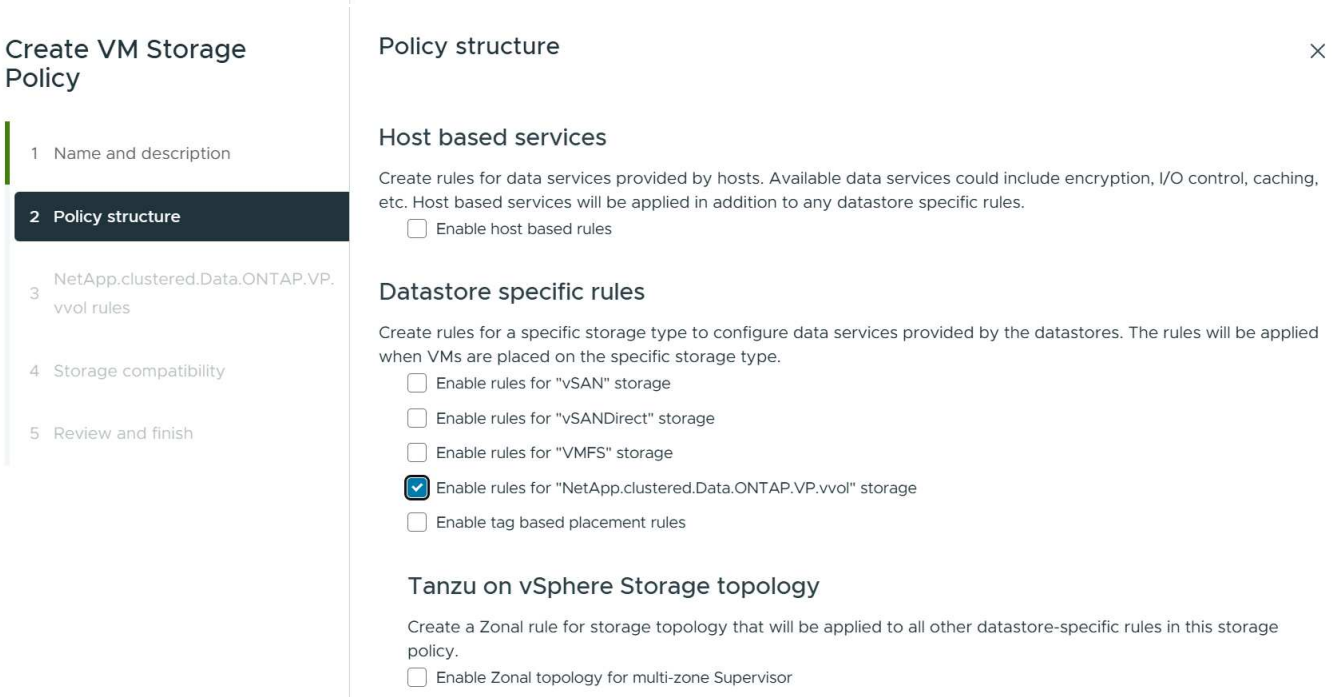
Creazione di policy di storage delle macchine virtuali

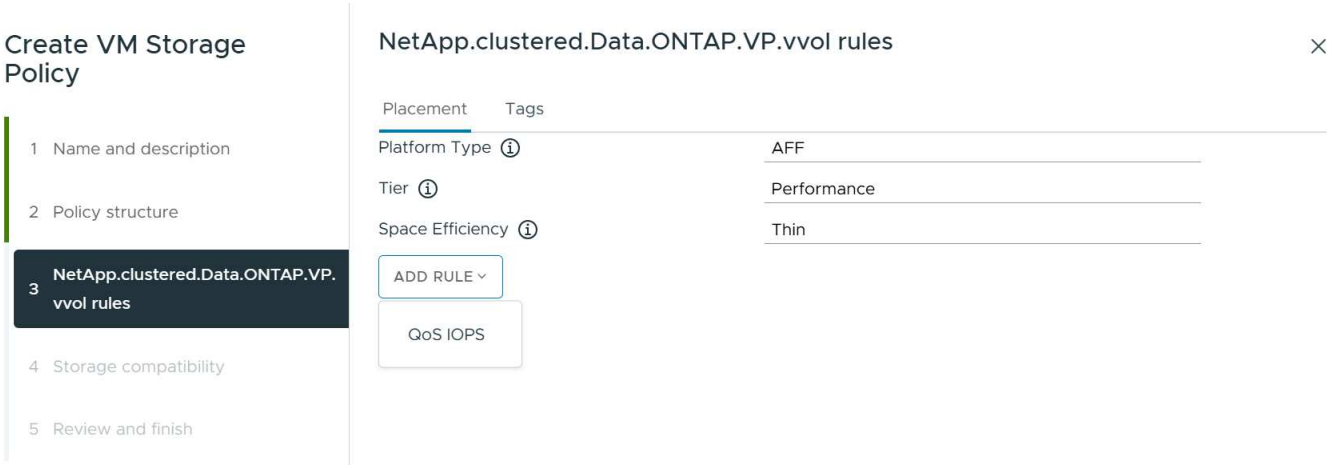
Le policy di storage delle macchine virtuali vengono utilizzate in vSphere per gestire funzionalità opzionali come Storage i/o Control o vSphere Encryption. Vengono inoltre utilizzati con vVol per applicare funzionalità di storage specifiche alla macchina virtuale. Utilizza il tipo di storage "NetApp.Clustered.Data.ONTAP.VP.vvol". Consulta esempio di configurazione di rete con vVol su NFS v3 per un esempio con il provider VASA degli strumenti ONTAP. Le regole per lo storage "NetApp.Clustered.Data.ONTAP.VP.VASA10" devono essere utilizzate con datastore non basati su vVol.

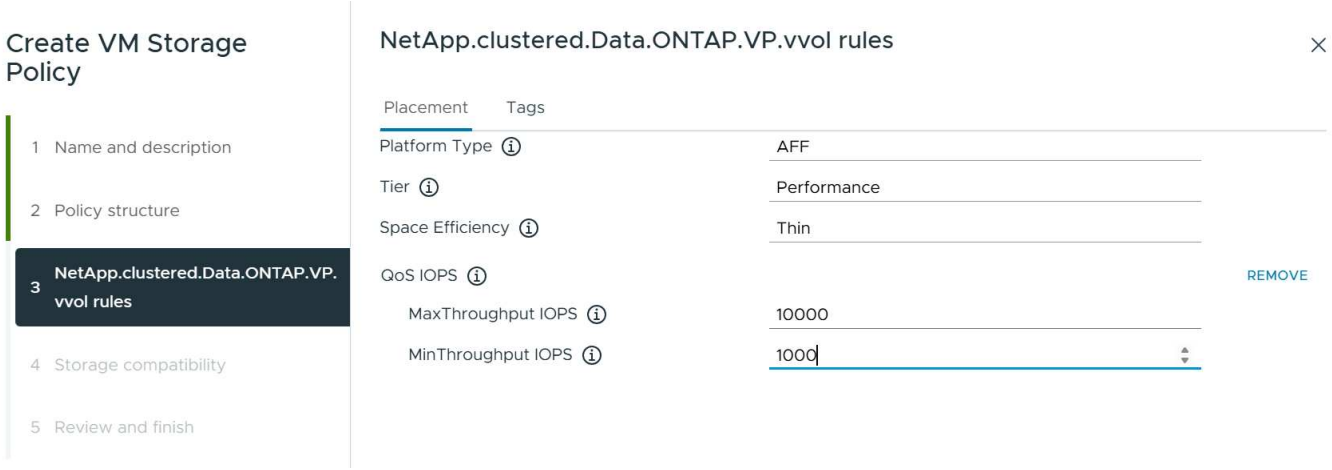
Una volta creato, il criterio storage può essere utilizzato per il provisioning di nuove macchine virtuali.

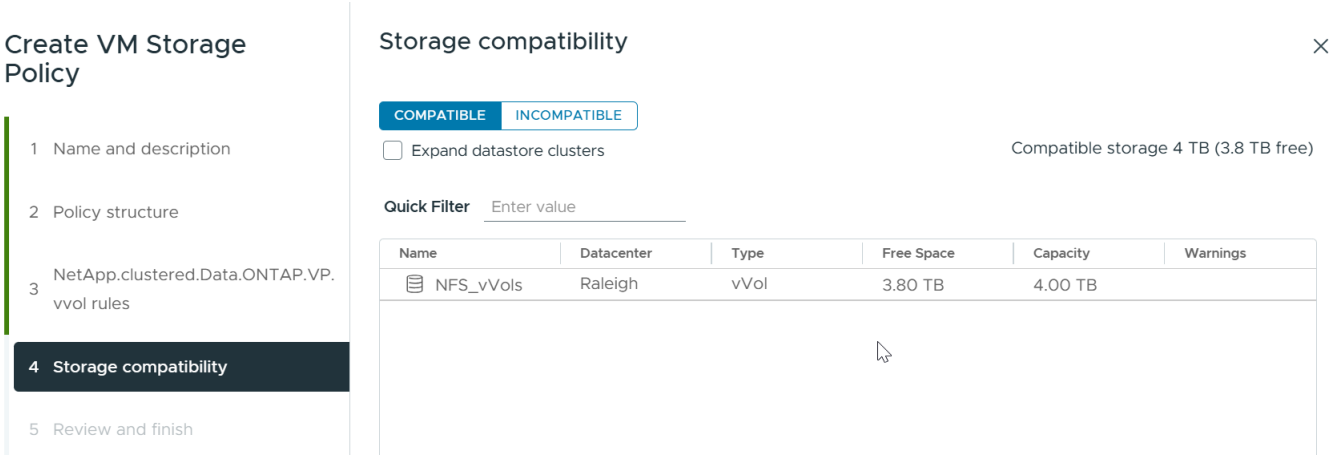

vSphere Client
Search in all environments
Policies and Profiles
VM Storage Policies
VM Customization Specifications
Host Profiles
Compute Policies
Storage Policy Components
VM Storage Policies
CREATE
Quick Filter
Enter value
Name
VC
VM Encryption Policy
vSAN Default Storage Policy
VVol No Requirements Policy
Management Storage Policy - Regular
Management Storage policy - Thin
Management Storage Policy - Large
Management Storage Policy - Stretched
Management Storage Policy - Stretched Lite
Management Storage Policy - Single Node
Management Storage policy - Encryption
Host-local PMem Default Storage Policy
vSAN ESA Default Policy - RAID5
vcf-vc01.ontappmtme.openenglab.netapp.com
Deselect All

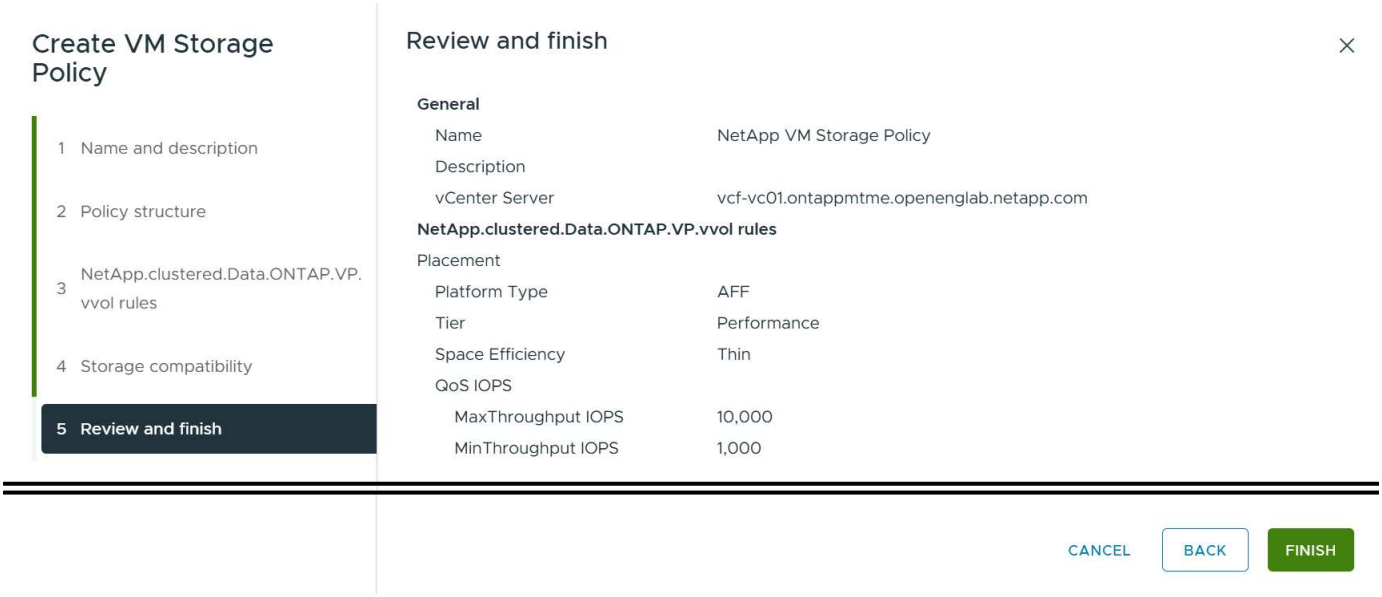

Create VM Storage Policy
1 Name and description
2 Policy structure
3 Storage compatibility
4 Review and finish
Name and description
vCenter Server:
VCF-VC01.ONTAPPMTME.OPENENGLAB.NETAPP.COM
Name:
NetApp VM Storage Policy
Description:


Create VM Storage Policy
1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP.vvol rules
4 Storage compatibility
5 Review and finish
Policy structure
Host based services
Create rules for data services provided by hosts. Available data services could include encryption, I/O control, caching, etc. Host based services will be applied in addition to any datastore specific rules.
Enable host based rules
Datastore specific rules
Create rules for a specific storage type to configure data services provided by the datastores. The rules will be applied when VMs are placed on the specific storage type.
Enable rules for "vSAN" storage
Enable rules for "vSANDirect" storage
Enable rules for "VMFS" storage
Enable rules for "NetApp.clustered.Data.ONTAP.VP.vvol" storage
Enable tag based placement rules
Tanzu on vSphere Storage topology
Create a Zonal rule for storage topology that will be applied to all other datastore-specific rules in this storage policy.
Enable Zonal topology for multi-zone Supervisor


Create VM Storage Policy
1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP. vvol rules
4 Storage compatibility
5 Review and finish
NetApp.clustered.Data.ONTAP.VP.vvol rules
Placement
Tags
Platform Type
AFF
Tier
Performance
Space Efficiency
Thin
ADD RULE
QoS IOPS


Create VM Storage Policy
1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP. vvol rules
4 Storage compatibility
5 Review and finish
NetApp.clustered.Data.ONTAP.VP.vvol rules
Placement
Tags
Platform Type
AFF
Tier
Performance
Space Efficiency
Thin
QoS IOPS
REMOVE
MaxThroughput IOPS
10000
MinThroughput IOPS
1000


Create VM Storage Policy
1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP. vvol rules
4 Storage compatibility
5 Review and finish
Storage compatibility
COMPATIBLE
INCOMPATIBLE
Expand datastore clusters
Compatible storage 4 TB (3.8 TB free)
Quick Filter
Enter value
Name
Datacenter
Type
Free Space
Capacity
Warnings
NFS_vVols
Raleigh
vVol
3.80 TB
4.00 TB



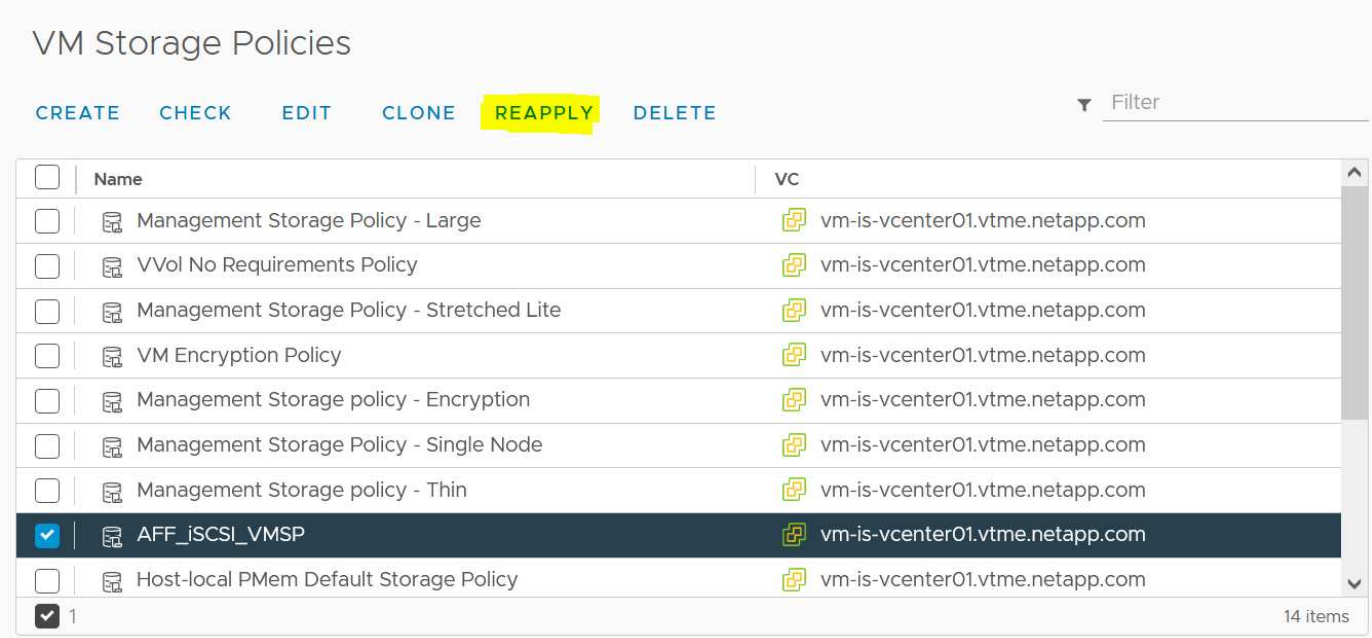
Gestione delle performance con tool ONTAP

I tool ONTAP utilizzano il proprio algoritmo di posizionamento bilanciato per posizionare un nuovo vVol nel BEST FlexVol volume, con sistemi ASA unificati o classici, o Storage Availability zone (SAZ) con sistemi ASA R2, all'interno di un datastore vVol. Il posizionamento si basa sulla corrispondenza tra lo storage di backup e il criterio di archiviazione della VM. In questo modo si garantisce che il datastore e lo storage di backup soddisfino i requisiti di performance specificati.

La modifica delle funzionalità delle performance, ad esempio IOPS min e max, richiede un'attenzione particolare alla configurazione specifica.

- **IOPS minimi e massimi** possono essere specificati in un criterio VM.
  - La modifica degli IOPS della policy non modifica la QoS sui vVol fino a quando il criterio VM non viene riapplicato alle VM che lo utilizzano. In alternativa, è possibile creare una nuova policy con gli IOPS desiderati e applicarla alle macchine virtuali di destinazione. In genere, si consiglia di definire semplicemente policy di storage delle macchine virtuali separate per diversi Tier di servizio e di modificare semplicemente il criterio di storage delle macchine virtuali sulla macchina virtuale.
  - Le personalità ASA, ASA R2, AFF e FAS hanno diverse impostazioni di IOPS. Sia min che Max sono disponibili sui sistemi all-flash; tuttavia, i sistemi non AFF possono utilizzare solo le impostazioni IOPS massime.
- Gli strumenti ONTAP creano policy QoS individuali non condivise con le versioni attualmente supportate di ONTAP. Pertanto, ogni singolo VMDK riceverà la propria allocazione di IOPS.

Riapplicazione dei criteri di storage delle macchine virtuali



Protezione di vVol

Nelle seguenti sezioni vengono illustrate le procedure e le Best practice per l'utilizzo di vVol VMware con lo storage ONTAP.

ALTA disponibilità del provider VASA

NetApp VASA Provider viene eseguito come parte dell'appliance virtuale insieme al plug-in vCenter, al server REST API (precedentemente noto come Virtual Storage Console [VSC]) e allo Storage Replication Adapter. Se il provider VASA non è disponibile, le VM che utilizzano vVol continueranno a funzionare. Tuttavia, non è possibile creare nuovi datastore vVol e non è possibile creare o vinare vVol da vSphere. Ciò significa che le macchine virtuali che utilizzano vVol non possono essere attivate poiché vCenter non sarà in grado di richiedere la creazione dello swap vVol. Inoltre, le macchine virtuali in esecuzione non possono utilizzare vMotion per la migrazione a un altro host perché i vVol non possono essere associati al nuovo host.

VASA Provider 7.1 e versioni successive supportano nuove funzionalità per garantire la disponibilità dei servizi quando necessario. Include nuovi processi di controllo che monitorano il provider VASA e i servizi di database integrati. Se rileva un errore, aggiorna i file di registro e riavvia automaticamente i servizi.

L'amministratore di vSphere deve configurare un'ulteriore protezione utilizzando le stesse funzionalità di disponibilità utilizzate per proteggere le altre macchine virtuali mission-critical da guasti del software, dell'hardware host e della rete. Non è richiesta alcuna configurazione aggiuntiva sull'appliance virtuale per utilizzare queste funzionalità; è sufficiente configurarle utilizzando gli approcci standard vSphere. Sono stati testati e supportati da NetApp.

VSphere High Availability è facilmente configurabile per riavviare una macchina virtuale su un altro host nel cluster host in caso di guasto. VSphere Fault Tolerance offre una maggiore disponibilità creando una macchina virtuale secondaria che viene continuamente replicata e che può assumere il controllo in qualsiasi momento. Ulteriori informazioni su queste funzioni sono disponibili nella "Strumenti ONTAP per la documentazione di VMware vSphere (configurare l'alta disponibilità per i tool ONTAP)", Oltre alla documentazione VMware vSphere (cercare vSphere Availability sotto ESXi e vCenter Server).

Il provider VASA di ONTAP Tools esegue automaticamente il backup della configurazione vVol in tempo reale sui sistemi ONTAP gestiti in cui le informazioni vVol vengono memorizzate nei metadati dei volumi FlexVol. Nel

caso in cui l'appliance ONTAP Tools non fosse disponibile per qualsiasi motivo, è possibile implementarne una nuova e importarne la configurazione in modo semplice e rapido. Fare riferimento a questo articolo della Knowledge base per ulteriori informazioni sulle fasi di ripristino del provider VASA:

"Come eseguire un Disaster Recovery provider VASA - Guida alla risoluzione"

Replica di vVol

Molti clienti ONTAP replicano i propri datastore tradizionali su sistemi storage secondari utilizzando NetApp SnapMirror, quindi utilizzano il sistema secondario per ripristinare singole macchine virtuali o un intero sito in caso di disastro. Nella maggior parte dei casi, i clienti utilizzano uno strumento software per la gestione di questo tipo, ad esempio un prodotto software di backup come il plug-in NetApp SnapCenter per VMware vSphere o una soluzione di disaster recovery come Site Recovery Manager di VMware (insieme all'adattatore di replica dello storage negli strumenti ONTAP).

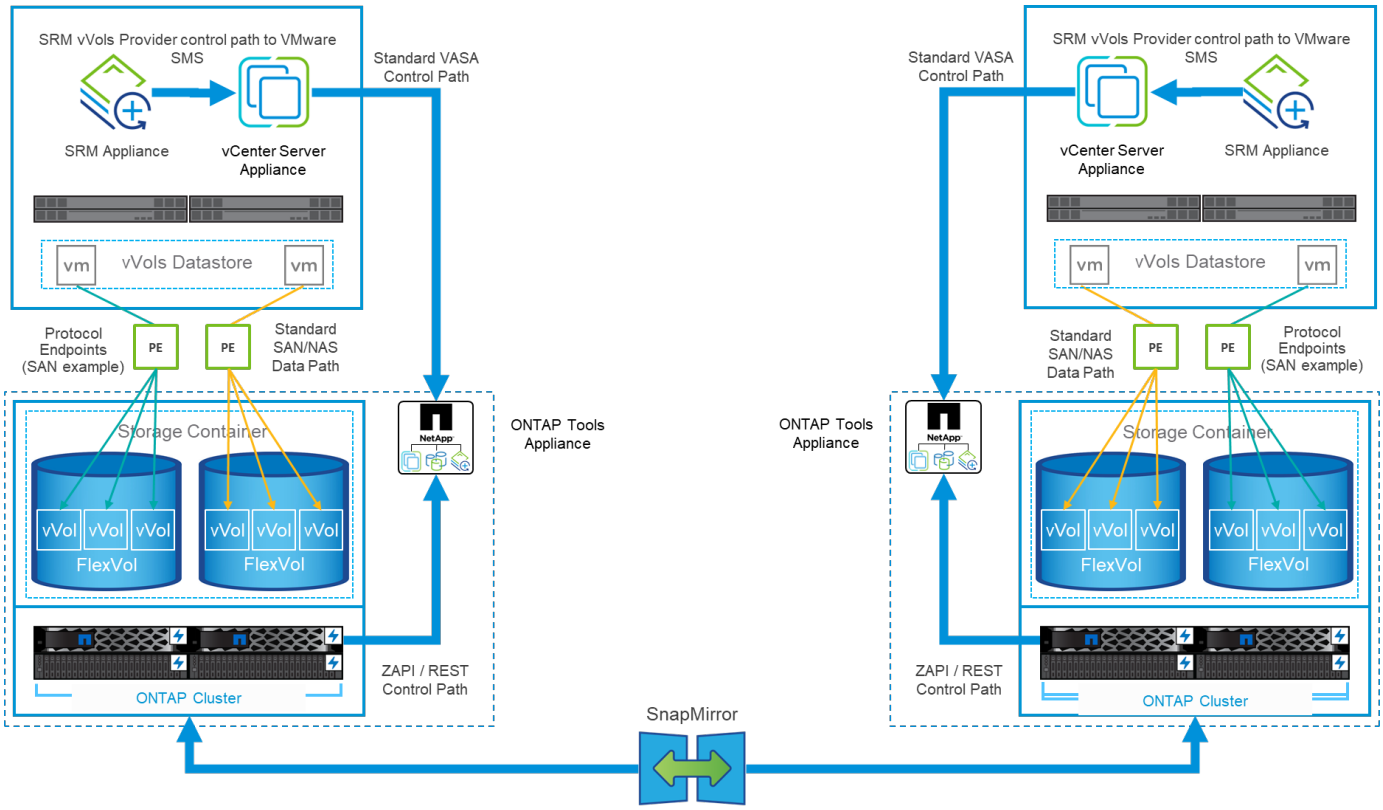
Questo requisito per uno strumento software è ancora più importante per gestire la replica di vVol. Sebbene alcuni aspetti possano essere gestiti da funzionalità native (ad esempio, le snapshot gestite da VMware di vVol vengono trasferite su ONTAP, che utilizza cloni di file o LUN rapidi ed efficienti), in generale l'orchestrazione è necessaria per gestire la replica e il ripristino. I metadati relativi ai vVol sono protetti da ONTAP e dal provider VASA, ma è necessaria un'ulteriore elaborazione per utilizzarli in un sito secondario.

I tool ONTAP 9.7.1, insieme alla release 8.3 di VMware Site Recovery Manager (SRM), hanno aggiunto il supporto per il disaster recovery e l'orchestrazione del flusso di lavoro di migrazione sfruttando la tecnologia SnapMirror di NetApp.

Nella release iniziale del supporto SRM con i tool ONTAP 9.7.1 era necessario creare in anticipo i volumi FlexVol e abilitare la protezione SnapMirror prima di utilizzarli come volumi di backup per un datastore vVol. A partire dagli strumenti ONTAP 9.10, questo processo non è più necessario. È ora possibile aggiungere la protezione SnapMirror ai volumi di backup esistenti e aggiornare le policy di storage delle macchine virtuali per sfruttare la gestione basata su policy con disaster recovery, orchestrazione e automazione della migrazione integrate con SRM.

Attualmente, VMware SRM è l'unica soluzione di disaster recovery e automazione della migrazione per vVol supportata da NetApp e i tool ONTAP verificheranno l'esistenza di un server SRM 8.3 o successivo registrato con vCenter prima di consentire la replica di vVol, Sebbene sia possibile sfruttare le API REST degli strumenti ONTAP per creare i propri servizi.

Replica di vVol con SRM



Supporto MetroCluster

Sebbene gli strumenti ONTAP non siano in grado di attivare uno switchover MetroCluster, supportano i sistemi NetApp MetroCluster per il backup dei volumi in una configurazione vMSC (vSphere Metro Storage Cluster) uniforme. La commutazione di un sistema MetroCluster viene gestita normalmente.

Anche se NetApp SnapMirror Business Continuity (SM-BC) può essere utilizzato come base per una configurazione vMSC, al momento non è supportato con vVol.

Consulta queste guide per ulteriori informazioni su NetApp MetroCluster:

"*Architettura e progettazione della soluzione IP TR-4689 MetroCluster*"

"*TR-4705 architettura e progettazione della soluzione NetApp MetroCluster*"

"*VMware KB 2031038 supporto VMware vSphere con NetApp MetroCluster*"

Panoramica del backup di vVol

Esistono diversi approcci per la protezione delle macchine virtuali, ad esempio l'utilizzo di agenti di backup in-guest, l'aggiunta di file di dati delle macchine virtuali a un proxy di backup o l'utilizzo di API definite come VMware VADP. I vVol possono essere protetti utilizzando gli stessi meccanismi e molti partner NetApp supportano i backup delle macchine virtuali, inclusi i vVol.

Come accennato in precedenza, le snapshot gestite da VMware vCenter vengono trasferite a cloni di file/LUN ONTAP efficienti in termini di spazio e veloci. Questi possono essere utilizzati per backup manuali e rapidi, ma sono limitati da vCenter a un massimo di 32 snapshot. È possibile utilizzare vCenter per creare snapshot e ripristinarli in base alle necessità.

A partire dal plug-in SnapCenter per VMware vSphere (SCV) 4.6, se utilizzato insieme ai tool ONTAP 9.10 e versioni successive, aggiunge il supporto per backup e ripristino coerenti in caso di crash delle macchine

virtuali basate su vVol, sfruttando le snapshot dei volumi ONTAP FlexVol con il supporto per SnapMirror e la replica SnapVault. Sono supportati fino a 1023 snapshot per volume. SCV può anche memorizzare più snapshot con una maggiore conservazione sui volumi secondari utilizzando SnapMirror con una policy di vault mirror.

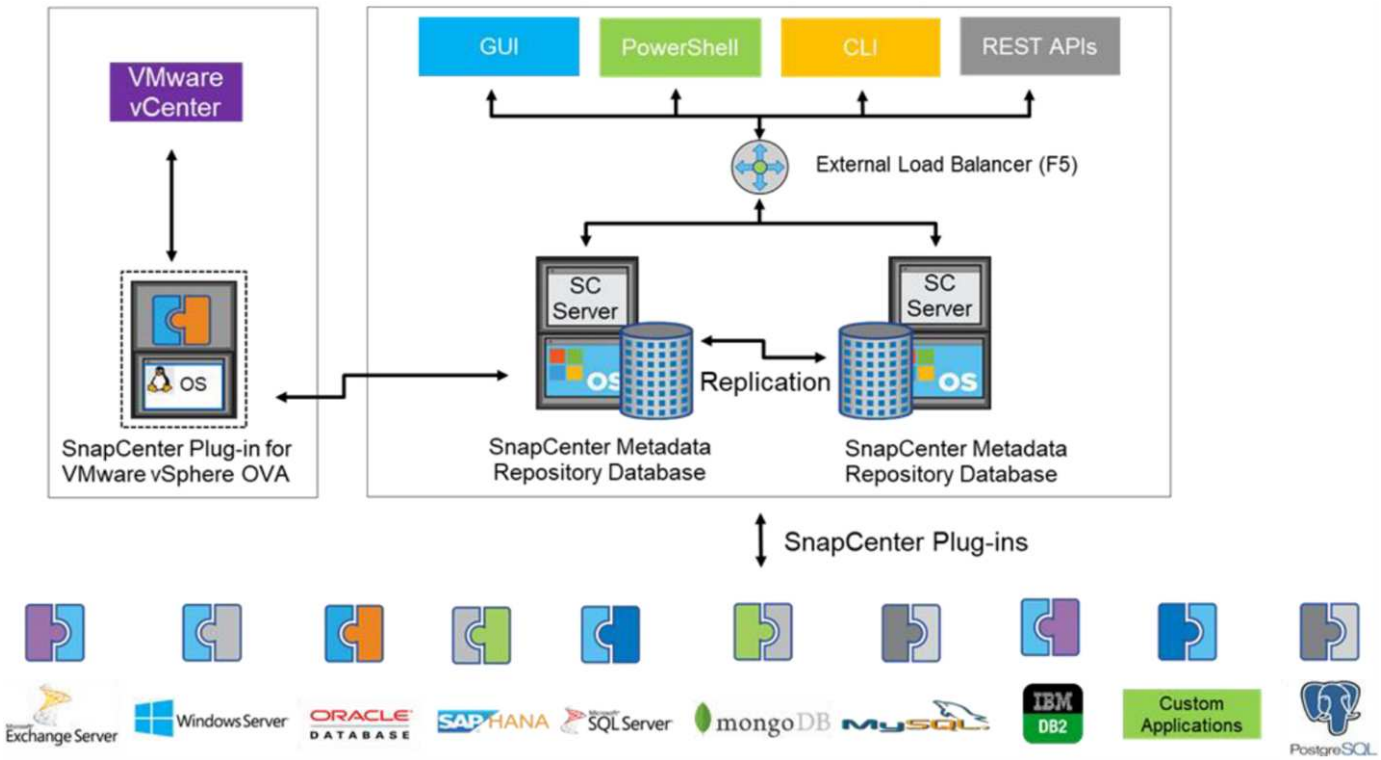
Il supporto di vSphere 8.0 è stato introdotto con SCV 4.7, che utilizzava un'architettura di plug-in locale isolata. Il supporto di vSphere 8.0U1 è stato aggiunto a SCV 4.8, che ha completato la transizione alla nuova architettura di plug-in remoto.

Backup vVol con plug-in SnapCenter per VMware vSphere

Con NetApp SnapCenter puoi ora creare gruppi di risorse per i vVol basati su tag e/o cartelle per sfruttare automaticamente le snapshot basate su FlexVol di ONTAP per macchine virtuali basate su vVol. Ciò consente di definire servizi di backup e ripristino che proteggeranno automaticamente le macchine virtuali man mano che vengono sottoposte a provisioning dinamico all'interno dell'ambiente.

Il plug-in SnapCenter per VMware vSphere viene implementato come appliance standalone registrata come estensione vCenter, gestita tramite l'interfaccia utente di vCenter o tramite API REST per l'automazione dei servizi di backup e recovery.

Architettura SnapCenter



Poiché gli altri plug-in di SnapCenter non supportano ancora i vVol al momento di questa scrittura, in questo documento ci concentreremo sul modello di distribuzione standalone.

Poiché SnapCenter utilizza snapshot ONTAP FlexVol, non è previsto alcun overhead su vSphere, né penalità in termini di performance, come si può vedere con le macchine virtuali tradizionali che utilizzano snapshot gestite da vCenter. Inoltre, poiché le funzionalità di SCV sono esposte attraverso le API REST, è semplice creare workflow automatizzati utilizzando tool come VMware aria Automation, Ansible, Terraform e virtualmente qualsiasi altro tool di automazione in grado di utilizzare le API REST standard.

Per informazioni sulle API REST di SnapCenter, vedere "Panoramica delle API REST"

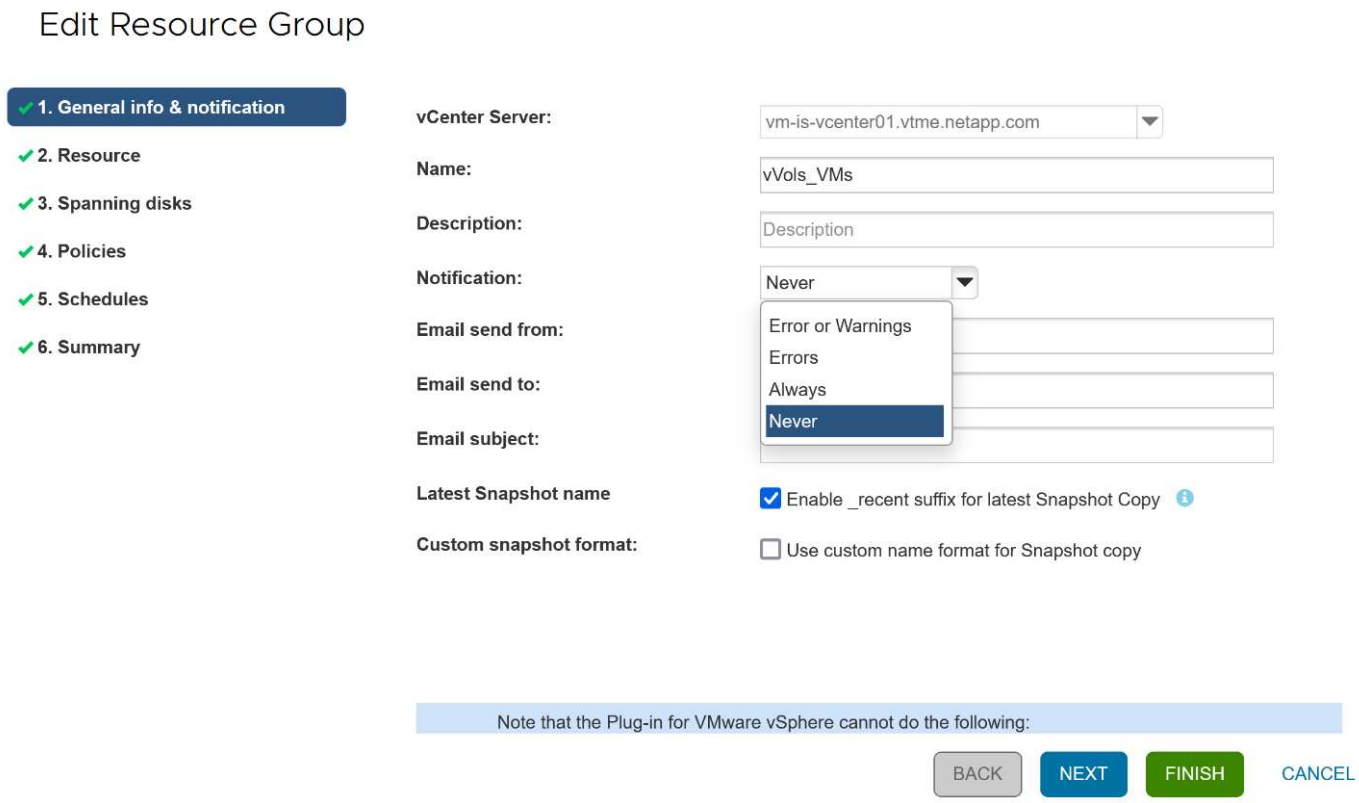
Per informazioni sulle API REST del plug-in SnapCenter per VMware vSphere, vedere "Plug-in SnapCenter per le API REST di VMware vSphere"

Best Practice

Le seguenti Best practice possono aiutarti a ottenere il massimo dalla tua implementazione SnapCenter.

- SCV supporta sia vCenter Server RBAC che ONTAP RBAC e include ruoli vCenter predefiniti che vengono creati automaticamente al momento della registrazione del plug-in. Ulteriori informazioni sui tipi di RBAC supportati "qui."
  - Utilizzare l'interfaccia utente di vCenter per assegnare l'accesso agli account con privilegi minimi utilizzando i ruoli predefiniti descritti "qui".
  - Se si utilizza SCV con il server SnapCenter, è necessario assegnare il ruolo *SnapCenterAdmin*.
  - ONTAP RBAC si riferisce all'account utente utilizzato per aggiungere e gestire i sistemi di storage utilizzati da SCV. Il role-based access control ONTAP non si applica ai backup basati su vVol. Scopri di più su ONTAP RBAC e SCV "qui".
- Replica i set di dati di backup su un secondo sistema utilizzando SnapMirror per repliche complete dei volumi di origine. Come indicato in precedenza, è anche possibile utilizzare policy di vault mirror per la conservazione a lungo termine dei dati di backup indipendentemente dalle impostazioni di conservazione delle snapshot del volume di origine. Entrambi i meccanismi sono supportati con vVol.
- Poiché SCV richiede anche strumenti ONTAP per la funzionalità vVol di VMware vSphere, controllare sempre lo strumento matrice di interoperabilità NetApp (IMT) per verificare la compatibilità delle versioni specifiche
- Se si utilizza la replica vVol con VMware SRM, prestare attenzione all'RPO delle policy e alla pianificazione del backup
- Progettare le policy di backup con impostazioni di conservazione che soddisfino gli obiettivi dei punti di ripristino (RPO) definiti dall'organizzazione
- Configurare le impostazioni di notifica sui gruppi di risorse per ricevere una notifica dello stato durante l'esecuzione dei backup (vedere la figura 10 di seguito)

Opzioni di notifica del gruppo di risorse



Iniziare a utilizzare SCV utilizzando questi documenti

"Scopri di più sul plug-in SnapCenter per VMware vSphere"

"Implementare il plug-in SnapCenter per VMware vSphere"

Risoluzione dei problemi

Sono disponibili diverse risorse per la risoluzione dei problemi con ulteriori informazioni.

Sito di supporto NetApp

Oltre a una vasta gamma di articoli della Knowledge base per i prodotti di virtualizzazione NetApp, il sito di supporto NetApp offre anche una comoda landing page per il "Strumenti ONTAP per VMware vSphere" prodotto. Questo portale fornisce link ad articoli, download, report tecnici e discussioni sulle soluzioni VMware sulla community NetApp. È disponibile all'indirizzo:

"*Sito di supporto NetApp*"

La documentazione aggiuntiva sulla soluzione è disponibile qui:

"*Soluzioni NetApp per la virtualizzazione con VMware di Broadcom*"

Risoluzione dei problemi del prodotto

I vari componenti degli strumenti ONTAP, come il plugin vCenter, il provider VASA e l'adattatore di replica dello storage, sono tutti documentati insieme nell'archivio dei documenti NetApp. Tuttavia, ciascuno di essi dispone di una sottosezione separata della Knowledge base e può disporre di procedure specifiche per la risoluzione dei problemi. Queste soluzioni risolvono i problemi più comuni che potrebbero verificarsi con il provider VASA.

Problemi dell'interfaccia utente del provider VASA

Occasionalmente, il client Web vCenter vSphere incontra problemi con i componenti di Serenity, causando la mancata visualizzazione delle voci di menu del provider VASA per ONTAP. Consultare la sezione risoluzione dei problemi di registrazione del provider VASA nella Guida all'implementazione o nella presente Knowledge base "articolo".

Il provisioning del datastore di vVol non riesce

Occasionalmente, i servizi vCenter potrebbero subire un timeout durante la creazione del datastore vVols. Per correggerlo, riavviare il servizio vmware-sps e rimontare il datastore vVols utilizzando i menu vCenter (Storage > New Datastore). Questo argomento viene trattato in vVols datastore provisioning fails with vCenter Server 6.5 nella Administration Guide.

L'aggiornamento di Unified Appliance non riesce a montare ISO

A causa di un bug in vCenter, l'ISO utilizzato per aggiornare Unified Appliance da una release alla successiva potrebbe non essere in grado di eseguire il montaggio. Se è possibile collegare l'ISO all'appliance in vCenter, seguire la procedura descritta in questa Knowledge base "articolo" per risolvere il problema.

VMware Site Recovery Manager con ONTAP

VMware Live Site Recovery con ONTAP

ONTAP è una soluzione di storage leader per VMware vSphere e, più di recente, per Cloud Foundation, da quando ESX è stato introdotto nei data center moderni più di due decenni fa. NetApp continua a introdurre sistemi innovativi, come l'ultima generazione della serie ASA A, insieme a funzionalità come la sincronizzazione attiva SnapMirror . Questi progressi semplificano la gestione, migliorano la resilienza e riducono il costo totale di proprietà (TCO) della tua infrastruttura IT.

Questo documento presenta la soluzione ONTAP per VMware Live Site Recovery (VLSR), precedentemente nota come Site Recovery Manager (SRM), il software di disaster recovery (DR) leader del settore di VMware, comprese le informazioni più recenti sul prodotto e le best practice per semplificare la distribuzione, ridurre i rischi e semplificare la gestione continua.

> Questa documentazione sostituisce il rapporto tecnico precedentemente pubblicato *TR-4900: VMware Site Recovery Manager con ONTAP*

Le Best practice integrano altri documenti come guide e strumenti di compatibilità. Sono sviluppati in base a test di laboratorio e a un'ampia esperienza sul campo da parte di tecnici e clienti NetApp. In alcuni casi, le Best practice consigliate potrebbero non essere adatte al tuo ambiente; tuttavia, sono generalmente le soluzioni più semplici che soddisfano le esigenze della maggior parte dei clienti.

Questo documento si concentra sulle funzionalità delle versioni recenti di ONTAP 9 utilizzate insieme agli strumenti ONTAP per VMware vSphere 10,4 (che include l'adattatore di replica dello storage NetApp [SRA] e il provider VASA [VP]), nonché su VMware Live Site Recovery 9.

Perché utilizzare ONTAP con VLSR o SRM?

Le piattaforme di gestione dati NetApp basate su ONTAP sono tra le soluzioni di storage più ampiamente adottate per VLSR. Le ragioni sono molteplici: una piattaforma di gestione dei dati sicura, ad alte prestazioni e

con protocollo unificato (NAS e SAN insieme) che fornisce efficienza di archiviazione senza pari nel settore, multi-tenancy, controlli della qualità del servizio, protezione dei dati con snapshot efficienti in termini di spazio e replica con SnapMirror. Tutto ciò sfrutta l'integrazione nativa multi-cloud ibrida per la protezione dei carichi di lavoro VMware e una miriade di strumenti di automazione e orchestrazione a portata di mano.

Quando si utilizza SnapMirror per la replica basata su array, si sfrutta una delle tecnologie più collaudate e mature di ONTAP. SnapMirror offre il vantaggio di trasferimenti di dati sicuri e altamente efficienti, copiando solo i blocchi del file system modificati e non intere VM o datastore. Anche questi blocchi sfruttano il risparmio di spazio, tramite deduplicazione, compressione e compattazione. I moderni sistemi ONTAP ora utilizzano SnapMirror indipendente dalla versione, consentendo flessibilità nella selezione dei cluster di origine e di destinazione. SnapMirror è diventato davvero uno degli strumenti più potenti disponibili per il disaster recovery.

Indipendentemente dal fatto che si utilizzino datastore tradizionali collegati a NFS, iSCSI o Fibre Channel (ora con supporto per datastore vVols), VLSR fornisce un'offerta proprietaria affidabile che sfrutta il meglio delle funzionalità ONTAP per il disaster recovery o la pianificazione e l'orchestrazione della migrazione del data center.

In che modo VLSR sfrutta ONTAP 9

VLSR sfrutta le tecnologie avanzate di gestione dei dati dei sistemi ONTAP integrandosi con i tool ONTAP per VMware vSphere, un'appliance virtuale che include tre componenti principali:

- Il plug-in vCenter dei tool ONTAP, in precedenza noto come Virtual Storage Console (VSC), semplifica le funzionalità di gestione ed efficienza dello storage, aumenta la disponibilità e riduce i costi dello storage e l'overhead operativo, sia che si stia utilizzando SAN o NAS. Utilizza le Best practice per il provisioning degli archivi dati e ottimizza le impostazioni degli host ESXi per gli ambienti di storage a blocchi e NFS. Per tutti questi vantaggi, NetApp consiglia questo plug-in quando si utilizza vSphere con sistemi che eseguono ONTAP.
- Il ONTAP provider VASA supporta le API vStorage di VMware per il framework VASA (Storage Awareness). IL provider VASA connette vCenter Server a ONTAP per facilitare il provisioning e il monitoraggio dello storage delle macchine virtuali. In questo modo, è stato abilitato il supporto dei volumi virtuali VMware (vVol) e la gestione delle policy storage delle macchine virtuali e delle performance dei singoli vVol delle macchine virtuali. Fornisce inoltre allarmi per il monitoraggio della capacità e della conformità con i profili.
- SRA viene utilizzato insieme a VLSR per gestire la replica dei dati delle macchine virtuali tra siti di produzione e disaster recovery per datastore VMFS e NFS tradizionali e per il test senza interruzioni delle repliche DR. Consente di automatizzare le attività di rilevamento, ripristino e protezione. Include sia un server appliance SRA che adattatori SRA per il server Windows SRM e l'appliance VLSR.

Dopo aver installato e configurato gli adattatori SRA sul server VLSR per la protezione dei datastore non vVols, è possibile iniziare l'attività di configurazione dell'ambiente vSphere per il disaster recovery.

SRA fornisce un'interfaccia di comando e controllo per il server VLSR per la gestione dei volumi ONTAP FlexVol che contengono le macchine virtuali (VM) VMware, nonché la replica SnapMirror che le protegge.

VLSR può testare il tuo piano DR in modo non invasivo utilizzando la tecnologia proprietaria FlexClone di NetApp per creare cloni quasi istantanei dei tuoi datastore protetti nel tuo sito DR. VLSR crea un sandbox per effettuare test in modo sicuro, in modo che la tua organizzazione e i tuoi clienti siano protetti in caso di un vero disastro, dandoti fiducia nella capacità della tua organizzazione di eseguire un failover durante un disastro.

In caso di disastro reale o persino di migrazione pianificata, VLSR consente di inviare eventuali modifiche dell'ultimo minuto al dataset tramite un aggiornamento finale di SnapMirror (se si sceglie di farlo). Quindi, interrompe il mirror e monta il datastore sugli host DR. A questo punto, le VM possono essere alimentate automaticamente in qualsiasi ordine in base alla strategia prepianificata.

Mentre i sistemi ONTAP permettono di accoppiare le SVM nello stesso cluster per la replica SnapMirror, questo scenario non viene testato e certificato con VLSR. Pertanto, si consiglia di utilizzare solo SVM di cluster diversi quando si utilizza VLSR.

VLSR con ONTAP e altri casi di utilizzo: Cloud ibrido e migrazione

L'integrazione della distribuzione VLSR con le funzionalità avanzate di gestione dei dati ONTAP consente di ottenere prestazioni e scalabilità notevolmente migliorate rispetto alle opzioni di archiviazione locale. Ma più di questo, offre la flessibilità del cloud ibrido. Il cloud ibrido consente di risparmiare denaro suddividendo i blocchi di dati inutilizzati dal tuo array ad alte prestazioni al tuo hyperscaler preferito tramite FabricPool, che potrebbe essere un archivio S3 locale come NetApp StorageGRID. È inoltre possibile utilizzare SnapMirror per sistemi edge-based con ONTAP Select definito dal software o DR basato su cloud utilizzando Cloud Volumes ONTAP (CVO) o "Storage NetApp su Equinix Metal®" o altri servizi ONTAP ospitati.

Quindi, grazie a FlexClone, è possibile eseguire un failover di test nel data center di un cloud service provider con un impatto dello storage prossimo allo zero. Proteggere la tua organizzazione può ora costare meno che mai.

VLSR può anche essere utilizzato per eseguire migrazioni pianificate sfruttando SnapMirror per trasferire in modo efficiente le macchine virtuali da un data center all'altro o anche all'interno dello stesso data center, sia esso il tuo, o tramite un numero qualsiasi di partner service provider NetApp.

Best practice per l'implementazione

Nelle sezioni seguenti vengono illustrate le Best practice per la distribuzione con ONTAP e VMware SRM.

Utilizzare la versione più recente di ONTAP Tools 10

Gli strumenti ONTAP 10 forniscono miglioramenti significativi rispetto alle versioni precedenti, tra cui:

- failover dei test 8x volte più veloce*
- pulizia e protezione 2x volte più veloci*
- failover più veloce del 32%*
- Maggiore scalabilità
- Supporto nativo per layout di siti condivisi

*Questi miglioramenti si basano su test interni e possono variare in base all'ambiente in uso.

Layout e segmentazione SVM per SMT

Con ONTAP, il concetto di storage virtual machine (SVM) offre una segmentazione rigorosa in ambienti multi-tenant sicuri. Gli utenti SVM su una SVM non possono accedere o gestire le risorse da un'altra. In questo modo, è possibile sfruttare la tecnologia ONTAP creando SVM separate per diverse business unit che gestiscono i propri flussi di lavoro SRM sullo stesso cluster per una maggiore efficienza dello storage globale.

Valutare la possibilità di gestire ONTAP utilizzando account con ambito SVM e LIF di gestione SVM per non solo migliorare i controlli di sicurezza, ma anche le performance. Le performance sono intrinsecamente maggiori quando si utilizzano connessioni con ambito SVM perché l'SRA non è richiesto per elaborare tutte le risorse di un intero cluster, incluse le risorse fisiche. Al contrario, l'IT deve solo comprendere le risorse logiche astratte dalla specifica SVM.

Best practice per la gestione dei sistemi ONTAP 9

Come indicato in precedenza, è possibile gestire i cluster ONTAP utilizzando credenziali cluster o SVM con ambito e LIF di gestione. Per performance ottimali, puoi prendere in considerazione l'utilizzo delle credenziali con ambito SVM ogni volta che non utilizzi vVol. Tuttavia, in questo modo, è necessario conoscere alcuni requisiti e perdere alcune funzionalità.

- L'account SVM vsadmin predefinito non dispone del livello di accesso richiesto per eseguire le attività degli strumenti ONTAP. Pertanto, devi creare un nuovo account SVM. "Configurare i ruoli e i privilegi degli utenti ONTAP" Utilizzando il file JSON incluso. Può essere utilizzato per account SVM o con ambito cluster.
- Poiché il plug-in dell'interfaccia utente vCenter, il provider VASA e il server SRA sono tutti microservizi completamente integrati, devi aggiungere storage all'adattatore SRA in SRM nello stesso modo in cui Aggiungi lo storage nell'interfaccia utente di vCenter per i tool ONTAP. In caso contrario, il server SRA potrebbe non riconoscere le richieste inviate da SRM tramite l'adattatore SRA.
- Il controllo del percorso NFS non viene eseguito quando si utilizzano credenziali con ambito SVM, a meno che non si abbia la precedenza nel gestore dei tool ONTAP e non le si "cluster integrati" associ ai vCenter. Questo perché la posizione fisica è logicamente astratta dalla SVM. Tuttavia, questo non è motivo di preoccupazione, in quanto i sistemi ONTAP moderni non subiscono più alcun calo significativo delle performance quando si utilizzano percorsi indiretti.
- Il risparmio di spazio aggregato dovuto all'efficienza dello storage potrebbe non essere segnalato.
- Se supportati, i mirror di condivisione del carico non possono essere aggiornati.
- La registrazione EMS potrebbe non essere eseguita sui sistemi ONTAP gestiti con credenziali SVM con ambito.

Best practice operative

Nelle seguenti sezioni vengono illustrate le Best practice operative per lo storage SRM e ONTAP di VMware.

Datastore e protocolli

- Se possibile, utilizza sempre gli strumenti ONTAP per eseguire il provisioning di datastore e volumi. In questo modo si garantisce che volumi, percorsi di giunzione, LUN, igroups, policy di esportazione, e altre impostazioni sono configurate in modo compatibile.
- SRM supporta iSCSI, Fibre Channel e NFS versione 3 con ONTAP 9 quando si utilizza la replica basata su array tramite SRA. SRM non supporta la replica basata su array per NFS versione 4.1 con datastore tradizionali o vVols.
- Per confermare la connettività, verificare sempre che sia possibile montare e smontare un nuovo datastore di test sul sito DR dal cluster ONTAP di destinazione. Verificare ogni protocollo che si intende utilizzare per la connettività del datastore. Una Best practice consiste nell'utilizzare gli strumenti ONTAP per creare il datastore di test, poiché sta eseguendo tutta l'automazione del datastore come indicato da SRM.
- I protocolli SAN devono essere omogenei per ciascun sito. È possibile combinare NFS e SAN, ma i protocolli SAN non devono essere combinati all'interno di un sito. Ad esempio, è possibile utilizzare FCP nel sito A e iSCSI nel sito B. non utilizzare sia FCP che iSCSI nel sito A.
- Le guide precedenti hanno consigliato la creazione di una LIF in una località dati. Vale a dire, montare sempre un datastore utilizzando una LIF situata sul nodo che fisicamente possiede il volume. Sebbene questa sia ancora la Best practice, non è più un requisito nelle moderne versioni di ONTAP 9. Quando possibile e se specifiche credenziali di ambito del cluster, i tool ONTAP continueranno a scegliere di bilanciare il carico tra le LIF locali dei dati, ma non è un requisito di high Availability o performance.

- ONTAP 9 può essere configurato in modo da rimuovere automaticamente le istantanee per preservare l'uptime in caso di esaurimento dello spazio quando il dimensionamento automatico non è in grado di fornire una capacità di emergenza sufficiente. L'impostazione predefinita di questa funzionalità non elimina automaticamente le snapshot create da SnapMirror. Se le snapshot SnapMirror vengono eliminate, il servizio SRA di NetApp non può invertire e risincronizzare la replica per il volume interessato. Per evitare che ONTAP elimini gli snapshot SnapMirror, configurare la funzionalità di eliminazione automatica degli snapshot su 'Try'.

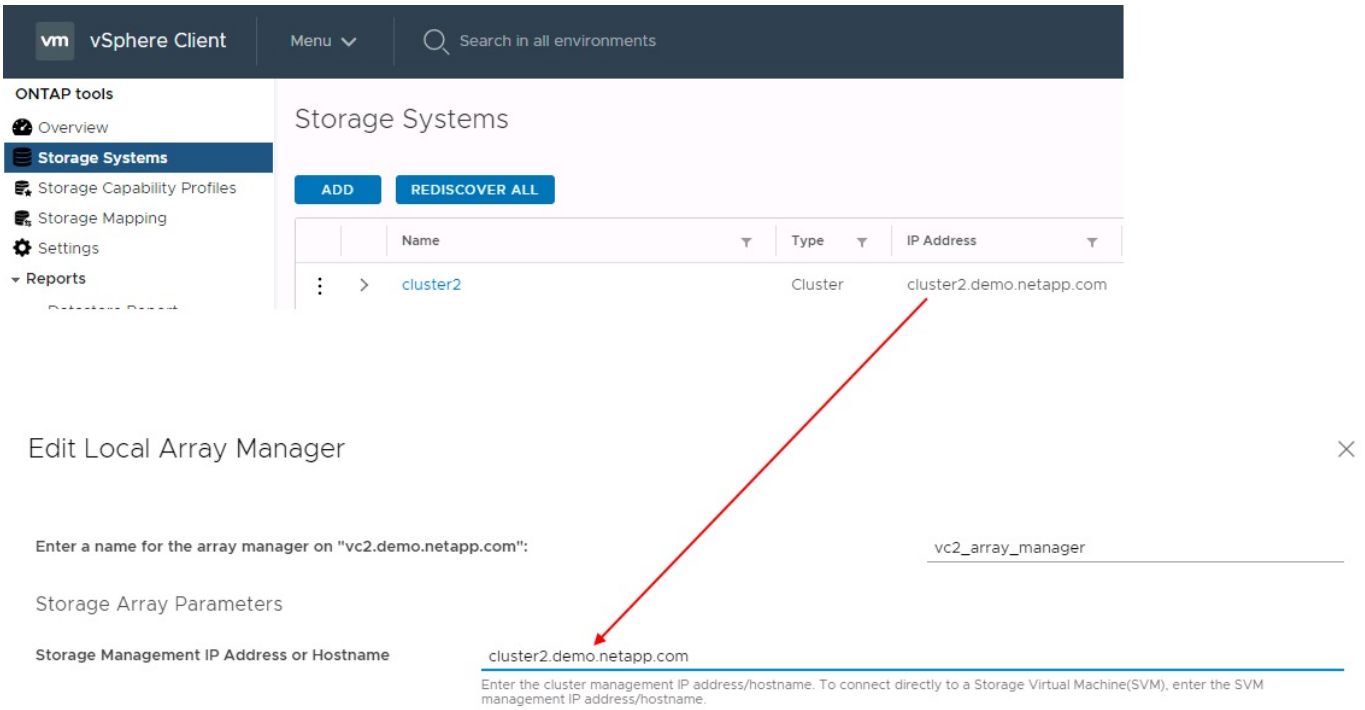
```
snap autodelete modify -volume -commitment try
```

- Il dimensionamento automatico del volume deve essere impostato su `grow` per i volumi che contengono datastore SAN e `grow_shrink` per i datastore NFS. Ulteriori informazioni su questo argomento sono disponibili all'indirizzo "Configurare i volumi per aumentare e ridurre automaticamente le dimensioni".
- SRM funziona al meglio quando il numero di datastore e quindi di gruppi di protezione viene ridotto al minimo nei piani di ripristino. È quindi opportuno prendere in considerazione l'ottimizzazione della densità delle macchine virtuali negli ambienti protetti con SRM in cui l'RTO è fondamentale.
- Utilizza DRS (Distributed Resource Scheduler) per bilanciare il carico sui cluster ESXi protetti e di recovery. Tenere presente che se si prevede di eseguire il failback, quando si esegue una nuova protezione i cluster precedentemente protetti diventeranno i nuovi cluster di ripristino. Il DRS aiuterà a bilanciare il posizionamento in entrambe le direzioni.
- Ove possibile, evitare di utilizzare la personalizzazione IP con SRM, poiché ciò può aumentare il vostro RTO.

Informazioni sulle coppie di array

Viene creato un gestore di array per ogni coppia di array. Con gli strumenti SRM e ONTAP, ogni accoppiamento di array viene eseguito con l'ambito di una SVM, anche se si utilizzano le credenziali del cluster. Ciò consente di segmentare i flussi di lavoro DR tra tenant in base alle SVM assegnate per la gestione. È possibile creare più array manager per un determinato cluster e possono essere asimmetrici. È possibile eseguire il fan-out o il fan-in tra diversi cluster di ONTAP 9. Ad esempio, è possibile utilizzare SVM-A e SVM-B nel cluster-1 in replica su SVM-C nel cluster-2, SVM-D nel cluster-3 o viceversa.

Quando si configurano le coppie di array in SRM, è necessario aggiungerle sempre in SRM nello stesso modo in cui sono state aggiunte agli strumenti ONTAP, ovvero devono utilizzare lo stesso nome utente, password e LIF di gestione. Questo requisito garantisce che SRA comunichi correttamente con l'array. La seguente schermata illustra come potrebbe essere visualizzato un cluster negli strumenti ONTAP e come potrebbe essere aggiunto a un gestore di array.



Informazioni sui gruppi di replica

I gruppi di replica contengono raccolte logiche di macchine virtuali che vengono ripristinate insieme. Poiché la replica di ONTAP SnapMirror avviene a livello di volume, tutte le macchine virtuali di un volume si trovano nello stesso gruppo di replica.

Esistono diversi fattori da considerare per i gruppi di replica e il modo in cui si distribuiscono le macchine virtuali tra i volumi FlexVol. Il raggruppamento di macchine virtuali simili nello stesso volume può aumentare l'efficienza dello storage con i sistemi ONTAP meno recenti che non dispongono di una deduplica a livello di aggregato, ma il raggruppamento aumenta la dimensione del volume e riduce l'simultaneità dell'i/O. Il miglior equilibrio tra performance ed efficienza dello storage si può ottenere negli attuali sistemi ONTAP distribuendo le VM su volumi FlexVol nello stesso aggregato, sfruttando così la deduplica a livello di aggregato e ottenendo una maggiore parallelizzazione i/o su più volumi. È possibile ripristinare le macchine virtuali nei volumi insieme perché un gruppo di protezione (discusso di seguito) può contenere più gruppi di replica. Lo svantaggio di questo layout è che i blocchi potrebbero essere trasmessi più volte via cavo perché SnapMirror non prende in considerazione la deduplica aggregata.

Un'ultima considerazione per i gruppi di replica è che ciascuno di essi è per sua natura un gruppo di coerenza logica (da non confondere con i gruppi di coerenza SRM). Questo perché tutte le VM nel volume vengono trasferite insieme utilizzando lo stesso snapshot. Pertanto, se si dispone di macchine virtuali che devono essere coerenti tra loro, è consigliabile memorizzarle nello stesso FlexVol.

A proposito dei gruppi di protezione

I gruppi di protezione definiscono macchine virtuali e datastore in gruppi che vengono ripristinati insieme dal sito protetto. Il sito protetto è il luogo in cui esistono le macchine virtuali configurate in un gruppo di protezione durante le normali operazioni in stato stazionario. È importante notare che anche se SRM potrebbe visualizzare più gestori di array per un gruppo di protezione, un gruppo di protezione non può estendersi a più gestori di array. Per questo motivo, non è necessario estendere i file delle macchine virtuali tra gli archivi dati su macchine virtuali SVM diverse.

Sui piani di recovery

I piani di recovery definiscono quali gruppi di protezione vengono ripristinati nello stesso processo. È possibile configurare più gruppi di protezione nello stesso piano di ripristino. Inoltre, per abilitare più opzioni per l'esecuzione dei piani di ripristino, è possibile includere un singolo gruppo di protezione in più piani di ripristino.

I piani di recovery consentono agli amministratori SRM di definire i flussi di lavoro di recovery assegnando le macchine virtuali a un gruppo di priorità da 1 (massimo) a 5 (minimo), con 3 (medio) come valore predefinito. All'interno di un gruppo di priorità, le VM possono essere configurate per le dipendenze.

Ad esempio, la tua azienda potrebbe disporre di un'applicazione business-critical Tier 1 che si affida a un server Microsoft SQL per il proprio database. Quindi, si decide di inserire le macchine virtuali nel gruppo di priorità 1. All'interno del gruppo di priorità 1, si inizia a pianificare l'ordine per visualizzare i servizi. Probabilmente si desidera che il controller di dominio Microsoft Windows si avvii prima del server Microsoft SQL, che deve essere online prima del server di applicazioni e così via. È necessario aggiungere tutte queste macchine virtuali al gruppo di priorità e quindi impostare le dipendenze perché le dipendenze si applicano solo all'interno di un determinato gruppo di priorità.

NetApp consiglia vivamente di collaborare con i team delle applicazioni per comprendere l'ordine delle operazioni richieste in uno scenario di failover e per costruire di conseguenza i piani di recovery.

Test del failover

Come Best practice, eseguire sempre un failover di test ogni volta che viene apportata una modifica alla configurazione dello storage protetto delle macchine virtuali. In questo modo, in caso di emergenza, è possibile verificare che Site Recovery Manager sia in grado di ripristinare i servizi entro la destinazione RTO prevista.

NetApp consiglia inoltre di confermare occasionalmente la funzionalità delle applicazioni in-guest, soprattutto dopo la riconfigurazione dello storage delle macchine virtuali.

Quando viene eseguita un'operazione di test recovery, viene creata una rete bubble di test privata sull'host ESXi per le macchine virtuali. Tuttavia, questa rete non è connessa automaticamente ad alcun adattatore di rete fisico e pertanto non fornisce connettività tra gli host ESXi. Per consentire la comunicazione tra macchine virtuali in esecuzione su host ESXi diversi durante il test di DR, viene creata una rete fisica privata tra gli host ESXi nel sito di DR. Per verificare che la rete di test sia privata, è possibile separare fisicamente la rete a bolle di test oppure utilizzando VLAN o tag VLAN. Questa rete deve essere separata dalla rete di produzione, in quanto non è possibile posizionare le macchine virtuali sulla rete di produzione con indirizzi IP che potrebbero entrare in conflitto con i sistemi di produzione effettivi. Quando viene creato un piano di ripristino in SRM, la rete di test creata può essere selezionata come rete privata a cui connettere le macchine virtuali durante il test.

Una volta convalidato il test e non più necessario, eseguire un'operazione di pulizia. L'esecuzione della pulizia riporta le macchine virtuali protette al loro stato iniziale e ripristina il piano di ripristino allo stato Pronta.

Considerazioni sul failover

Oltre all'ordine delle operazioni indicato in questa guida, è necessario considerare anche altri aspetti relativi al failover di un sito.

Un problema che potrebbe essere dovuto affrontare è rappresentato dalle differenze di rete tra i siti. Alcuni ambienti potrebbero essere in grado di utilizzare gli stessi indirizzi IP di rete sia nel sito primario che nel sito di DR. Questa capacità viene definita come una LAN virtuale estesa (VLAN) o una configurazione di rete estesa. Altri ambienti potrebbero richiedere l'utilizzo di indirizzi IP di rete diversi (ad esempio, in VLAN diverse) nel sito primario rispetto al sito di DR.

VMware offre diversi modi per risolvere questo problema. Per prima cosa, le tecnologie di virtualizzazione di

rete come VMware NSX-T Data Center astraggono l'intero stack di rete dai livelli 2 fino a 7 dall'ambiente operativo, consentendo soluzioni più portatili. Scopri di più "Opzioni NSX-T con SRM".

SRM consente inoltre di modificare la configurazione di rete di una macchina virtuale durante il ripristino. Questa riconfigurazione include impostazioni quali indirizzi IP, indirizzi gateway e impostazioni del server DNS. È possibile specificare diverse impostazioni di rete, che vengono applicate alle singole macchine virtuali non appena vengono recuperate, nelle impostazioni della proprietà di una macchina virtuale nel piano di ripristino.

Per configurare SRM in modo che applichi impostazioni di rete diverse a più macchine virtuali senza dover modificare le proprietà di ciascuna di esse nel piano di ripristino, VMware fornisce uno strumento chiamato dr-ip-customizer. Per informazioni sull'utilizzo di questa utilità, fare riferimento alla sezione "Documentazione di VMware".

Proteggere di nuovo

Dopo un ripristino, il sito di ripristino diventa il nuovo sito di produzione. Poiché l'operazione di ripristino ha rotto la replica di SnapMirror, il nuovo sito di produzione non è protetto da eventuali disastri futuri. Una Best practice consiste nel proteggere il nuovo sito di produzione in un altro sito immediatamente dopo un ripristino. Se il sito di produzione originale è operativo, l'amministratore di VMware può utilizzare il sito di produzione originale come nuovo sito di ripristino per proteggere il nuovo sito di produzione, invertendo efficacemente la direzione della protezione. La protezione è disponibile solo in caso di guasti non catastrofici. Pertanto, i server vCenter originali, i server ESXi, i server SRM e i database corrispondenti devono essere ripristinabili. Se non sono disponibili, è necessario creare un nuovo gruppo di protezione e un nuovo piano di ripristino.

Failback

Un'operazione di failback è fondamentalmente un failover in una direzione diversa rispetto a prima. Come Best practice, prima di tentare di eseguire il failback o, in altre parole, di eseguire il failover sul sito originale, è necessario verificare che il sito originale sia tornato a livelli di funzionalità accettabili. Se il sito originale è ancora compromesso, è necessario ritardare il failback fino a quando il guasto non viene risolto in modo adeguato.

Un'altra Best practice per il failback consiste nell'eseguire sempre un failover di test dopo aver completato la protezione e prima di eseguire il failback finale. In questo modo si verifica che i sistemi installati presso il sito originale possano completare l'operazione.

Protezione del sito originale

Dopo il failback, è necessario confermare con tutti gli stakeholder che i loro servizi sono stati riportati alla normalità prima di eseguire nuovamente la funzione di protezione,

L'esecuzione di una nuova protezione dopo il failback riporta sostanzialmente l'ambiente nello stato in cui si trovava all'inizio, con la replica di SnapMirror nuovamente in esecuzione dal sito di produzione al sito di ripristino.

Topologie di replica

In ONTAP 9, i componenti fisici di un cluster sono visibili agli amministratori del cluster, ma non sono direttamente visibili alle applicazioni e agli host che utilizzano il cluster. I componenti fisici forniscono un pool di risorse condivise da cui vengono costruite le risorse del cluster logico. Le applicazioni e gli host accedono ai dati solo tramite SVM che contengono volumi e LIF.

Ogni SVM NetApp viene trattata come array in VMware vCenter Site Recovery Manager. VLSR supporta determinati layout di replica array-to-array (o SVM-to-SVM).

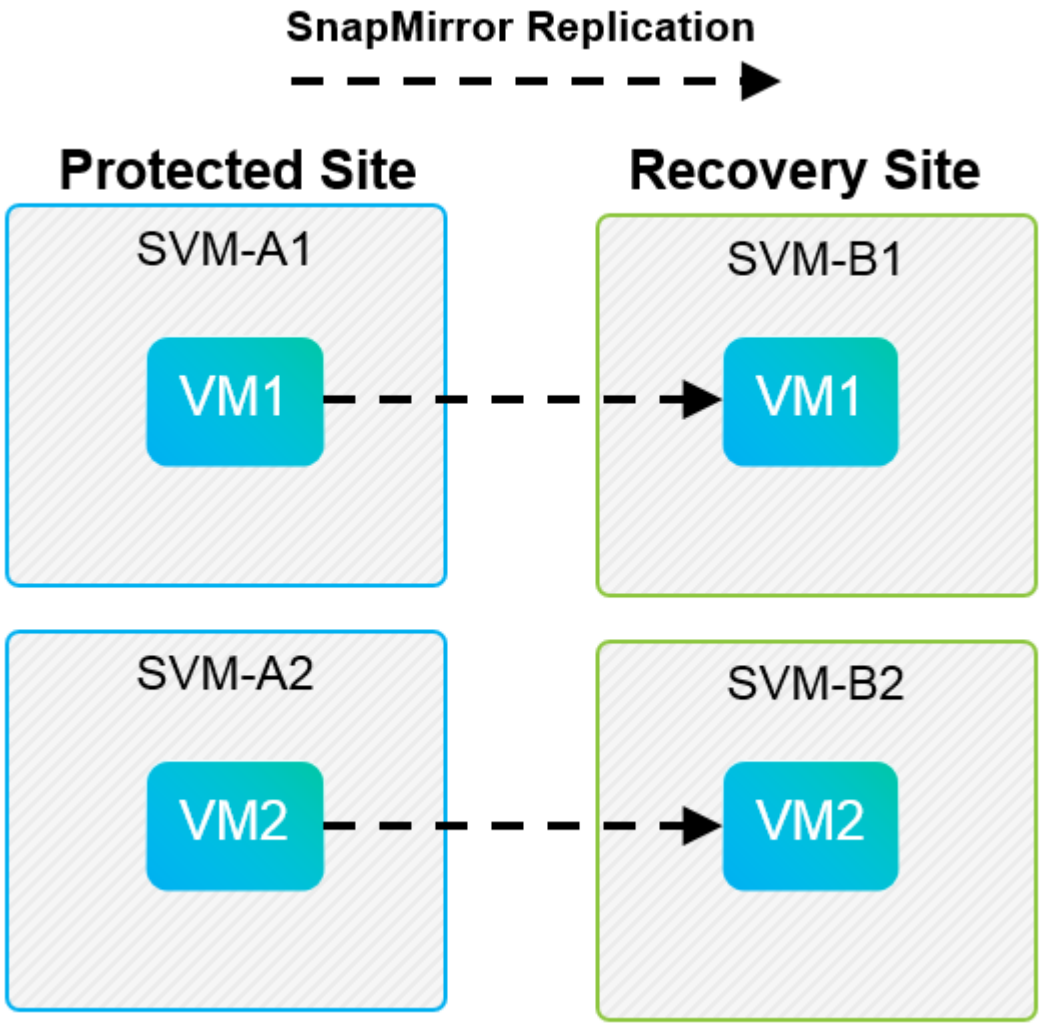
Una singola macchina virtuale non è in grado di gestire i dati (VMDK) o RDM) su più array VLSR per i seguenti motivi:

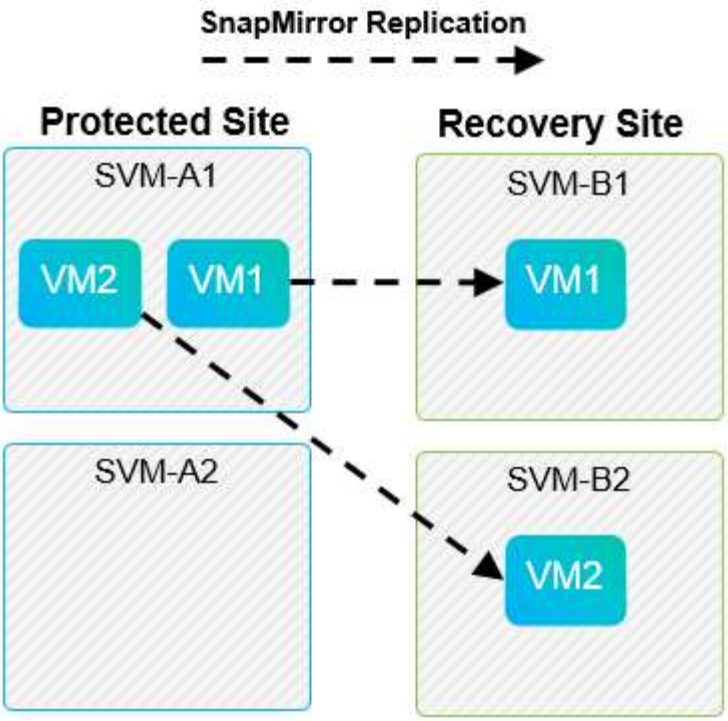
- VLSR vede solo la SVM, non un singolo controller fisico.
- Una SVM può controllare LUN e volumi che si estendono su più nodi in un cluster.

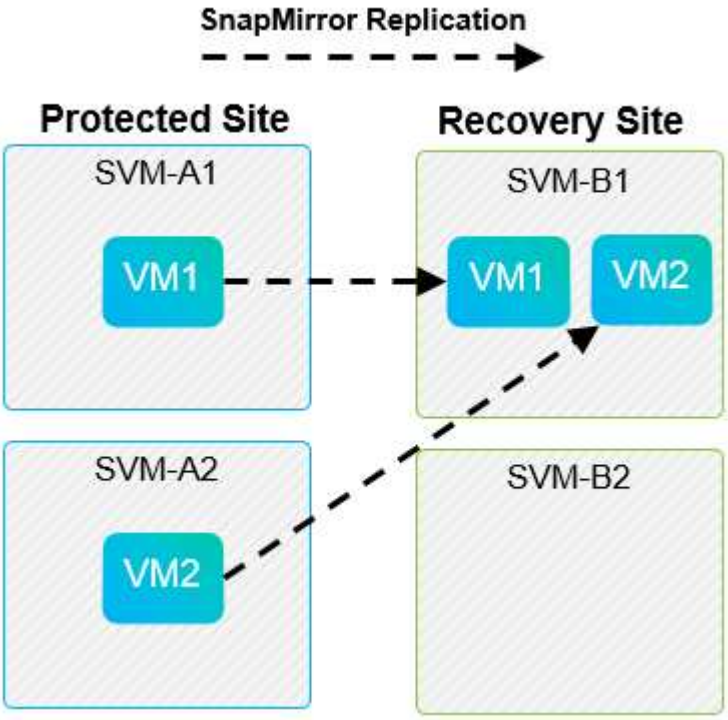
Best practice
Per determinare la supportabilità, tenere presente questa regola: Per proteggere una macchina virtuale utilizzando VLSR e NetApp SRA, tutte le parti della macchina virtuale devono esistere su un solo SVM. Questa regola si applica sia al sito protetto che al sito di ripristino.

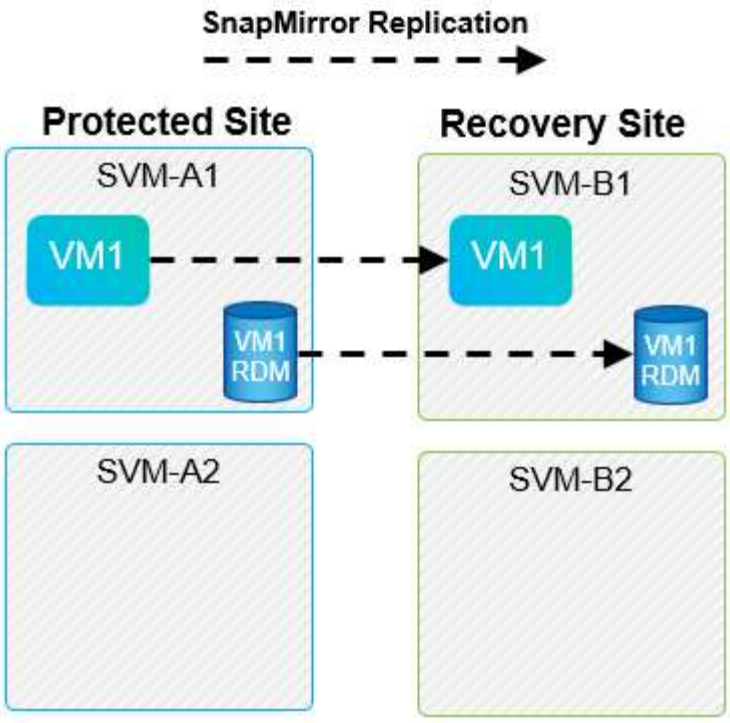
Layout SnapMirror supportati

Le seguenti figure mostrano gli scenari di layout delle relazioni SnapMirror supportati da VLSR e SRA. Ogni macchina virtuale nei volumi replicati possiede i dati su un solo array VLSR (SVM) in ogni sito.



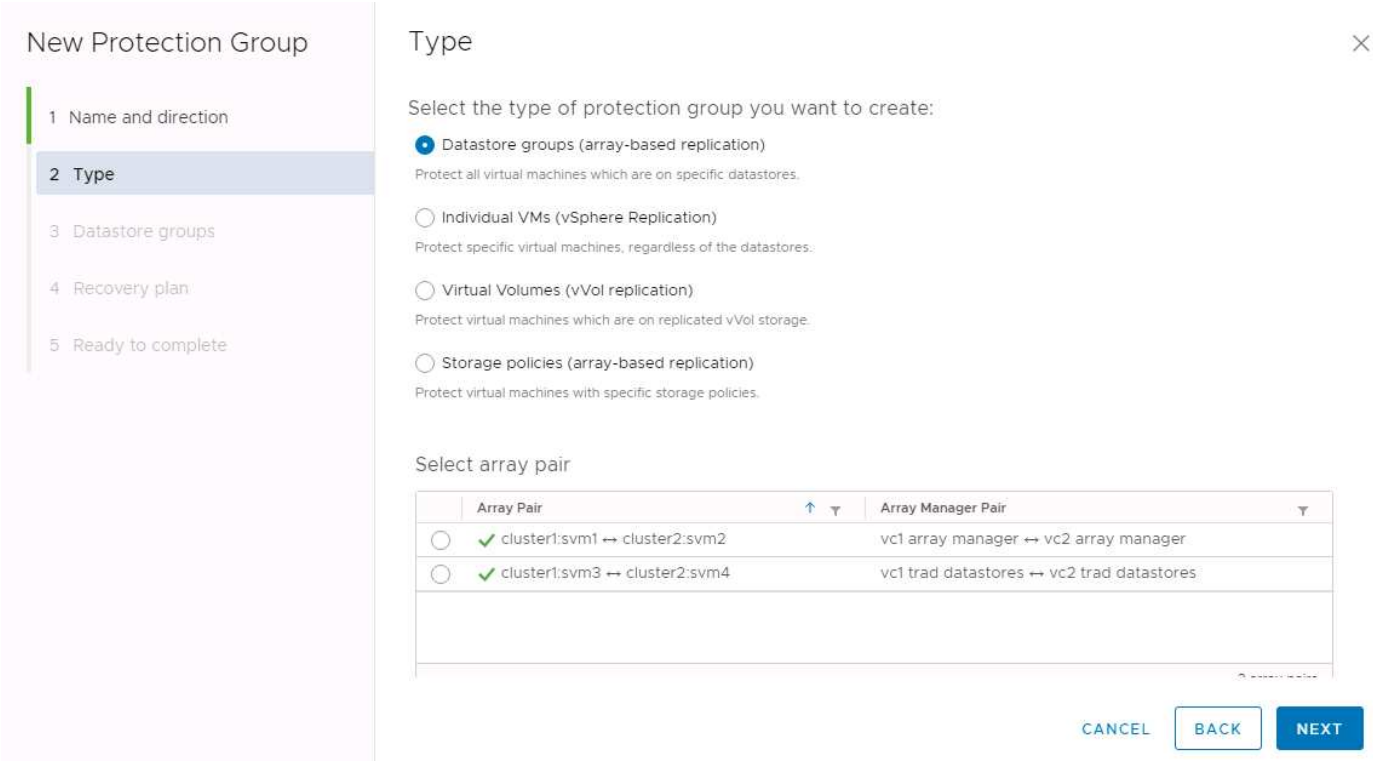

SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
SVM-B1
VM2
VM1
VM1
SVM-A2
SVM-B2
VM2


SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
SVM-B1
VM1
VM1
VM2
SVM-A2
SVM-B2
VM2



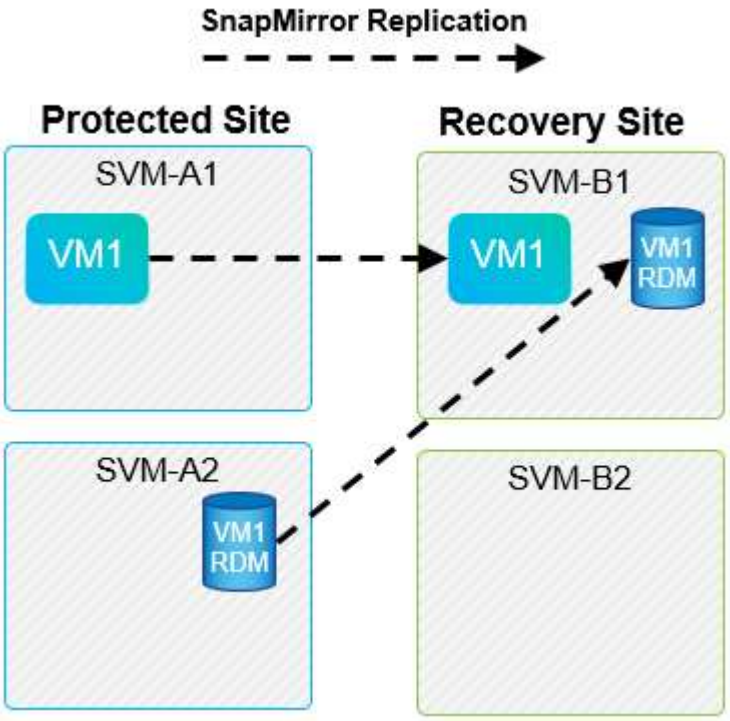
Layout di Array Manager supportati

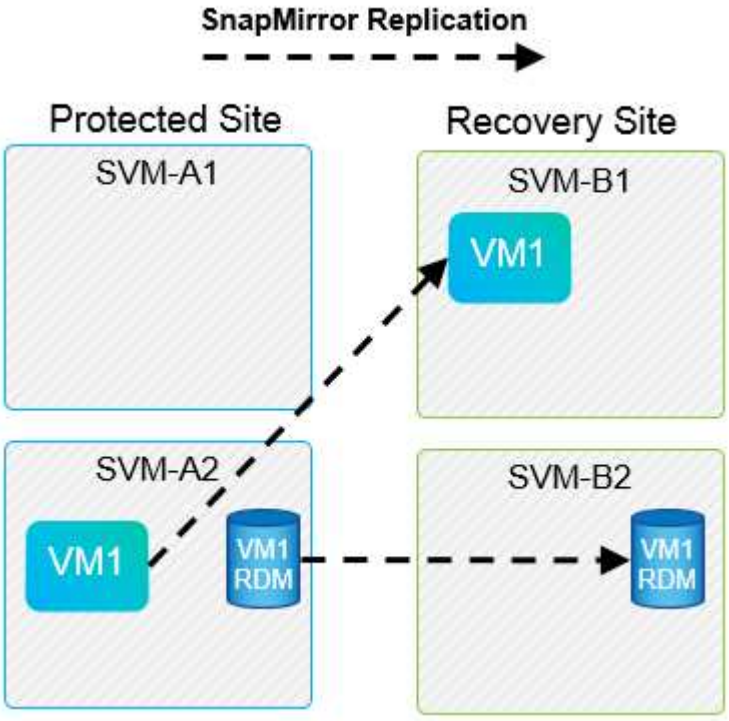
Quando si utilizza la replica basata su array (ABR) in VLSR, i gruppi di protezione vengono isolati in una singola coppia di array, come illustrato nella seguente schermata. In questo scenario, `SVM1` e `SVM2` vengono sottoposti a peed con `SVM3` e `SVM4` nel sito di recovery. Tuttavia, è possibile selezionare solo una delle due coppie di array quando si crea un gruppo di protezione.



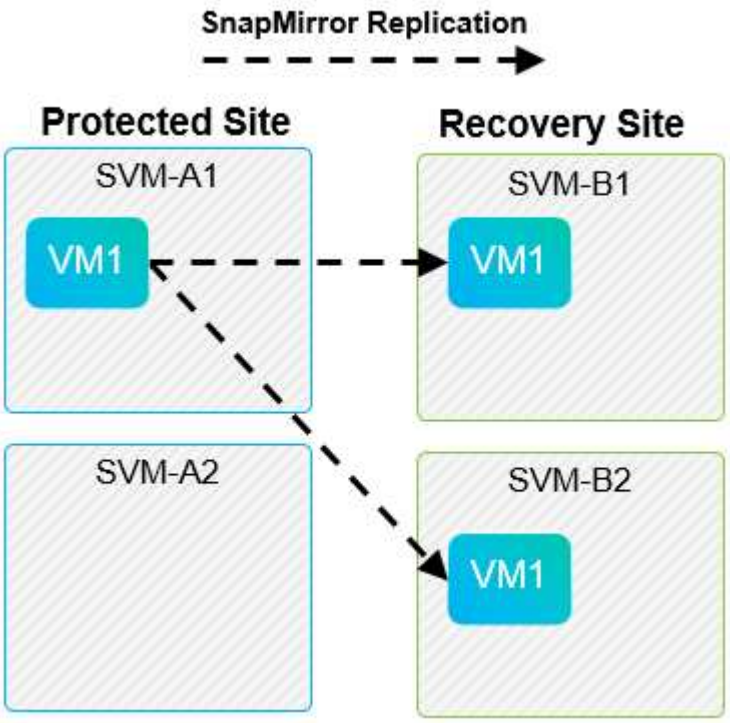
Layout non supportati

Le configurazioni non supportate dispongono di dati (VMDK o RDM) su più SVM di proprietà di una singola macchina virtuale. Negli esempi mostrati nelle seguenti figure, `VM1` non è possibile configurare la protezione con VLSR perché `VM1` dispone di dati su due SVM.



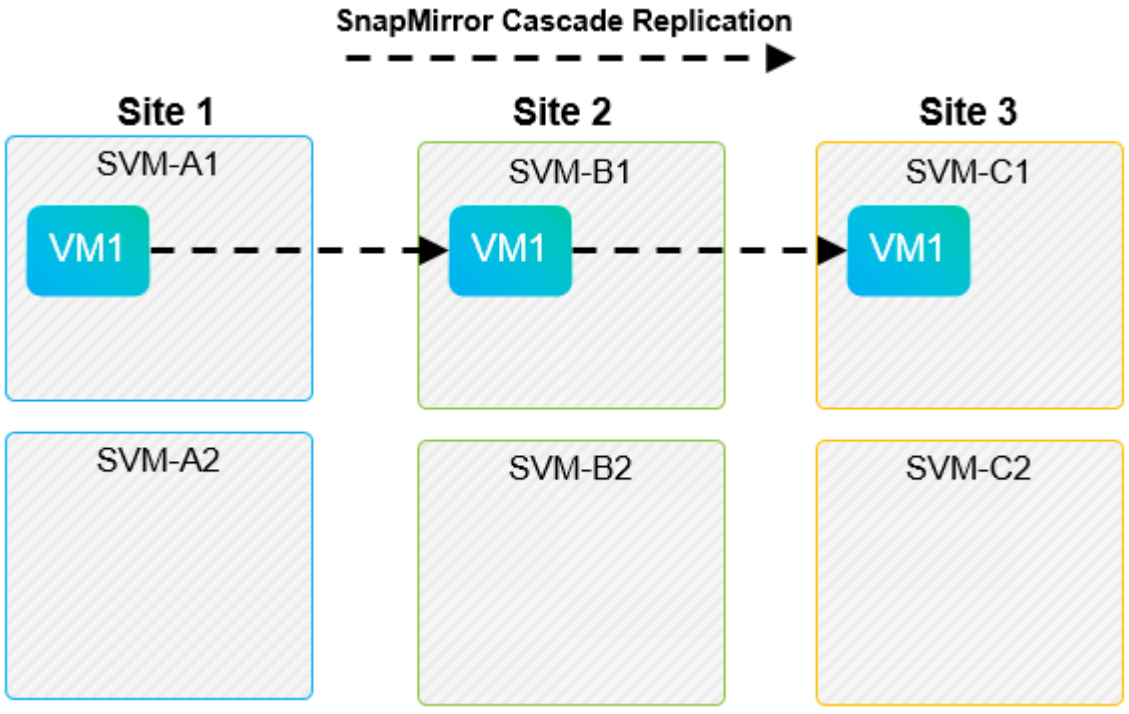


Qualsiasi relazione di replica in cui un singolo volume NetApp viene replicato da una SVM di origine a più destinazioni nella stessa SVM o in SVM differenti viene definita fan-out di SnapMirror. Fan-out non supportato con VLSR. Nell'esempio mostrato nella figura seguente, `VM1` non è possibile configurare la protezione in VLSR perché viene replicata con SnapMirror in due posizioni diverse.



Cascata di SnapMirror

VLSR non supporta la sovrapposizione delle relazioni SnapMirror, in cui un volume di origine viene replicato in un volume di destinazione e tale volume di destinazione viene replicato anche con SnapMirror in un altro volume di destinazione. Nello scenario illustrato nella figura seguente, VLSR non può essere utilizzato per il failover tra siti.



SnapMirror e SnapVault

Il software NetApp SnapVault consente il backup basato su disco dei dati aziendali tra i sistemi storage

NetApp. SnapVault e SnapMirror possono coesistere nello stesso ambiente; tuttavia, VLSR supporta il failover solo delle relazioni SnapMirror.

NetApp SRA supporta `mirror-vault` tipo di policy.

SnapVault è stato ricostruito da zero per ONTAP 8.2. Anche se gli utenti di Data ONTAP 7-Mode precedenti dovrebbero trovare delle analogie, in questa versione di SnapVault sono stati apportati importanti miglioramenti. Un importante progresso è la capacità di preservare l'efficienza dello storage sui dati primari durante i trasferimenti SnapVault.

Un'importante modifica architetturale è che SnapVault in ONTAP 9 replica a livello di volume anziché a livello di qtree, come nel caso di 7-Mode SnapVault. Questa configurazione indica che l'origine di una relazione SnapVault deve essere un volume e che tale volume deve replicarsi nel proprio volume sul sistema secondario SnapVault.

In un ambiente in cui viene utilizzato SnapVault, vengono create snapshot specificatamente denominate sul sistema di storage primario. A seconda della configurazione implementata, gli snapshot denominati possono essere creati sul sistema primario da una pianificazione SnapVault o da un'applicazione come NetApp Active IQ Unified Manager. Gli Snapshot con nome creati sul sistema primario vengono quindi replicati nella destinazione SnapMirror, da dove vengono trasferiti in un vault nella destinazione SnapVault.

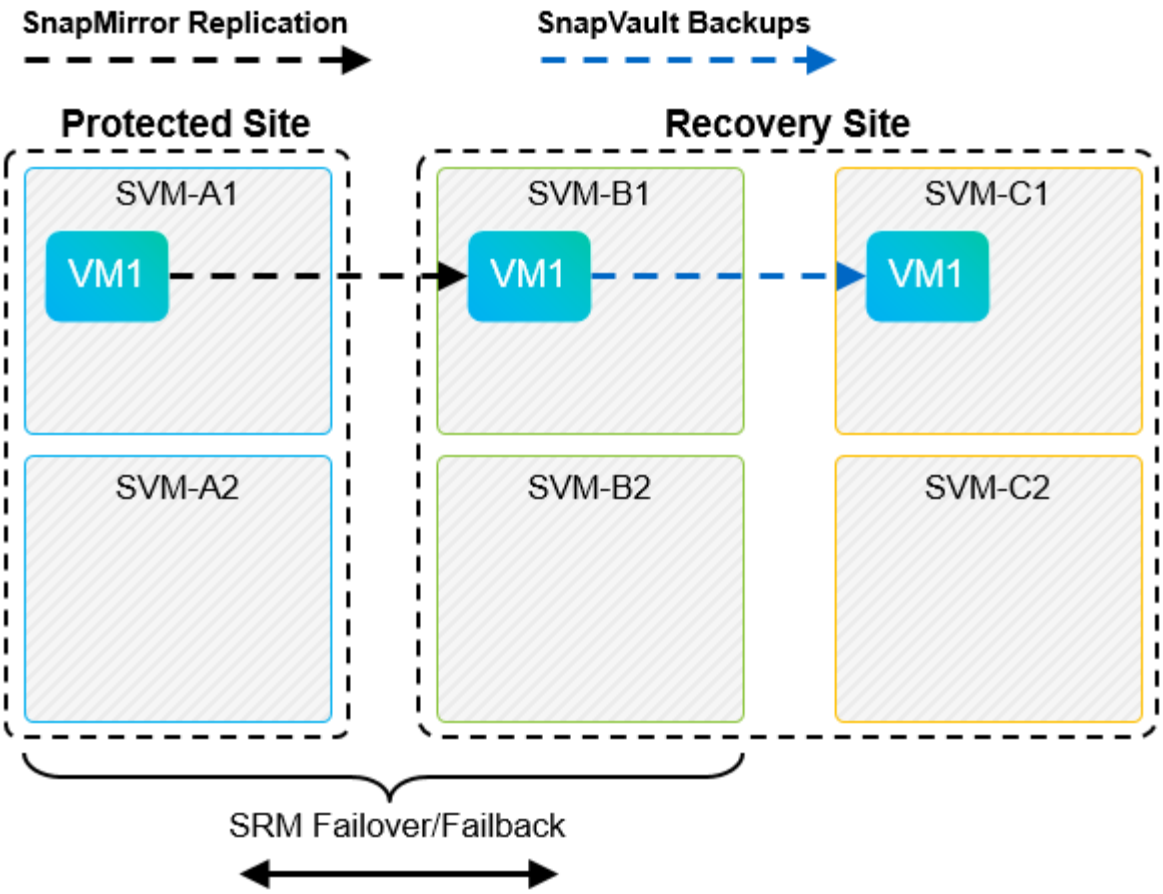
È possibile creare un volume di origine in una configurazione a cascata in cui un volume viene replicato in una destinazione SnapMirror nel sito DR e da qui viene vault in una destinazione SnapVault. È possibile creare un volume di origine anche in una relazione fan-out in cui una destinazione è una destinazione SnapMirror e l'altra destinazione è una destinazione SnapVault. Tuttavia, SRA non riconfigurerà automaticamente la relazione SnapVault per utilizzare il volume di destinazione SnapMirror come origine per il vault quando si verifica il failover VLSR o l'inversione della replica.

Per informazioni aggiornate su SnapMirror e SnapVault per ONTAP 9, vedere "Guida alle Best practice per la configurazione di SnapMirror TR-4015 per ONTAP 9."

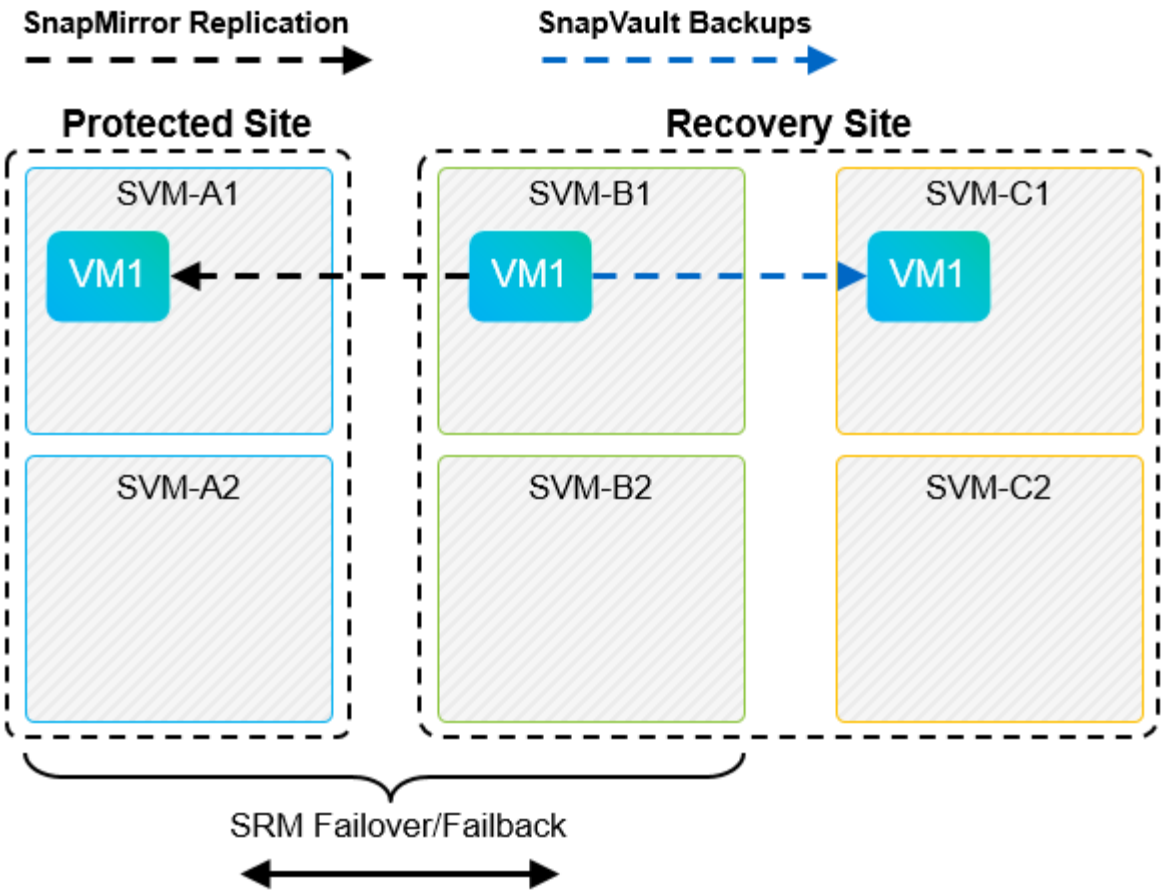
Best practice
Se SnapVault e VLSR vengono utilizzati nello stesso ambiente, NetApp consiglia di utilizzare una configurazione a cascata da SnapMirror a SnapVault in cui i backup di SnapVault vengono normalmente eseguiti dalla destinazione di SnapMirror nel sito di DR. In caso di disastro, questa configurazione rende il sito primario inaccessibile. Mantenendo la destinazione SnapVault nel sito di recovery, è possibile riconfigurare i backup SnapVault dopo il failover in modo che i backup SnapVault possano continuare mentre si opera nel sito di recovery.

In un ambiente VMware, ogni datastore dispone di un UUID (Universal Unique Identifier) e ogni VM dispone di un MOID (Managed Object ID) univoco. Questi ID non vengono gestiti da VLSR durante il failover o il failback. Poiché gli UUID degli archivi di dati e i MOID delle macchine virtuali non vengono mantenuti durante il failover da VLSR, tutte le applicazioni che dipendono da questi ID devono essere riconfigurate dopo il failover di VLSR. Un'applicazione di esempio è NetApp Active IQ Unified Manager, che coordina la replica SnapVault con l'ambiente vSphere.

La figura seguente mostra una configurazione a cascata da SnapMirror a SnapVault. Se la destinazione SnapVault si trova nel sito di DR o in un sito terzo che non è interessato da un'interruzione nel sito primario, l'ambiente può essere riconfigurato per consentire ai backup di continuare dopo il failover.



La seguente figura illustra la configurazione dopo l'utilizzo di VLSR per eseguire il reverse della replica di SnapMirror nel sito primario. L'ambiente è stato anche riconfigurato in modo che i backup di SnapVault si verifichino da quella che ora è l'origine di SnapMirror. Questa configurazione è una configurazione fan-out di SnapMirror SnapVault.



Dopo che vsrm esegue il failback e una seconda inversione delle relazioni SnapMirror, i dati di produzione vengono ripristinati nel sito primario. Questi dati sono ora protetti nello stesso modo in cui erano prima del failover al sito di DR, tramite i backup SnapMirror e SnapVault.

Utilizzo di Qtree in ambienti Site Recovery Manager

I qtree sono directory speciali che consentono l'applicazione delle quote del file system per NAS. ONTAP 9 consente la creazione di qtree e qtree possono esistere in volumi replicati con SnapMirror. Tuttavia, SnapMirror non consente la replica di singoli qtree o replica a livello di qtree. Tutte le repliche di SnapMirror sono solo a livello di volume. Per questo motivo, NetApp sconsiglia l'utilizzo di qtree con VLSR.

Ambienti misti FC e iSCSI

Con i protocolli SAN supportati (FC, FCoE e iSCSI), ONTAP 9 offre servizi LUN, ovvero la possibilità di creare e mappare LUN agli host collegati. Poiché il cluster è costituito da più controller, esistono più percorsi logici gestiti da i/o multipath verso qualsiasi LUN individuale. L'ALUA (Asymmetric Logical Unit Access) viene utilizzato sugli host in modo che il percorso ottimizzato per un LUN sia selezionato e reso attivo per il trasferimento dei dati. Se il percorso ottimizzato per qualsiasi LUN cambia (ad esempio, perché il volume contenente viene spostato), ONTAP 9 riconosce automaticamente e regola senza interruzioni per questa modifica. Se il percorso ottimizzato non è disponibile, ONTAP può passare senza interruzioni a qualsiasi altro percorso disponibile.

VMware VLSR e NetApp SRA supportano l'utilizzo del protocollo FC in un sito e del protocollo iSCSI nell'altro. Tuttavia, non supporta la combinazione di datastore FC-attached e datastore iSCSI-attached nello stesso host ESXi o in host diversi nello stesso cluster. Questa configurazione non è supportata con VLSR perché, durante il failover VLSR o il failover di test, VLSR include tutti gli iniziatori FC e iSCSI negli host ESXi nella richiesta.

Best practice
VLSR e SRA supportano protocolli FC e iSCSI misti tra i siti protetti e di ripristino. Tuttavia, ogni sito deve essere configurato con un solo protocollo, FC o iSCSI, non entrambi nello stesso sito. Se esiste un requisito per la configurazione dei protocolli FC e iSCSI nello stesso sito, NetApp consiglia che alcuni host utilizzino iSCSI e altri host utilizzino FC. In questo caso, NetApp consiglia anche di configurare le mappature delle risorse VLSR in modo che le macchine virtuali siano configurate per il failover in un gruppo di host o nell'altro.

Risoluzione dei problemi relativi a VLSRM/SRM quando si utilizza la replica vVols

Quando si utilizzano gli strumenti ONTAP 9.13P2, il flusso di lavoro all'interno di VLSR e SRM è notevolmente diverso quando si utilizza la replica vVol da ciò che viene utilizzato con SRA e i datastore tradizionali. Ad esempio, non esiste alcun concetto di gestore di array. Come tali, `discoverarrays` e `discoverdevices` i comandi non vengono mai visti.

Durante la risoluzione dei problemi, è utile comprendere i nuovi flussi di lavoro, elencati di seguito:

1. QueryReplicationPeer: Rileva gli accordi di replica tra due domini di errore.
2. QueryFaultDomain: Rileva la gerarchia di dominio di errore.
3. QueryReplicationGroup: Consente di individuare i gruppi di replica presenti nei domini di origine o di destinazione.
4. SyncReplicationGroup: Sincronizza i dati tra origine e destinazione.
5. QueryPointInTimeReplica: Consente di rilevare le repliche point-in-time di una destinazione.
6. TestFailoverReplicationGroupStart: Avvia il failover del test.
7. TestFailoverReplicationGroupStop: Termina il failover del test.
8. PromoteReplicationGroup: Promuove un gruppo attualmente in fase di test in produzione.
9. PrepareFailoverReplicationGroup: Prepara per un disaster recovery.
10. FailoverReplicationGroup: Esegue il disaster recovery.
11. ReverseReplicateGroup: Avvia la replica inversa.
12. QueryMatchingContainer: Trova i container (insieme agli host o ai gruppi di replica) che potrebbero soddisfare una richiesta di provisioning con una determinata policy.
13. QueryResourceMetadata: Rileva i metadati di tutte le risorse dal provider VASA, l'utilizzo delle risorse può essere restituito come risposta alla funzione QueryMatchingContainer.

L'errore più comune riscontrato durante la configurazione della replica di vVol è il mancato rilevamento delle relazioni di SnapMirror. Ciò si verifica perché i volumi e le relazioni di SnapMirror vengono creati al di fuori dell'ambito di applicazione degli strumenti ONTAP. Pertanto, è consigliabile assicurarsi sempre che la relazione di SnapMirror sia completamente inizializzata e che sia stata eseguita una riscoperta negli strumenti ONTAP in entrambi i siti prima di tentare di creare un datastore vVol replicato.

Ulteriori informazioni

Per ulteriori informazioni sulle informazioni descritte in questo documento, consultare i seguenti documenti e/o siti Web:

- Tool ONTAP per le risorse di VMware vSphere 10.x
"https://mysupport.netapp.com/site/products/all/details/otv10/docs-tab"
- Tool ONTAP per le risorse di VMware vSphere 9.x
"https://mysupport.netapp.com/site/products/all/details/otv/docsandkb-tab"
- TR-4597: VMware vSphere per ONTAP
"https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vsphere-overview.html"
- TR-4400: Volumi virtuali VMware vSphere con ONTAP
"https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vvols-overview.html"
- Guida alle Best practice per la configurazione di SnapMirror TR-4015 per ONTAP 9
https://www.netapp.com/media/17229-tr4015.pdf?v=127202175503P
- Documentazione di VMware Live Site Recovery "https://techdocs.broadcom.com/us/en/vmware-cis/live-recovery/live-site-recovery/9-0.html"

Per verificare se le versioni delle funzionalità e dei prodotti descritti nel presente documento sono supportate nel proprio ambiente specifico, fare riferimento "Tool di matrice di interoperabilità (IMT)"alla sul sito di supporto NetApp. NetApp IMT definisce i componenti e le versioni dei prodotti che possono essere utilizzati per costruire configurazioni supportate da NetApp. I risultati specifici dipendono dall'installazione di ciascun cliente in conformità alle specifiche pubblicate.

VSphere Metro Storage Cluster con ONTAP

VSphere Metro Storage Cluster con ONTAP

L'hypervisor vSphere leader del settore di VMware può essere implementato come cluster stretched indicato come vSphere Metro Storage Cluster (vMSC).

Le soluzioni vMSC sono supportate sia con NetApp® MetroCluster™ che con SnapMirror Active Sync (precedentemente noto come SnapMirror Business Continuity o SMBC) e forniscono una business continuity avanzata se uno o più domini di errore subiscono un'interruzione totale. La resilienza alle diverse modalità di errore dipende dalle opzioni di configurazione scelte.

Questa documentazione sostituisce i report tecnici precedentemente pubblicati *TR-4128: VSphere on NetApp MetroCluster*

Soluzioni di disponibilità continua per ambienti vSphere

L'architettura ONTAP è una piattaforma di storage flessibile e scalabile che fornisce servizi SAN (FCP, iSCSI e NVMe-of) e NAS (NFS v3 e v4,1) per datastore. I sistemi storage NetApp AFF, ASA e FAS utilizzano il sistema operativo ONTAP per offrire protocolli aggiuntivi per l'accesso allo storage guest, come S3 e SMB/CIFS.

NetApp MetroCluster utilizza la funzione di ha (failover del controller o CFO) di NetApp per la protezione dai guasti dei controller. Include inoltre la tecnologia SyncMirror locale, il failover cluster in caso di disastro (Cluster failover on Disaster o CFOD), la ridondanza hardware e la separazione geografica per ottenere livelli elevati di disponibilità. SyncMirror esegue il mirroring sincrono dei dati tra le due metà della configurazione MetroCluster scrivendo i dati su due plessi: Il plesso locale (sullo shelf locale) fornendo attivamente i dati e il plesso remoto (sullo shelf remoto) normalmente non fornendo i dati. La ridondanza hardware viene implementata per tutti i componenti MetroCluster, come controller, storage, cavi, switch (utilizzati con Fabric MetroCluster) e adattatori.

La sincronizzazione attiva di NetApp SnapMirror, disponibile su sistemi non MetroCluster e su sistemi ASA R2, offre una protezione granulare dei datastore con protocolli SAN FCP e iSCSI. Consente di proteggere l'intero

vMSC o in modo selettivo i carichi di lavoro ad alta priorità. Offre l'accesso Active-Active ai siti locali e remoti, a differenza di NetApp MetroCluster, che è una soluzione Active-standby. A partire da ONTAP 9.15.1, SnapMirror Active Sync supporta una funzionalità Active/Active simmetrica, consentendo operazioni i/o in lettura e scrittura da entrambe le copie di un LUN protetto con replica sincrona bidirezionale, consentendo alle due copie LUN di servire le operazioni i/O. Prima di ONTAP 9.15.1, la sincronizzazione attiva di SnapMirror supporta solo configurazioni Active/Active asimmetriche, in cui i dati sul sito secondario sono sottoposti a proxy alla copia primaria di una LUN.

Per creare un cluster VMware ha/DRS su due siti, gli host ESXi vengono utilizzati e gestiti da un'appliance vCenter Server (VCSA). Le reti di gestione vSphere, vMotion® e delle macchine virtuali sono collegate tramite una rete ridondante tra i due siti. VCenter Server che gestisce il cluster ha/DRS può connettersi agli host ESXi in entrambi i siti e deve essere configurato utilizzando vCenter ha.

Fare riferimento a. "Come creare e configurare i cluster nel client vSphere" Per configurare vCenter ha.

Fare riferimento anche a "Procedure consigliate per VMware vSphere Metro Storage Cluster".

Che cos'è vSphere Metro Storage Cluster?

VSphere Metro Storage Cluster (vMSC) è una configurazione certificata che protegge le macchine virtuali (VM) e i container dai guasti. Ciò si ottiene utilizzando concetti di storage estesi insieme ai cluster di host ESXi, distribuiti in diversi domini di errore come rack, edifici, campus o persino città. Le tecnologie di storage sincrono attive di NetApp MetroCluster e SnapMirror vengono utilizzate per fornire ai cluster host una protezione con RPO (Recovery Point Objective) pari a zero (RPO=0). La configurazione vMSC è progettata per garantire che i dati siano sempre disponibili anche in caso di errore di un "sito" fisico o logico completo. Un dispositivo di storage che fa parte della configurazione vMSC deve essere certificato dopo aver superato un processo di certificazione vMSC di successo. Tutti i dispositivi di archiviazione supportati sono disponibili nella "Guida alla compatibilità dello storage VMware" .

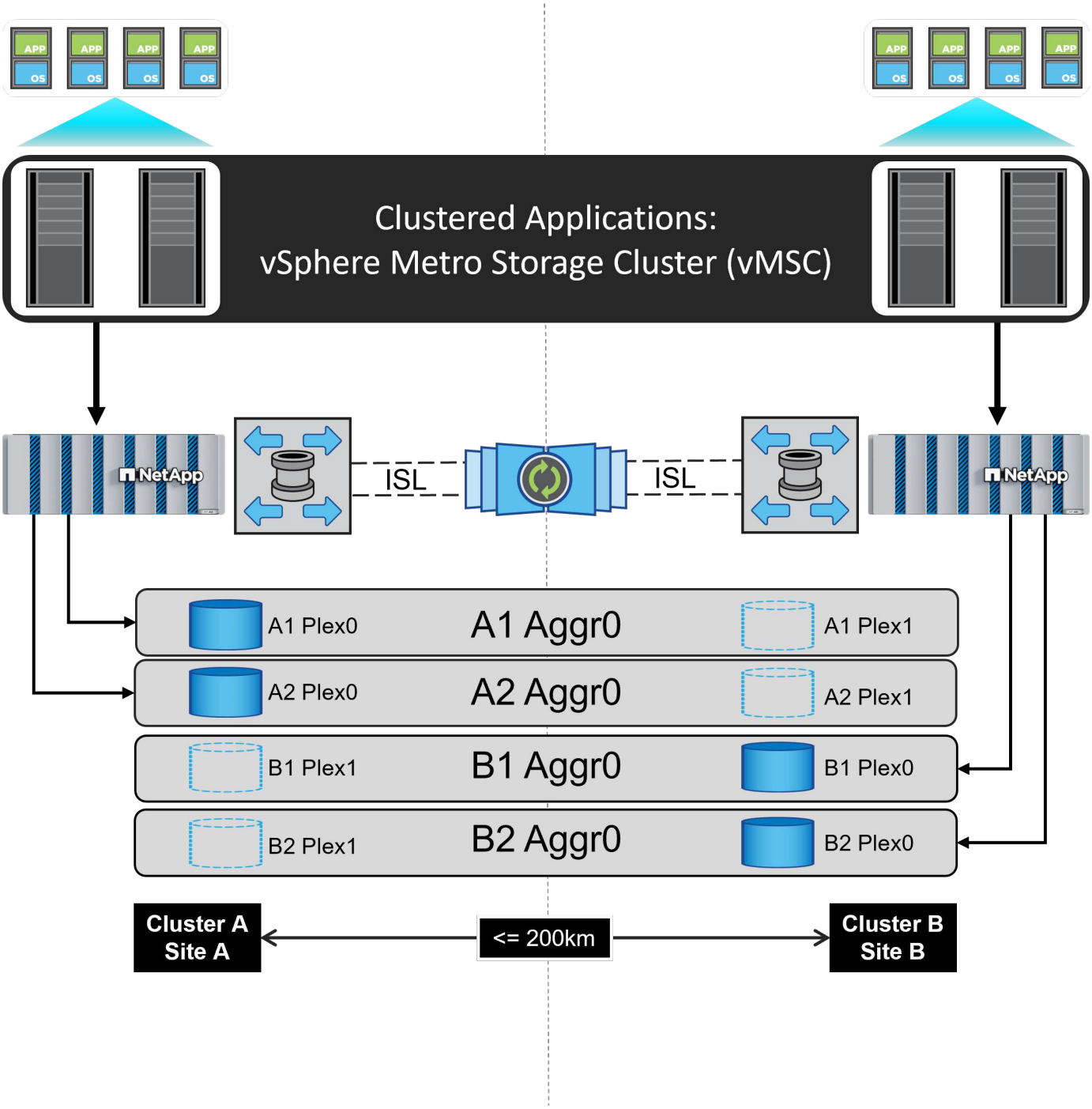
Per ulteriori informazioni sulle linee guida di progettazione per vSphere Metro Storage Cluster, consultare la seguente documentazione:

- "Supporto di VMware vSphere con NetApp MetroCluster"
- "Supporto di VMware vSphere con business continuity di NetApp SnapMirror" (Adesso noto come SnapMirror Active Sync)

NetApp MetroCluster può essere implementato in due configurazioni diverse per l'utilizzo con vSphere:

- Stretch MetroCluster
- Fabric MetroCluster

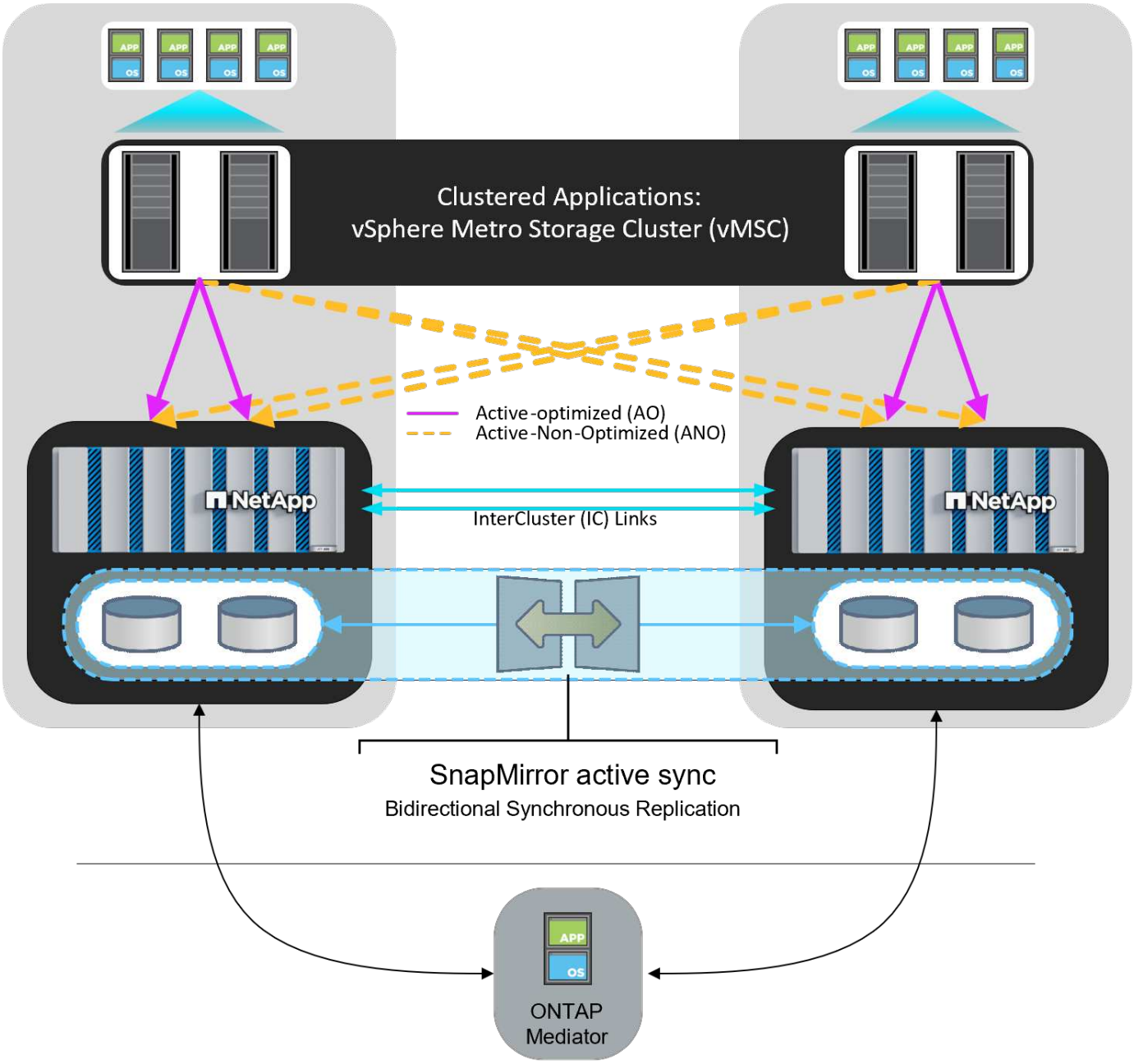
Di seguito viene illustrato uno schema topologico di alto livello di Stretch MetroCluster.



Fare riferimento a. "Documentazione MetroCluster" Per informazioni specifiche sulla progettazione e la distribuzione di MetroCluster.

SnapMirror Active Sync può anche essere implementato in due modi diversi.

- Asimmetrico
- Active Sync simmetrico (ONTAP 9.15.1)



Per informazioni specifiche sulla progettazione e la distribuzione della sincronizzazione attiva di SnapMirror, consultare la sezione "Documenti NetApp".

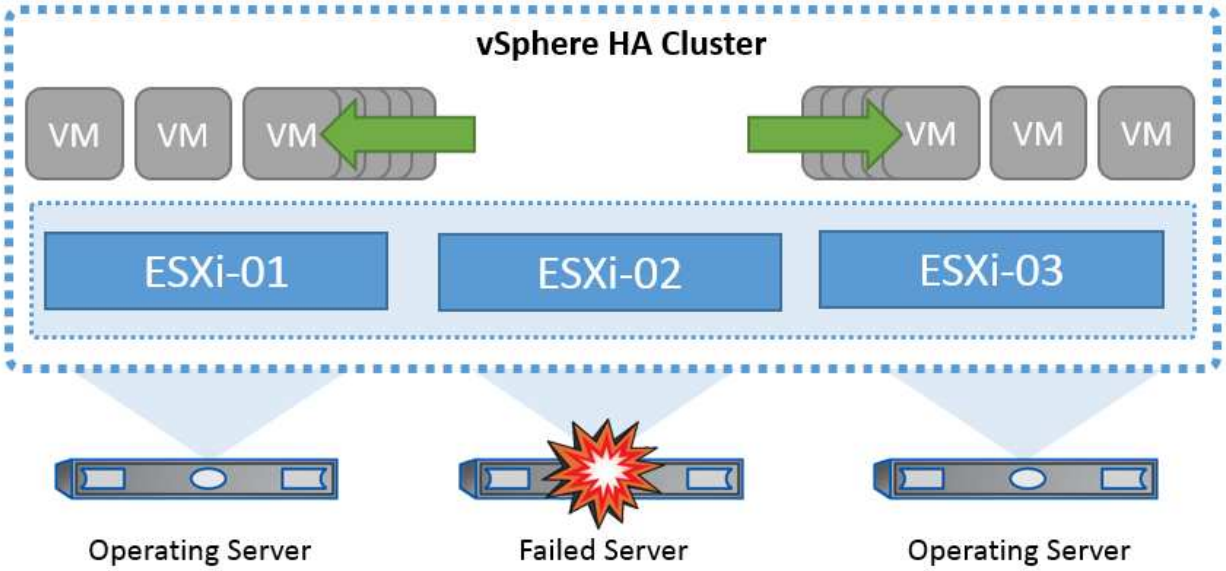
Panoramica della soluzione VMware vSphere

VCenter Server Appliance (VCSA) è un potente sistema di gestione centralizzato e un singolo pannello di controllo per vSphere che consente agli amministratori di utilizzare in modo efficace i cluster ESXi. Agevola le funzioni chiave come provisioning delle macchine virtuali, funzionamento di vMotion, alta disponibilità (ha), Distributed Resource Scheduler (DRS), Tanzu Kubernetes Grid e altro ancora. Si tratta di un componente essenziale negli ambienti cloud VMware e deve essere progettato tenendo presente la disponibilità del servizio.

Alta disponibilità vSphere

La tecnologia cluster di VMware raggruppa i server ESXi in pool di risorse condivise per le macchine virtuali e fornisce vSphere High Availability (ha), offrendo un'elevata disponibilità e una semplicità d'uso per le applicazioni eseguite su macchine virtuali. Quando la funzionalità ha è abilitata sul cluster, ogni server ESXi mantiene la comunicazione con altri host in modo che, se un host ESXi non risponde o si isola, il cluster di ha può negoziare il recovery delle macchine virtuali in esecuzione sull'host ESXi tra gli host sopravvissuti nel cluster. In caso di errore del sistema operativo guest, vSphere ha può riavviare la macchina virtuale interessata sullo stesso server fisico. VSphere ha consente di ridurre i tempi di inattività pianificati, prevenire i tempi di inattività non pianificati e ripristinare rapidamente i dati in caso di interruzioni.

Cluster vSphere ha in grado di ripristinare le VM dal server guasto.



È importante comprendere che VMware vSphere non conosce la sincronizzazione attiva di NetApp MetroCluster o SnapMirror e vede tutti gli host ESXi nel cluster vSphere come host idonei per le operazioni del cluster ha in base alle configurazioni di affinità dei gruppi host e VM.

Rilevamento errori host

Non appena viene creato il cluster ha, tutti gli host nel cluster partecipano alle elezioni e uno degli host diventa un master. Ogni slave esegue heartbeat di rete al master, e il master a sua volta esegue heartbeat di rete su tutti gli host slave. L'host master di un cluster vSphere ha è responsabile del rilevamento del guasto degli host slave.

A seconda del tipo di errore rilevato, potrebbe essere necessario eseguire il failover delle macchine virtuali in esecuzione sugli host.

In un cluster vSphere ha, vengono rilevati tre tipi di errore dell'host:

- Errore - Un host smette di funzionare.
- Isolamento - Un host diventa isolato dalla rete.
- Partizione - Un host perde la connettività di rete con l'host master.

L'host master monitora gli host slave nel cluster. Questa comunicazione viene fatta attraverso lo scambio di heartbeat di rete ogni secondo. Quando l'host master smette di ricevere questi heartbeat da un host slave, controlla la liveness dell'host prima di dichiarare che l'host non è riuscito. Il controllo liveness che l'ospite

principale effettua è di determinare se l'ospite secondario sta scambiando i heartbeat con uno dei datastore. Inoltre, l'host master verifica se l'host risponde ai ping ICMP inviati ai propri indirizzi IP di gestione per rilevare se è semplicemente isolato dal suo nodo master o completamente isolato dalla rete. Per farlo, eseguire il ping del gateway predefinito. È possibile specificare manualmente uno o più indirizzi di isolamento per migliorare l'affidabilità della convalida dell'isolamento.

NetApp consiglia di specificare un minimo di due indirizzi di isolamento aggiuntivi e che ciascuno di questi indirizzi sia locale al sito. Ciò migliorerà l'affidabilità della convalida dell'isolamento.

Risposta di isolamento dell'host

Risposta di isolamento è un'impostazione in vSphere ha che determina l'azione attivata sulle macchine virtuali quando un host in un cluster vSphere ha perde le connessioni di rete di gestione ma continua a essere eseguito. Sono disponibili tre opzioni per questa impostazione: "Disabilitato", "Arresta e riavvia le macchine virtuali" e "Spegni e riavvia le macchine virtuali".

"Spegni" è meglio di "Spegni", che non elimina le modifiche più recenti al disco o alle transazioni di commit. Se le macchine virtuali non si sono arrestate entro 300 secondi, vengono spente. Per modificare il tempo di attesa, utilizzare l'opzione avanzata das.isolationshutdowntimeout.

Prima che ha avvii la risposta di isolamento, verifica prima se l'agente master ha vSphere è proprietario del datastore che contiene i file di configurazione della VM. In caso contrario, l'host non attiverà la risposta di isolamento, poiché non vi è alcun master per riavviare le VM. L'host controllerà periodicamente lo stato del datastore per determinare se viene richiesto da un agente vSphere ha che detiene il ruolo master.

NetApp consiglia di impostare la risposta di isolamento dell'host su Disabilitato.

Una condizione split-brain può verificarsi se un host viene isolato o partizionato dall'host master vSphere ha e il master non è in grado di comunicare tramite datastore heartbeat o tramite ping. Il master dichiara l'host isolato inattivo e riavvia le macchine virtuali su altri host nel cluster. Esiste ora una condizione split-brain perché esistono due istanze della macchina virtuale in esecuzione, una sola delle quali è in grado di leggere o scrivere i dischi virtuali. Le condizioni split-brain possono ora essere evitate configurando VMCP (VM Component Protection).

Protezione dei componenti VM (VMCP)

Uno dei miglioramenti delle funzionalità di vSphere 6, relativi all'ha, è VMCP. VMCP fornisce una protezione avanzata da APD (All Path Down) e PDL (Permanent Device Loss) per lo storage a blocchi (FC, iSCSI, FCoE) e a file (NFS).

Perdita permanente del dispositivo (PDL)

PDL è una condizione che si verifica quando un dispositivo di memorizzazione si guasta in modo permanente o viene rimosso amministrativamente e non deve essere restituito. L'array di storage NetApp invia un codice di rilevamento SCSI a ESXi dichiarando che il dispositivo è perso in modo permanente. Nella sezione Condizioni di guasto e Risposta VM di vSphere ha, è possibile configurare la risposta che deve essere dopo il rilevamento di una condizione PDL.

NetApp consiglia di impostare "Risposta per datastore con PDL" su "**Spegni e riavvia VM**". Quando viene rilevata questa condizione, una VM viene riavviata istantaneamente su un host integro all'interno del cluster vSphere ha.

Tutti i percorsi verso il basso (APD)

APD è una condizione che si verifica quando un dispositivo di archiviazione diventa inaccessibile all'host e non sono disponibili percorsi all'array. ESXi considera questo un problema temporaneo con il dispositivo e si aspetta che diventi nuovamente disponibile.

Quando viene rilevata una condizione APD, viene avviato un timer. Dopo 140 secondi, la condizione APD viene dichiarata ufficialmente e il dispositivo viene contrassegnato come timeout APD. Una volta trascorsi i 140 secondi, ha inizia il conteggio dei minuti specificati nell'APD Delay for VM failover. Una volta trascorso il tempo specificato, ha riavvia le macchine virtuali interessate. È possibile configurare VMCP in modo che risponda in modo diverso, se lo si desidera (Disattivato, Eventi problema o Spegni e riavvia le macchine virtuali).

- NetApp consiglia di configurare "Risposta per datastore con APD" su "**Spegni e riavvia le VM (conservative)**".
- Conservative si riferisce alla probabilità che ha sia in grado di riavviare le VM. Quando è impostata su Conservative, ha riavvia la VM interessata dall'APD solo se sa che un altro host può riavviarla. In caso di problemi aggressivi, ha tenterà di riavviare la macchina virtuale anche se non conosce lo stato degli altri host. Ciò può comportare il mancato riavvio delle VM se non vi è alcun host con accesso al datastore su cui si trova.
- Se lo stato APD viene risolto e l'accesso allo storage viene ripristinato prima che sia trascorso il timeout, ha non riavvierà inutilmente la macchina virtuale a meno che non sia stata esplicitamente configurata. Se si desidera una risposta anche quando l'ambiente è stato ripristinato dalla condizione APD, è necessario configurare la risposta per il ripristino APD dopo il timeout APD in modo da ripristinare le VM.
- NetApp consiglia di configurare la risposta per il ripristino APD dopo il timeout APD su Disabilitato.

Implementazione di VMware DRS per NetApp SnapMirror Active Sync

VMware DRS è una funzionalità che aggrega le risorse host in un cluster e viene utilizzata principalmente per il bilanciamento del carico all'interno di un cluster in un'infrastruttura virtuale. VMware DRS calcola principalmente le risorse di CPU e memoria per eseguire il bilanciamento del carico in un cluster. Poiché vSphere non è consapevole del clustering allungato, considera tutti gli host in entrambi i siti durante il bilanciamento del carico.

Implementazione VMware DRS per NetApp MetroCluster

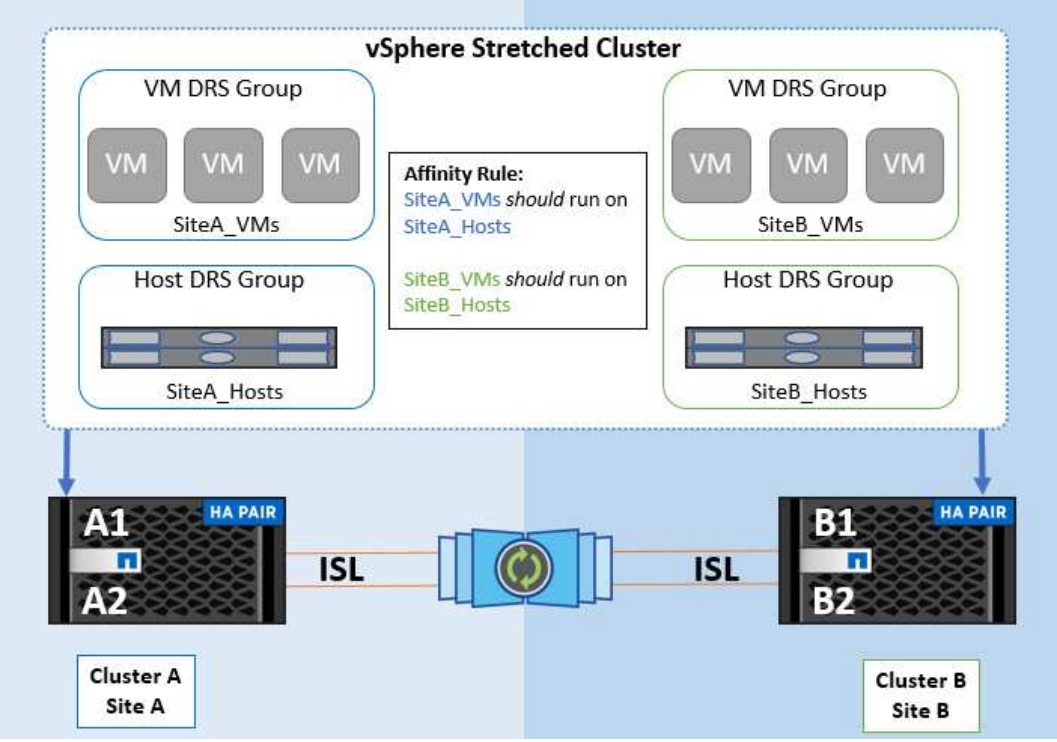
```
 To avoid cross-site traffic, NetApp recommends configuring DRS affinity
rules to manage a logical separation of VMs. This will ensure that unless
there is a complete site failure, HA and DRS will only use local hosts.
Se si crea una regola di affinità DRS per il cluster, è possibile
specificare in che modo vSphere applica tale regola durante il failover di
una macchina virtuale.
```

Esistono due tipi di regole che è possibile specificare il comportamento di failover di vSphere ha:

- Le regole di anti-affinità delle macchine virtuali costringono le macchine virtuali specificate a rimanere separate durante le azioni di failover.
- Le regole di affinità degli host VM collocano macchine virtuali specifiche su un host specifico o su un membro di un gruppo definito di host durante le azioni di failover.

Utilizzando le regole di affinità degli host delle macchine virtuali in VMware DRS, si può avere una separazione logica tra il sito A e il sito B in modo che la macchina virtuale venga eseguita sull'host nello stesso sito dell'array configurato come controller di lettura/scrittura principale per un determinato datastore. Inoltre, le regole di affinità degli host delle macchine virtuali consentono alle macchine virtuali di rimanere locali rispetto allo storage, il che a sua volta determina la connessione della macchina virtuale in caso di errori di rete tra i siti.

Di seguito è riportato un esempio di gruppi di host VM e regole di affinità.



Best practice

NetApp consiglia di implementare regole "should" invece di regole "must", in quanto vengono violate da vSphere ha in caso di errore. L'utilizzo di regole "must" potrebbe potenzialmente causare interruzioni del servizio.

La disponibilità dei servizi dovrebbe sempre prevalere sulle prestazioni. Nello scenario in cui un data center completo si guasta, le regole "must" devono scegliere gli host dal gruppo di affinità host VM e, quando il data center non è disponibile, le macchine virtuali non si riavviano.

Implementazione di VMware Storage DRS con NetApp MetroCluster

La funzionalità VMware Storage DRS consente l'aggregazione di datastore in una singola unità e bilancia i dischi della macchina virtuale quando vengono superate le soglie di controllo i/o di storage (SIOC).

Il controllo i/o dello storage è abilitato per impostazione predefinita sui cluster DRS abilitati per Storage DRS. Il controllo i/o dello storage consente a un amministratore di controllare la quantità di i/o dello storage allocata alle macchine virtuali nei periodi di congestione dell'i/o e di conseguenza le macchine virtuali più importanti possono preferire le macchine virtuali meno importanti per l'allocazione delle risorse i/O.

Storage DRS utilizza Storage vMotion per migrare le macchine virtuali in datastore diversi all'interno di un cluster di datastore. In un ambiente NetApp MetroCluster, la migrazione di una macchina virtuale deve essere controllata all'interno dei datastore di quel sito. Ad esempio, la macchina virtuale A, in esecuzione su un host

nel sito A, dovrebbe idealmente migrare all'interno dei datastore della SVM nel sito A. In caso contrario, la macchina virtuale continuerà a funzionare ma con prestazioni ridotte, poiché la lettura/scrittura del disco virtuale avverrà dal sito B attraverso collegamenti tra siti.

*Quando si utilizza l'archiviazione ONTAP, si consiglia di disattivare l'archiviazione DRS.

- I DRS di archiviazione non sono generalmente necessari o consigliati per l'uso con i sistemi di archiviazione ONTAP.
- ONTAP offre proprie funzionalità di efficienza dello storage, come deduplica, compressione e compaction, che possono essere influenzate dallo Storage DRS.
- Se si utilizzano snapshot ONTAP, storage vMotion lascerebbe la copia della macchina virtuale nella snapshot, aumentando potenzialmente l'utilizzo dello storage e potrebbe avere un impatto sulle applicazioni di backup, come NetApp SnapCenter, che tengono traccia delle macchine virtuali e delle relative snapshot ONTAP.

Linee guida per la progettazione e l'implementazione di vMSC

Questo documento delinea le linee guida di progettazione e implementazione per vMSC con i sistemi di storage ONTAP.

Configurazione dello storage NetApp

Le istruzioni per l'installazione di NetApp MetroCluster sono disponibili all'indirizzo "Documentazione MetroCluster". Le istruzioni per SnapMirror Active Sync (SMA) sono disponibili anche all'indirizzo "Panoramica di SnapMirror Business Continuity".

Una volta configurato MetroCluster, gestirlo è come gestire un ambiente ONTAP tradizionale. Puoi configurare Storage Virtual Machine (SVM) utilizzando vari strumenti come l'interfaccia a riga di comando (CLI), System Manager o Ansible. Una volta configurate le SVM, occorre creare nel cluster interfacce logiche (LIF), volumi e LUN (Logical Unit Number) da utilizzare per le normali operazioni. Questi oggetti verranno replicati automaticamente sull'altro cluster utilizzando la rete di peering del cluster.

Se non utilizzi MetroCluster, o se disponi di sistemi ONTAP non supportati per MetroCluster, come ad esempio i sistemi ASA R2, puoi utilizzare SnapMirror Active Sync che fornisce una protezione granulare dei datastore e l'accesso Active-Active su più cluster ONTAP in diversi domini di errore. SMA utilizza gruppi di coerenza (CGS) per garantire la coerenza dell'ordine di scrittura tra uno o più datastore ed è possibile creare più CGS in base ai requisiti dell'applicazione e del datastore. I gruppi di coerenza sono particolarmente utili per le applicazioni che richiedono la sincronizzazione dei dati tra datastore multipli. Ad esempio, LVM guest distribuiti tra datastore. SMA supporta inoltre RDM (Raw Device Mapping) e storage connesso al guest con iniziatori iSCSI in-guest. Per ulteriori informazioni sui gruppi di coerenza, visitare il sitoWeb all'indirizzo "Panoramica dei gruppi di coerenza".

Esiste una certa differenza nella gestione di una configurazione vMSC con sincronizzazione attiva SnapMirror rispetto a una MetroCluster. In primo luogo, SMA è una configurazione solo SAN, nessun datastore NFS può essere protetto con la sincronizzazione attiva di SnapMirror. In secondo luogo, è necessario mappare entrambe le copie delle LUN agli host ESXi per accedere ai datastore replicati in entrambi i domini di errore. Terzo, devi creare uno o più gruppi di coerenza per i datastore da proteggere con la sincronizzazione attiva di SnapMirror. Infine, è necessario creare un criterio SnapMirror per i gruppi di coerenza creati. Tutto questo può essere fatto facilmente usando la procedura guidata "Protect cluster" nel plug-in vCenter di ONTAP tools, o usando manualmente la CLI di ONTAP o System Manager.

Utilizzo del plug-in vCenter di ONTAP Tools per SnapMirror Active Sync

Il plug-in vCenter degli strumenti ONTAP fornisce un modo semplice e intuitivo per configurare SnapMirror Active Sync per vMSC. Puoi usare il plug-in vCenter di ONTAP Tools per creare e gestire relazioni di sincronizzazione attive di SnapMirror tra due cluster ONTAP. Questo plugin fornisce un'interfaccia di facile utilizzo per stabilire e gestire queste relazioni in modo efficiente. Per ulteriori informazioni sul plug-in vCenter degli strumenti ONTAP, visitare il sito Web all'indirizzo "Strumenti ONTAP per VMware vSphere", oppure accedere direttamente a "Proteggere utilizzando la protezione del cluster host".

Configurazione di VMware vSphere

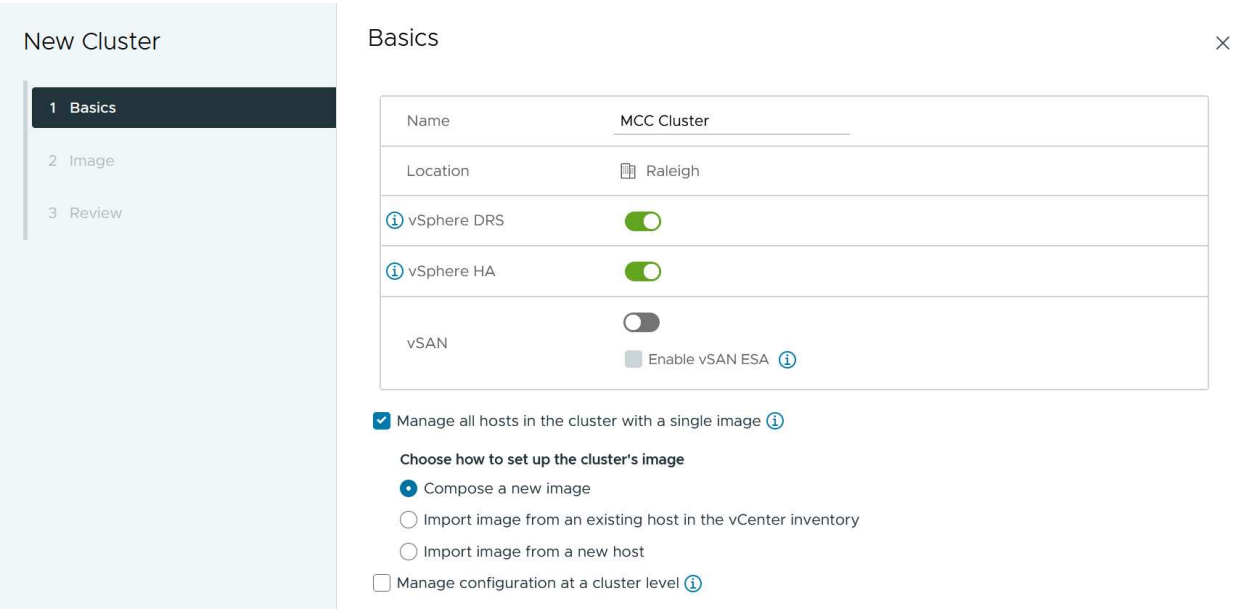
Creare un cluster vSphere ha

La creazione di un cluster vSphere ha è un processo in più fasi documentato all'indirizzo "Come creare e configurare i cluster nel client vSphere su docs.vmware.com". In poche parole, devi prima creare un cluster vuoto, quindi, utilizzando vCenter, devi aggiungere host e specificare l'ha vSphere del cluster e le altre impostazioni.

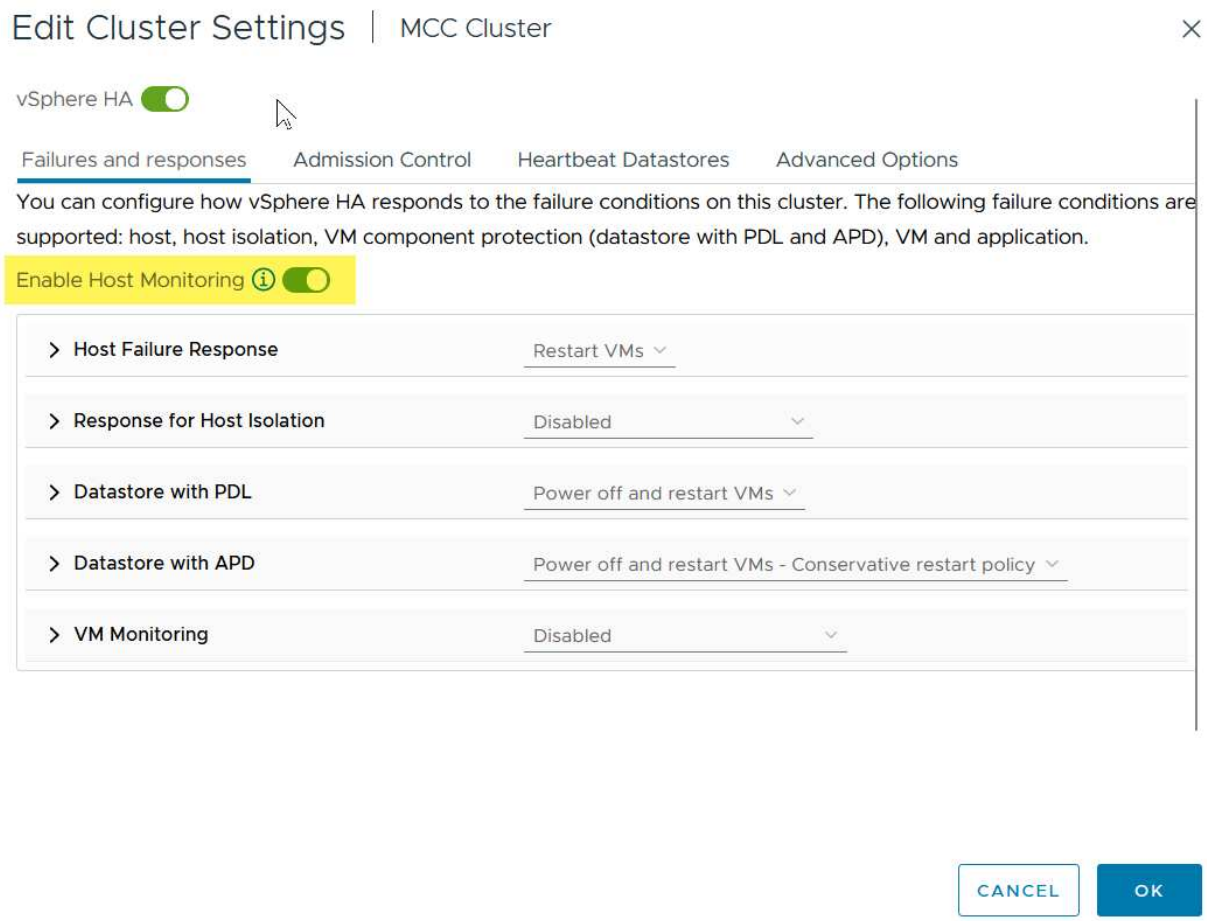
> Nessuna disposizione del presente documento sostituisce "Procedure consigliate per VMware vSphere Metro Storage Cluster". Questo contenuto viene fornito a scopo di riferimento e non sostituisce la documentazione ufficiale VMware.

Per configurare un cluster ha, completare i seguenti passaggi:

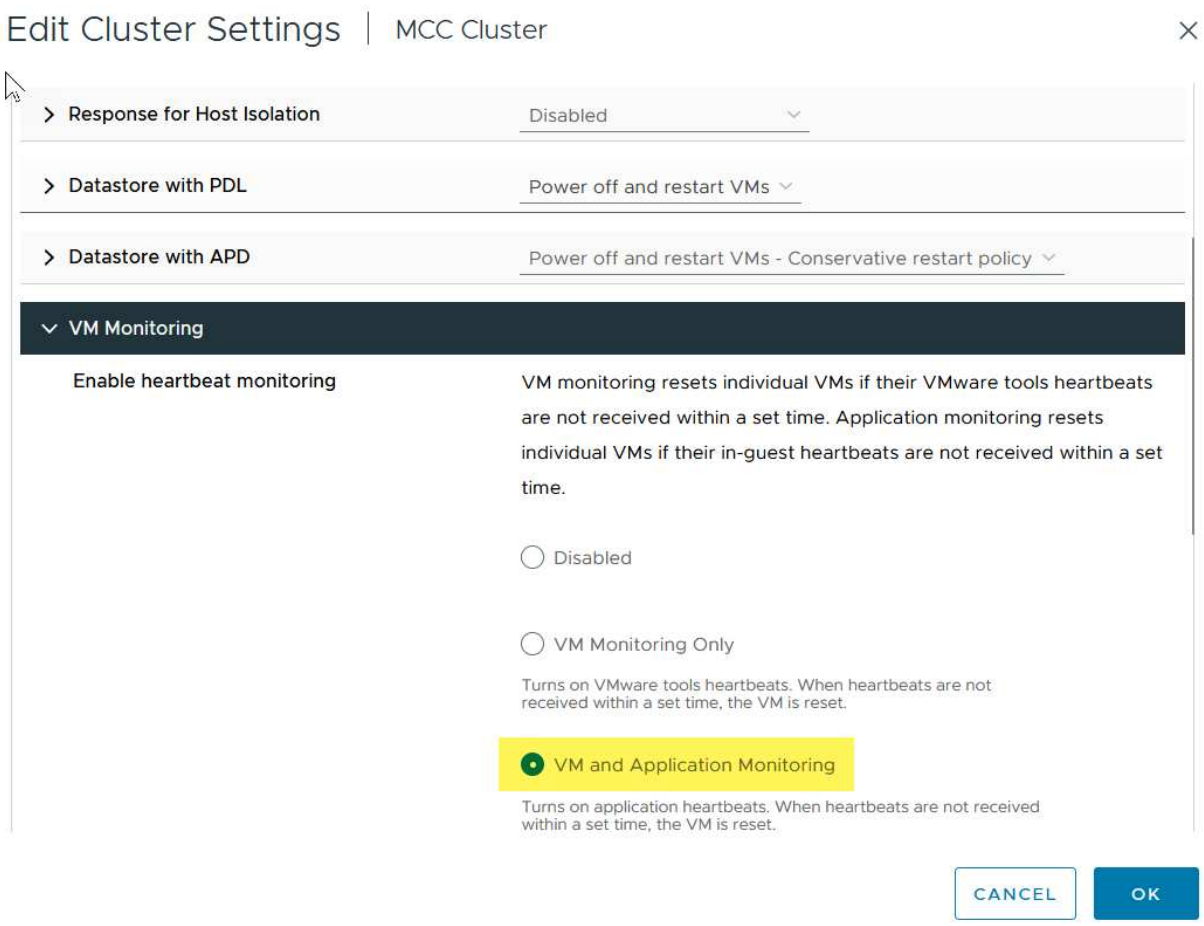
1. Connettersi all'interfaccia utente di vCenter.
2. In host e cluster, individuare il data center in cui si desidera creare il cluster ha.
3. Fare clic con il pulsante destro del mouse sull'oggetto del data center e selezionare nuovo cluster. In base alle nozioni di base, assicurarsi di aver abilitato vSphere DRS e vSphere ha. Completare la procedura guidata.



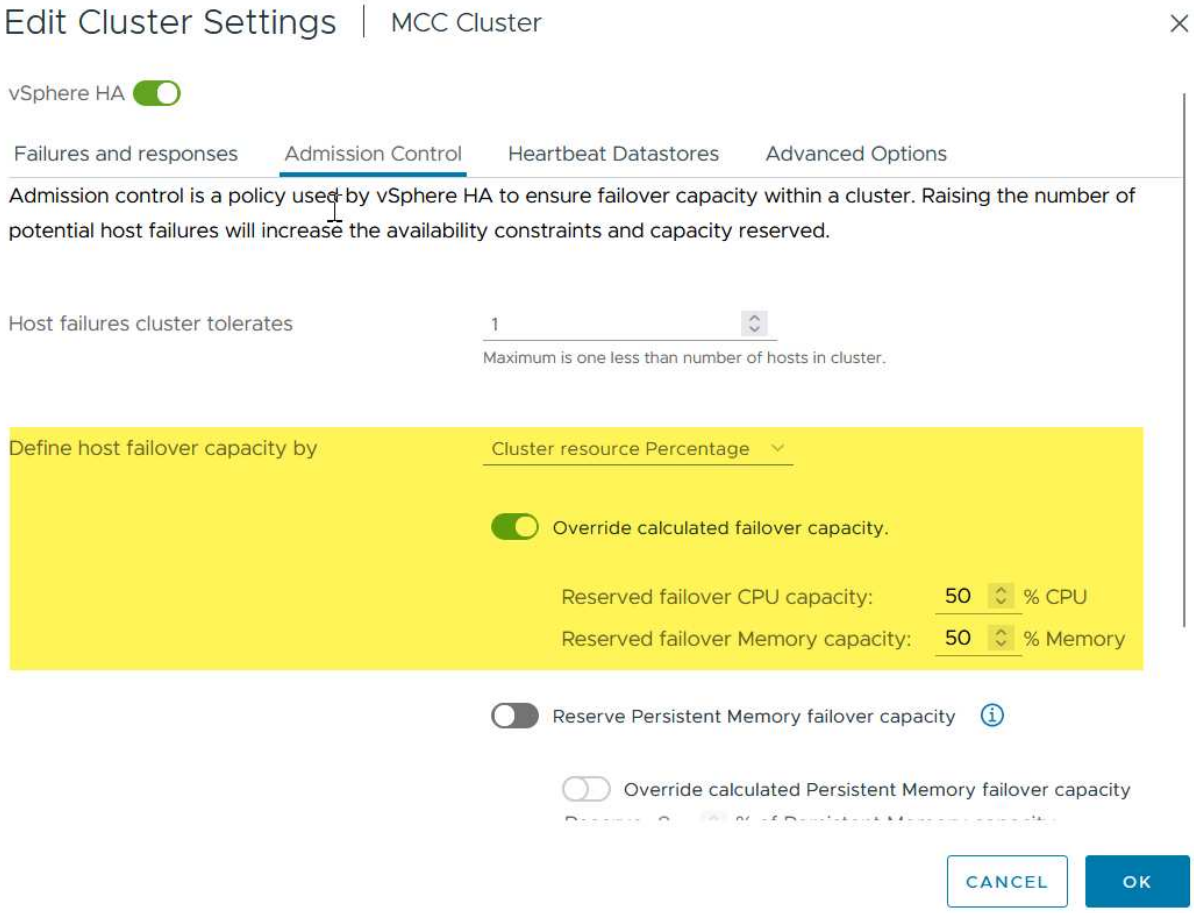
1. Selezionare il cluster e accedere alla scheda di configurazione. Selezionare vSphere ha e fare clic su Modifica.
2. In monitoraggio host, selezionare l'opzione attiva monitoraggio host.



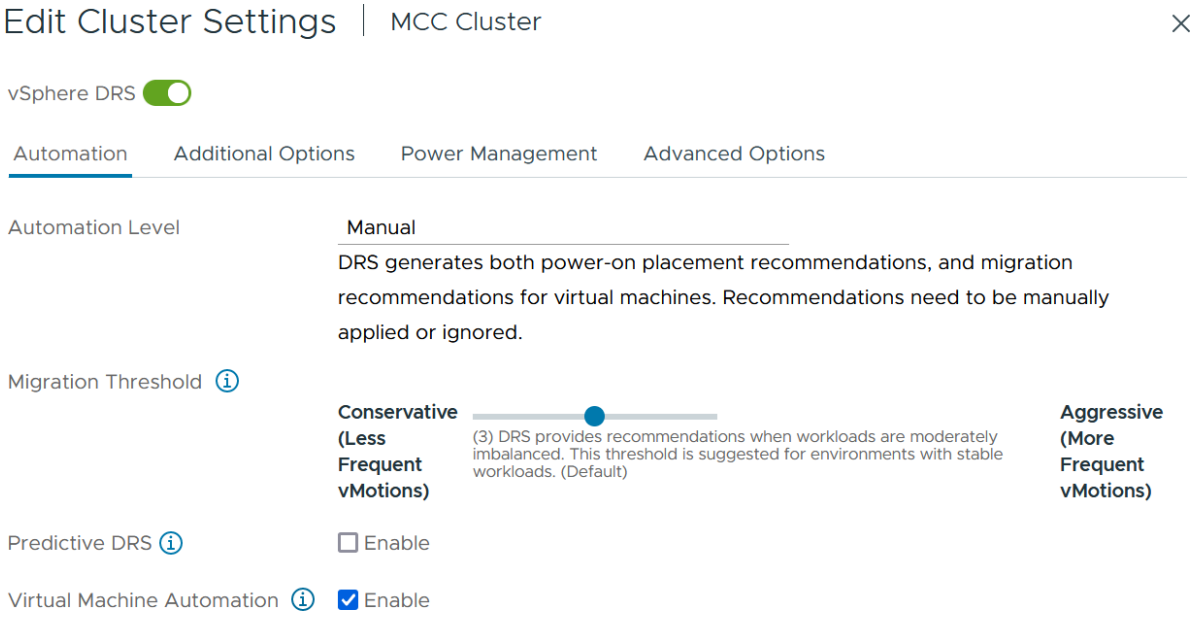
1. Nella scheda guasti e risposte, in monitoraggio VM, selezionare l'opzione solo monitoraggio VM o monitoraggio VM e applicazione.



1. In controllo ammissione, impostare l'opzione di controllo ammissione ha su Cluster Resource Reserve; utilizzare 50% CPU/MEM.



1. Fare clic su "OK".
2. Selezionare DRS e fare clic su MODIFICA.
3. Impostare il livello di automazione su manuale, a meno che non sia richiesto dalle applicazioni.



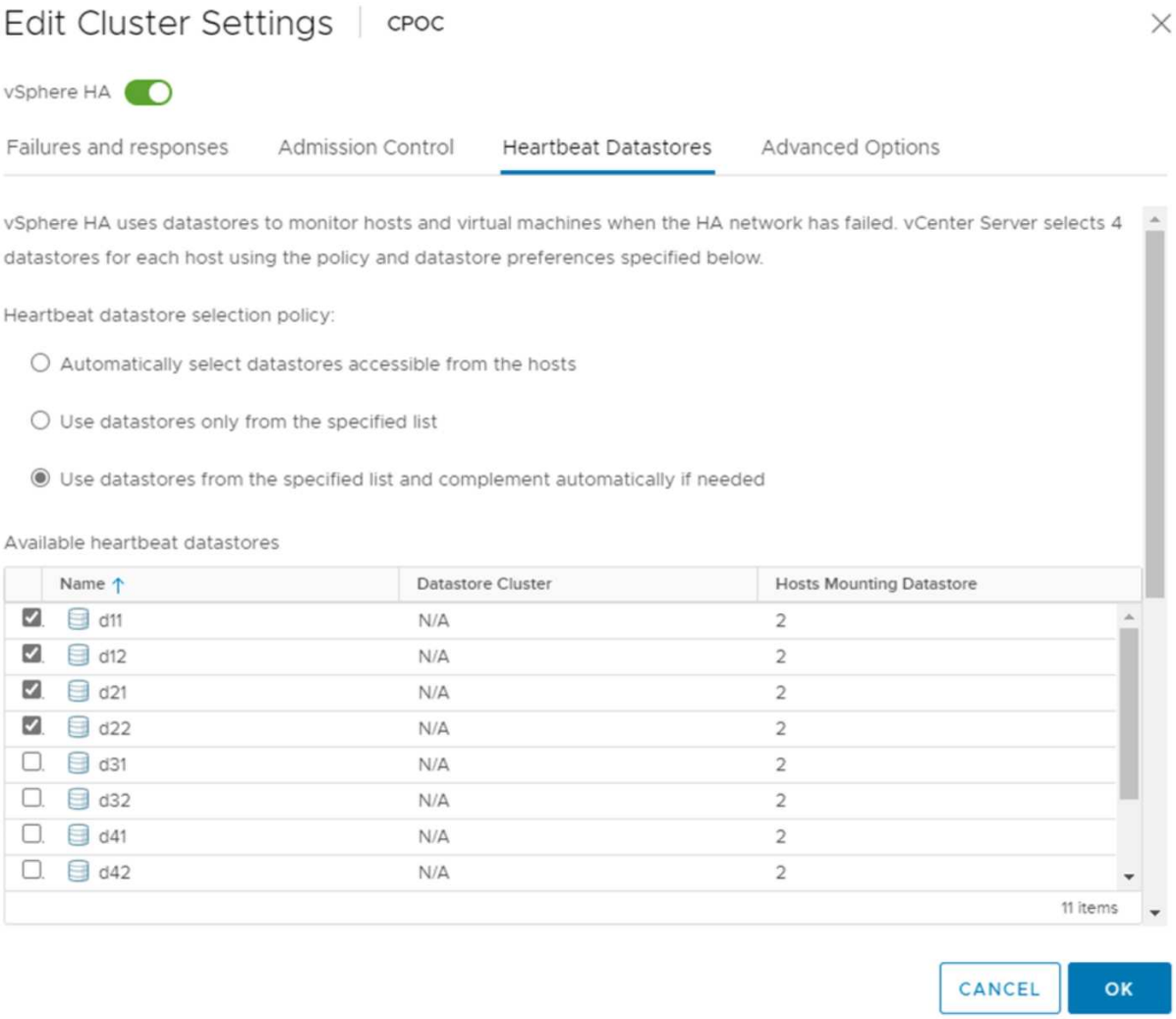
1. Abilitare la protezione dei componenti VM, fare riferimento a. "docs.vmware.com".
2. Le seguenti impostazioni aggiuntive di vSphere ha sono consigliate per vMSC con MetroCluster:

Guasto	Risposta
Errore host	Riavviare le VM
Isolamento degli host	Disattivato
Datastore con perdita permanente di dispositivi (PDL)	Spegnere e riavviare le macchine virtuali
Datastore con tutti i percorsi verso il basso (APD)	Spegnere e riavviare le macchine virtuali
L'ospite non batte il cuore	Ripristinare le VM
Policy di riavvio della VM	Determinato dall'importanza della VM
Risposta per l'isolamento dell'host	Arrestare e riavviare le VM
Risposta per il datastore con PDL	Spegnere e riavviare le macchine virtuali
Risposta per datastore con APD	Spegnere e riavviare le macchine virtuali (conservative)
Ritardo del failover delle macchine virtuali per APD	3 minuti
Risposta per il ripristino APD con timeout APD	Disattivato
Sensibilità di monitoraggio VM	Preimpostazione alta

Configurare gli archivi dati per Heartbeating

VSphere ha utilizza i datastore per monitorare gli host e le macchine virtuali in caso di guasto alla rete di gestione. È possibile configurare in che modo vCenter seleziona i datastore heartbeat. Per configurare gli archivi dati per il heartbeat, completare i seguenti passaggi:

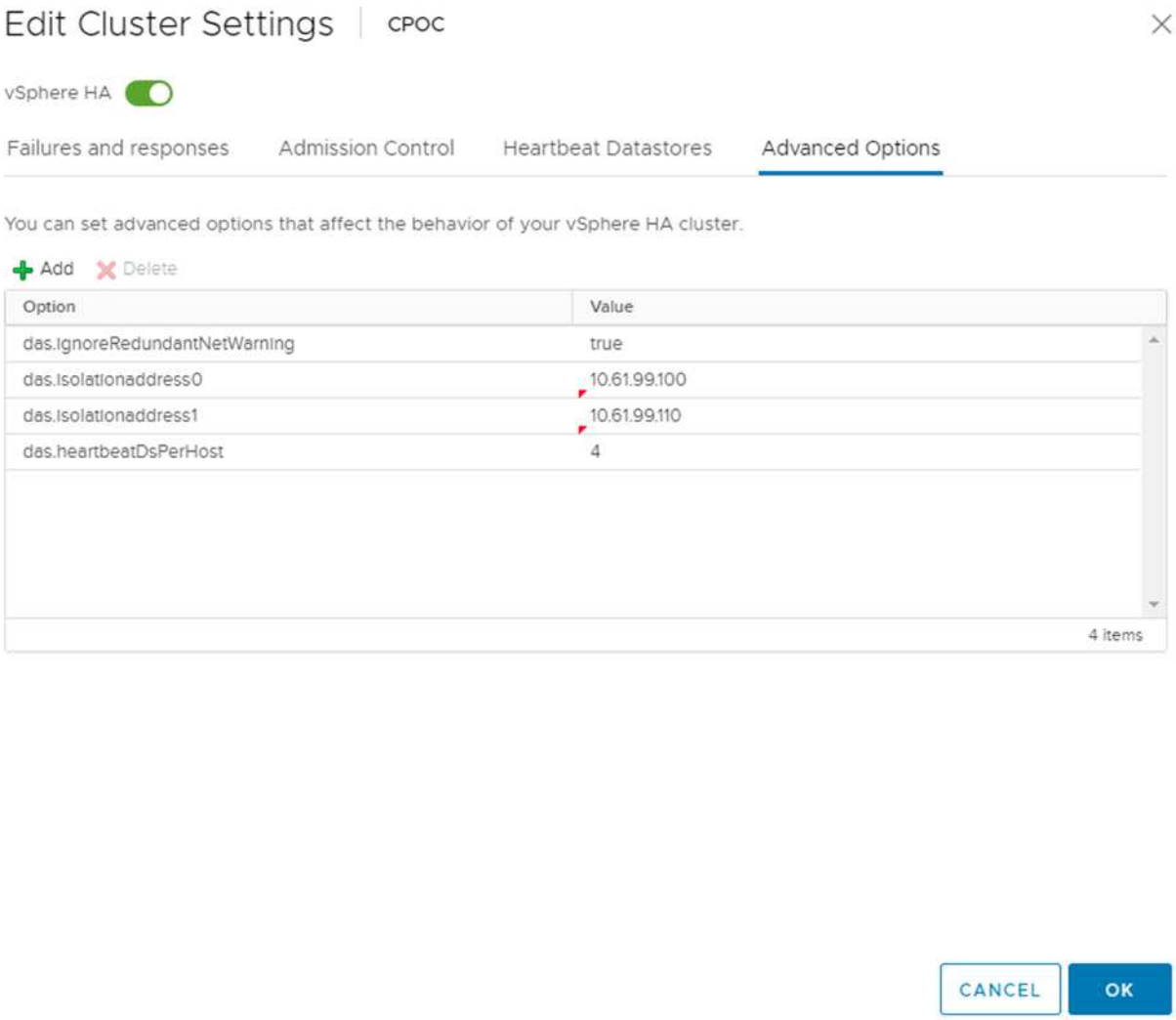
1. Nella sezione Heartbeating del datastore, selezionare Usa archivi dati dall'elenco specificato e completare automaticamente se necessario.
2. Seleziona i datastore che desideri utilizzare vCenter da entrambi i siti e premi OK.



Configurare le opzioni avanzate

Gli eventi di isolamento si verificano quando gli host all'interno di un cluster ha perdono la connettività alla rete o ad altri host nel cluster. Per impostazione predefinita, vSphere ha utilizzerà il gateway predefinito per la propria rete di gestione come indirizzo di isolamento predefinito. Tuttavia, è possibile specificare indirizzi di isolamento aggiuntivi per l'host al ping per determinare se deve essere attivata una risposta di isolamento. Aggiungere due IP di isolamento in grado di eseguire il ping, uno per sito. Non utilizzare l'indirizzo IP del gateway. L'impostazione avanzata vSphere ha utilizzata è das.isolationaddress. A tale scopo, è possibile utilizzare gli indirizzi IP ONTAP o Mediator.

Fare riferimento a "Procedure consigliate per VMware vSphere Metro Storage Cluster" per ulteriori informazioni.



L'aggiunta di un'impostazione avanzata denominata das.heartbeatDsPerHost può aumentare il numero di datastore heartbeat. Utilizzare quattro datastore heartbeat (HB DSS), due per sito. Utilizzare l'opzione "Seleziona dall'elenco ma complimento". Questo è necessario perché se un sito non funziona, è necessario ancora due HB DSS. Tuttavia, questi elementi non devono essere protetti con la sincronizzazione attiva di MetroCluster o SnapMirror.

Fare riferimento a "Procedure consigliate per VMware vSphere Metro Storage Cluster" per ulteriori informazioni.

Affinità con VMware DRS per NetApp MetroCluster

In questa sezione vengono creati gruppi DRS per VM e host per ciascun sito/cluster nell'ambiente MetroCluster. Quindi configuriamo le regole VM\host per allineare l'affinità dell'host VM con le risorse di storage locali. Ad esempio, il sito A fa parte del gruppo VM sitea_VM e gli host del sito A appartengono al gruppo host sitea_hosts. Successivamente, in VM\host Rules, si afferma che sitea_vm deve essere eseguito sugli host in sitea_hosts.

- NetApp consiglia vivamente la specifica **deve essere eseguita sugli host nel gruppo** piuttosto che sulla specifica **deve essere eseguita sugli host nel gruppo**. In caso di guasto dell'host del sito A, è necessario riavviare le macchine virtuali del sito A sugli host del sito B attraverso vSphere ha, ma quest'ultima specifica non consente all'ha di riavviare le macchine virtuali sul sito B perché è una regola rigida. La specifica precedente è una regola debole e viene violata in caso di ha, abilitando in tal modo la disponibilità anziché le prestazioni.
- È possibile creare un allarme basato su eventi che viene attivato quando una macchina virtuale viola una regola di affinità VM-host. Nel client vSphere, aggiungere un nuovo allarme per la macchina virtuale e selezionare "VM is Violating VM-host Affinity Rule" (VM viola la regola di affinità VM-host) come trigger dell'evento. Per ulteriori informazioni sulla creazione e la modifica degli allarmi, consultare "Monitoraggio e performance di vSphere"la documentazione.

Creare gruppi host DRS

Per creare gruppi di host DRS specifici per il sito A e il sito B, attenersi alla seguente procedura:

1. Nel client web vSphere, fare clic con il pulsante destro del mouse sul cluster nell'inventario e selezionare Impostazioni.
2. Fare clic su VM\host Groups.
3. Fare clic su Aggiungi.
4. Digitare il nome del gruppo (ad esempio, sitea_hosts).
5. Dal menu tipo, selezionare Gruppo host.
6. Fare clic su Aggiungi e selezionare gli host desiderati dal sito A, quindi fare clic su OK.
7. Ripetere questi passaggi per aggiungere un altro gruppo di host per il sito B.
8. Fare clic su OK.

Creare gruppi DRS VM

Per creare gruppi di macchine virtuali DRS specifici per il sito A e il sito B, attenersi alla seguente procedura:

1. Nel client web vSphere, fare clic con il pulsante destro del mouse sul cluster nell'inventario e selezionare Impostazioni.
2. Fare clic su VM\host Groups.
3. Fare clic su Aggiungi.
4. Digitare il nome del gruppo (ad esempio, sitea_vm).
5. Dal menu tipo, selezionare Gruppo VM.
6. Fare clic su Add (Aggiungi) e selezionare le VM desiderate dal sito A, quindi fare clic su OK.
7. Ripetere questi passaggi per aggiungere un altro gruppo di host per il sito B.
8. Fare clic su OK.

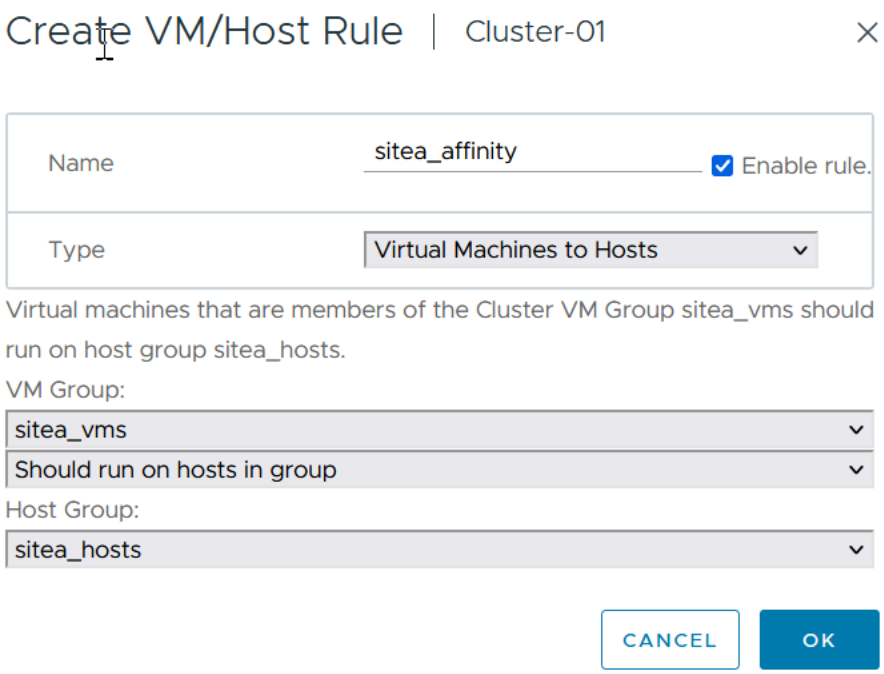
Crea regole host VM

Per creare regole di affinità DRS specifiche per il sito A e il sito B, completare i seguenti passaggi:

1. Nel client web vSphere, fare clic con il pulsante destro del mouse sul cluster nell'inventario e selezionare

Impostazioni.

2. Fare clic su VM\host Rules.
3. Fare clic su Aggiungi.
4. Digitare il nome della regola (ad esempio, sitea_Affinity).
5. Verificare che l'opzione Enable Rule (attiva regola) sia selezionata.
6. Dal menu Type (tipo), selezionare Virtual Machines to hosts (macchine virtuali a host).
7. Selezionare il gruppo VM (ad esempio, sitea_vm).
8. Selezionare il gruppo host (ad esempio, sitea_hosts).
9. Ripetere questi passaggi per aggiungere un'altra VM\regola host per il sito B.
10. Fare clic su OK.



Crea cluster di datastore se necessario

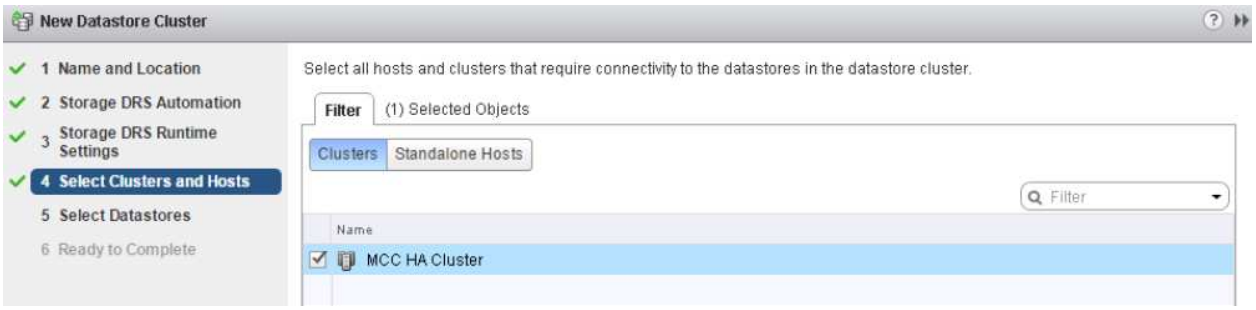
Per configurare un cluster di datastore per ciascun sito, attenersi alla seguente procedura:

1. Utilizzando il client web vSphere, individuare il data center in cui risiede il cluster ha in Storage.
2. Fare clic con il pulsante destro del mouse sull'oggetto del data center e selezionare Storage > New Datastore Cluster.

*Quando si utilizza l'archiviazione ONTAP, si consiglia di disattivare l'archiviazione DRS.

- I DRS di archiviazione non sono generalmente necessari o consigliati per l'uso con i sistemi di archiviazione ONTAP.
- ONTAP offre proprie funzionalità di efficienza dello storage, come deduplica, compressione e compaction, che possono essere influenzate dallo Storage DRS.
- Se si utilizzano snapshot ONTAP, storage vMotion lascerebbe la copia della macchina virtuale nella snapshot, aumentando potenzialmente l'utilizzo dello storage e potrebbe avere un impatto sulle applicazioni di backup, come NetApp SnapCenter, che tengono traccia delle macchine virtuali e delle relative snapshot ONTAP.

1. Selezionare il cluster ha e fare clic su Next.



1. Selezionare gli archivi dati appartenenti al sito A e fare clic su Avanti.

1. Rivedere le opzioni e fare clic su fine.
2. Ripetere questa procedura per creare il cluster di datastore del sito B e verificare che siano selezionati solo i datastore del sito B.

Disponibilità di vCenter Server

Le appliance vCenter Server (VCSA) devono essere protette con vCenter ha. VCenter ha ti consente di implementare due VCSA in una coppia ha Active-passive. Uno in ogni dominio di errore. Puoi leggere ulteriori informazioni su vCenter ha all'indirizzo "docs.vmware.com".

Resilienza per eventi pianificati e non pianificati

NetApp MetroCluster e SnapMirror Active Sync sono potenti strumenti che migliorano l'alta disponibilità e le operazioni senza interruzioni dell'hardware NetApp e del software ONTAP®.

Questi strumenti garantiscono una protezione a livello di sito per l'intero ambiente di storage, garantendo che i tuoi dati siano sempre disponibili. Che si stia utilizzando server standalone, cluster di server ad alta disponibilità, container o server virtualizzati, la tecnologia NetApp permette di conservare perfettamente la disponibilità dello storage in caso di black-out totale causato da black-out, raffreddamento o connettività di rete, arresto dello storage array o errori operativi.

La sincronizzazione attiva di MetroCluster e SnapMirror offre tre metodi di base per la continuità dei dati in caso di eventi pianificati o non pianificati:

- Componenti ridondanti per la protezione contro i guasti a un singolo componente
- Takeover locale di ha in caso di eventi che colpiscono un singolo controller
- Protezione completa del sito: Rapida ripresa del servizio mediante il trasferimento dello storage e dell'accesso client dal cluster di origine al cluster di destinazione

Ciò significa che le operazioni continuano senza problemi in caso di guasto a un singolo componente e vengono ripristinate automaticamente al funzionamento ridondante una volta sostituito il componente guasto.

Tutti i cluster ONTAP, ad eccezione dei cluster a nodo singolo (in genere versioni software-defined, come ad esempio ONTAP Select), offrono funzionalità di ha integrate chiamate takeover e giveback. Ciascun controller del cluster è accoppiato con un altro controller in modo da formare una coppia ha. Queste coppie garantiscono che ogni nodo sia connesso localmente allo storage.

Il takeover è un processo automatizzato in cui un nodo assume il controllo dello storage dell'altro per la gestione dei servizi dati. Giveback è il processo inverso che ripristina il normale funzionamento. Il takeover può essere pianificato, ad esempio durante la manutenzione hardware o gli upgrade della ONTAP, o non pianificato, derivante da un nodo di panico o da un guasto dell'hardware.

Durante un takeover, le LIF NAS nelle configurazioni MetroCluster eseguono automaticamente il failover. Tuttavia, le LIF SAN non vengono sottoposte a failover e continueranno a utilizzare il percorso diretto dei LUN (Logical Unit Number).

Per ulteriori informazioni sul takeover e sullo sconto ha, consultare la "Panoramica sulla gestione delle coppie HA". È importante notare che questa funzionalità non è specifica per MetroCluster o SnapMirror Active Sync.

Lo switchover del sito con MetroCluster viene eseguito quando un sito è offline o come attività pianificata per la manutenzione di un intero sito. Il sito rimanente presuppone la proprietà delle risorse storage (dischi e aggregati) del cluster offline e le SVM del sito guasto vengono messe online e riavviate nel sito di disaster recovery, preservando la loro identità completa per l'accesso client e host.

Con la sincronizzazione attiva di SnapMirror, poiché entrambe le copie vengono utilizzate contemporaneamente in modo attivo, gli host esistenti continueranno a funzionare. Il ONTAP Mediator è necessario per garantire che il failover del sito avvenga correttamente.

Scenari di errore per vMSC con MetroCluster

Nelle sezioni seguenti vengono illustrati i risultati attesi da vari scenari di guasto con i sistemi vMSC e NetApp MetroCluster.

Errore singolo percorso di storage

In questo scenario, se componenti come la porta HBA, la porta di rete, la porta dello switch dati front-end o un cavo FC o Ethernet si guastano, quel particolare percorso al dispositivo di storage viene contrassegnato come inattivo dall'host ESXi. Se vengono configurati diversi percorsi per il dispositivo storage fornendo resilienza alla porta HBA/rete/switch, ESXi esegue uno switchover del percorso. Durante questo periodo, le macchine virtuali rimangono in esecuzione senza alcun impatto, perché la disponibilità dello storage viene garantita attraverso l'offerta di più percorsi al dispositivo di storage.

In questo scenario non vi sono cambiamenti nel comportamento di MetroCluster e tutti i datastore continuano a essere intatti dai rispettivi siti.

Best practice

Negli ambienti in cui vengono utilizzati volumi NFS/iSCSI, NetApp consiglia di avere almeno due uplink di rete configurati per la porta vmkernel NFS nel vSwitch standard e lo stesso nel gruppo di porte in cui è mappata l'interfaccia vmkernel NFS per il vSwitch distribuito. Il raggruppamento NIC può essere configurato in modalità Active-Active o Active-standby.

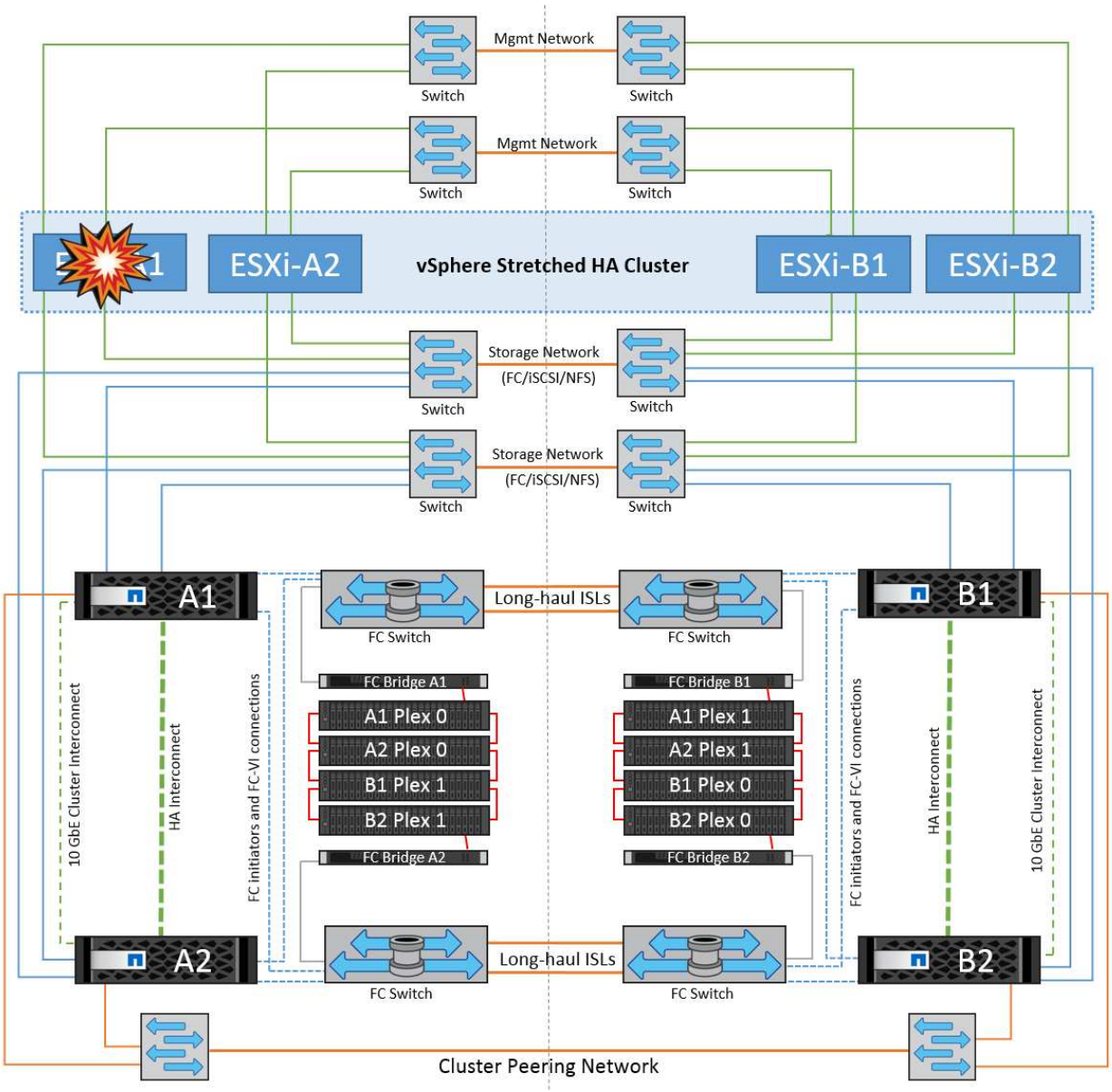
Inoltre, per i LUN iSCSI, il multipathing deve essere configurato legando le interfacce vmkernel agli adattatori di rete iSCSI. Per ulteriori informazioni, fai riferimento alla documentazione dello storage vSphere.

Best practice

Negli ambienti in cui vengono utilizzate le LUN Fibre Channel, NetApp consiglia di disporre di almeno due HBA, che garantiscono resilienza a livello di HBA/porta. NetApp consiglia inoltre di utilizzare lo zoning a destinazione singola come Best practice per la configurazione dello zoning.

È necessario utilizzare Virtual Storage Console (VSC) per impostare policy di multipathing, perché imposta policy per tutti i dispositivi storage NetApp nuovi ed esistenti.

Errore host ESXi singolo



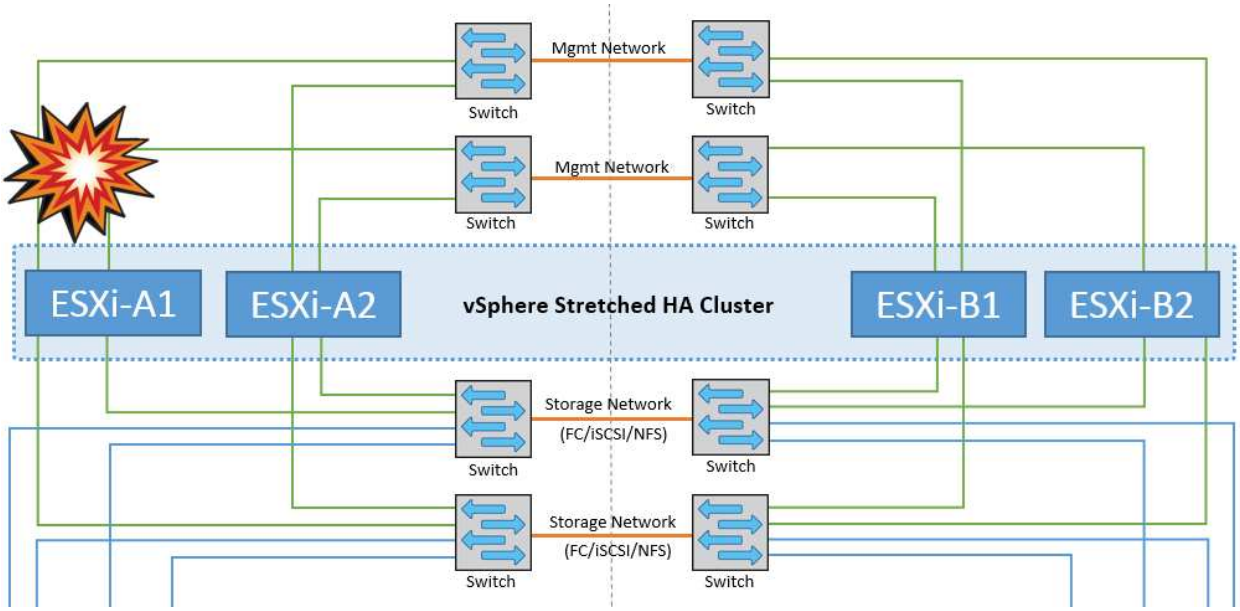
In questo scenario, se si verifica un guasto dell'host ESXi, il nodo master nel cluster VMware ha rileva il guasto dell'host in quanto non riceve più gli heartbeat di rete. Per determinare se l'host è effettivamente inattivo o solo una partizione di rete, il nodo master monitora gli heartbeat del datastore e, se sono assenti, esegue un controllo finale eseguendo il ping degli indirizzi IP di gestione dell'host guasto. Se tutti questi controlli sono negativi, il nodo master dichiara l'host un host guasto e tutte le macchine virtuali in esecuzione su questo host guasto vengono riavviate sull'host rimasto nel cluster.

Se sono state configurate le regole di affinità per DRS VM e host (le VM nel gruppo VM sitea_VM devono eseguire gli host nel gruppo host sitea_hosts), il master ha controlla prima le risorse disponibili nel sito A. Se non ci sono host disponibili nel sito A, il master tenta di riavviare le VM sugli host nel sito B.

È possibile che le macchine virtuali vengano avviate sugli host ESXi nell'altro sito se è presente un vincolo di risorse nel sito locale. Tuttavia, le regole di affinità definite per DRS VM e host verranno corrette in caso di violazione di regole mediante la migrazione delle macchine virtuali a qualsiasi host ESXi rimasto nel sito locale. Nei casi in cui DRS è impostato su manuale, NetApp consiglia di richiamare DRS e applicare le raccomandazioni per correggere il posizionamento della macchina virtuale.

In questo scenario, non vi sono cambiamenti nel comportamento di MetroCluster e tutti i datastore continuano a essere intatti dai rispettivi siti.

Isolamento dell'host ESXi

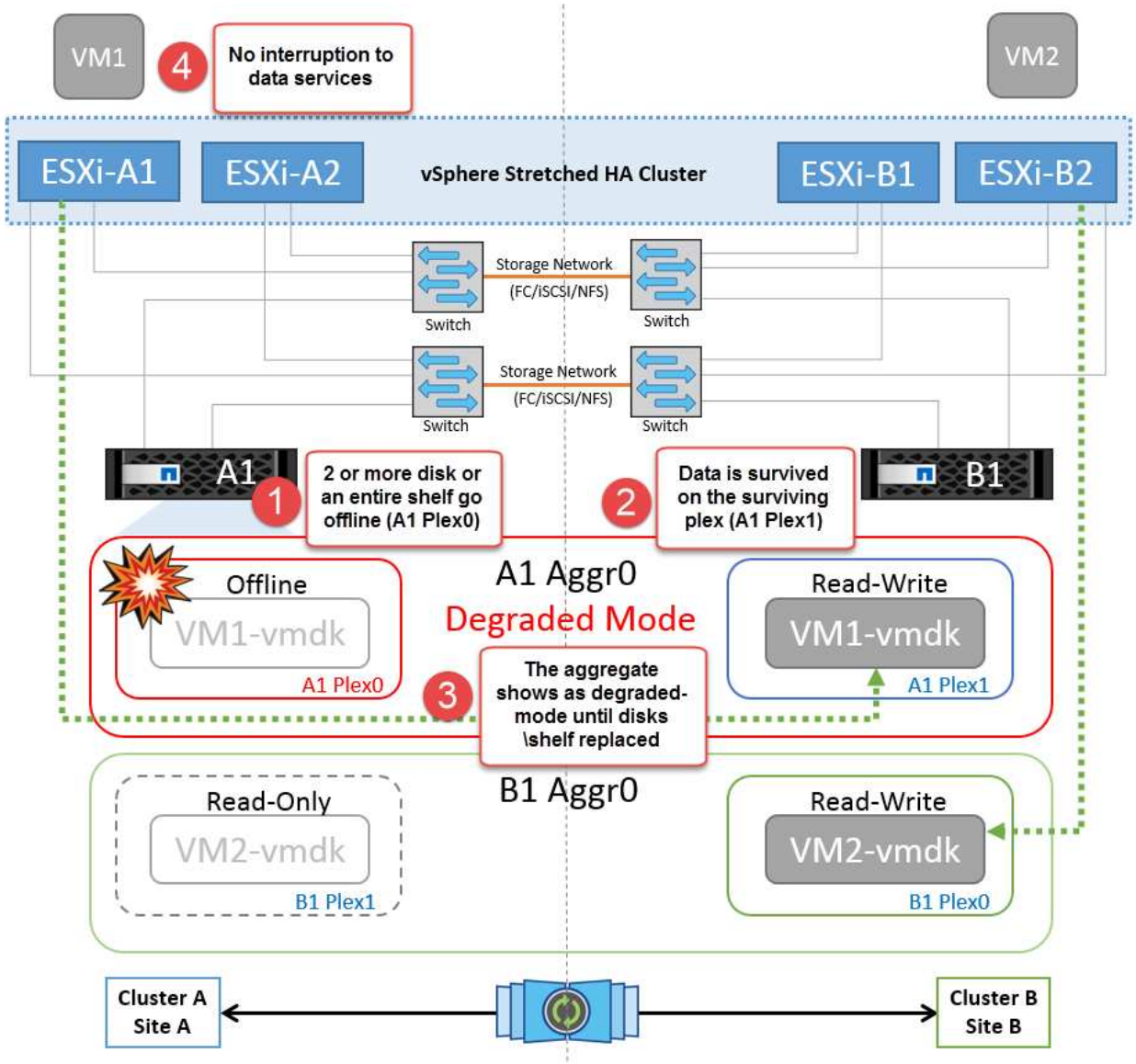


In questo scenario, se la rete di gestione dell'host ESXi non è attiva, il nodo master nel cluster ha non riceverà alcun heartbeat, pertanto l'host viene isolato nella rete. Per determinare se si è verificato un errore o se è solo isolato, il nodo master inizia a monitorare l'heartbeat del datastore. Se è presente, l'host viene dichiarato isolato dal nodo master. A seconda della risposta di isolamento configurata, l'host può scegliere di spegnere, spegnere le macchine virtuali o persino lasciare accese le macchine virtuali. L'intervallo predefinito per la risposta di isolamento è di 30 secondi.

In questo scenario, non vi sono cambiamenti nel comportamento di MetroCluster e tutti i datastore continuano a essere intatti dai rispettivi siti.

Guasto a shelf di dischi

In questo scenario, si verifica un errore di più di due dischi o di un intero shelf. I dati vengono distribuiti dal plesso restante senza alcuna interruzione dei servizi dati. Il guasto del disco potrebbe influire su un plesso locale o remoto. Gli aggregati vengono visualizzati come modalità degradata perché è attivo un solo plesso. Una volta sostituiti i dischi guasti, gli aggregati interessati si risincronizzano automaticamente per ricostruire i dati. Dopo la risincronizzazione, gli aggregati tornano automaticamente alla normale modalità con mirroring. Se più di due dischi all'interno di un singolo gruppo RAID si sono guastati, il plex deve essere ricostruito.

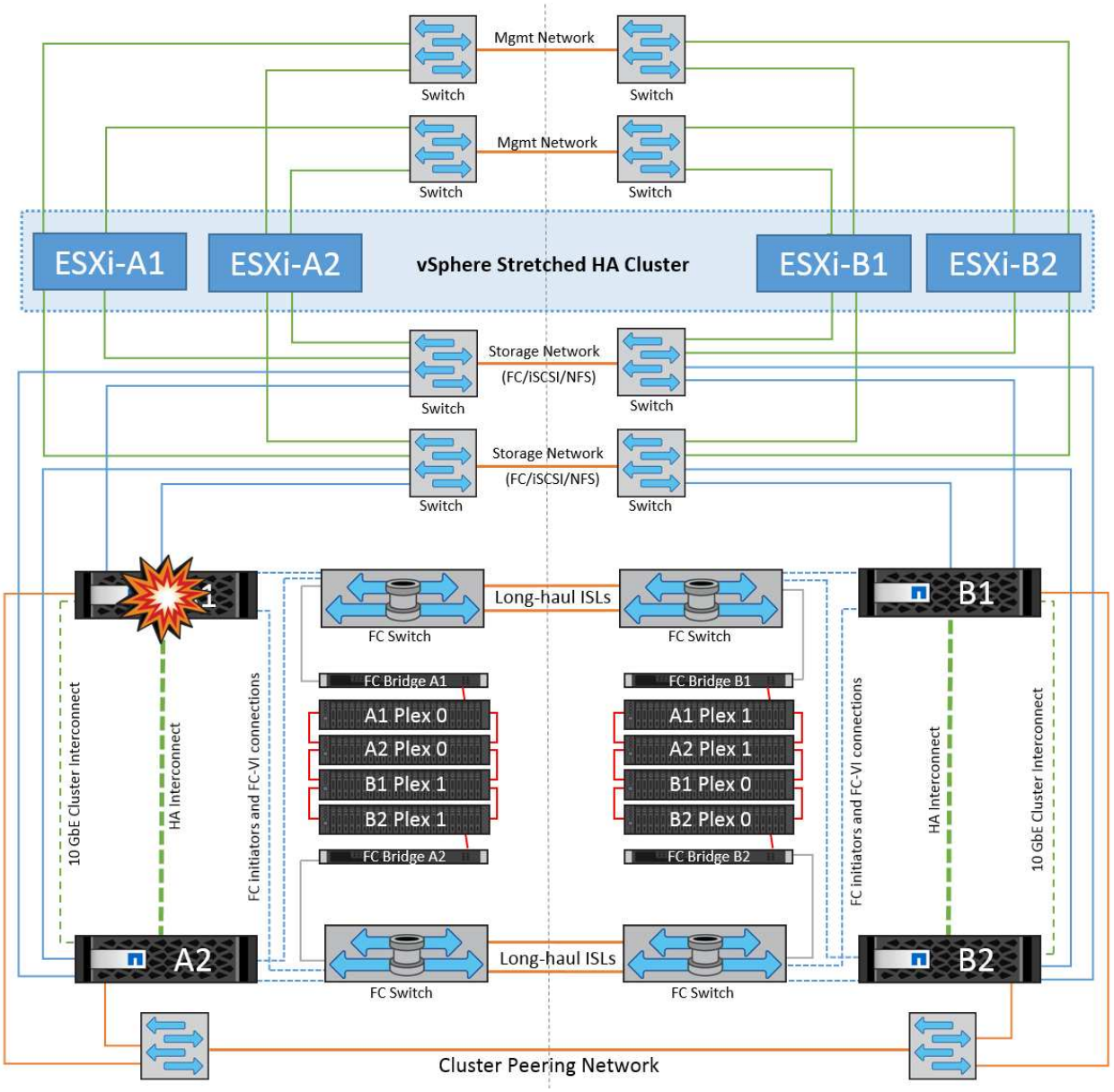


*[NOTA]

- Durante questo periodo, non si verifica alcun impatto sulle operazioni di i/o della macchina virtuale, tuttavia le performance sono peggiorate a causa dell'accesso ai dati dallo shelf di dischi remoto attraverso link ISL.

Guasto a un singolo storage controller

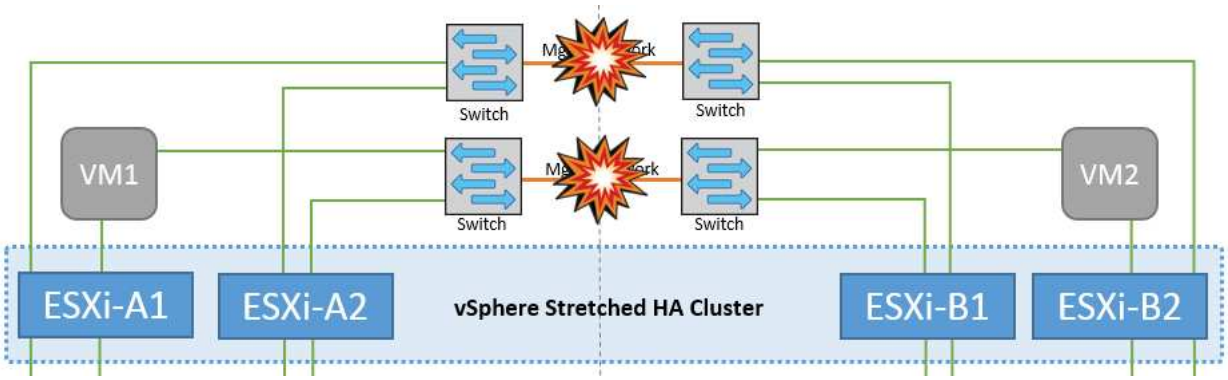
In questo scenario, uno dei due storage controller si guasta in un solo sito. Poiché è presente una coppia ha in ciascun sito, un guasto di un nodo attiva automaticamente il failover sull'altro nodo. Ad esempio, in caso di guasto al nodo A1, il relativo storage e carichi di lavoro vengono trasferiti automaticamente al nodo A2. Le macchine virtuali non saranno interessate perché tutti i plessi rimangono disponibili. I nodi del secondo sito (B1 e B2) non sono interessati. Inoltre, vSphere ha non intraprenderà alcuna azione perché il nodo master nel cluster riceverà comunque gli heartbeat di rete.



Se il failover fa parte di un rolling disaster (il nodo A1 esegue il failover su A2) e si verifica un successivo guasto di A2 o il guasto completo del sito A, è possibile eseguire lo switchover in seguito a un disastro nel sito B.

Errori del collegamento interswitch

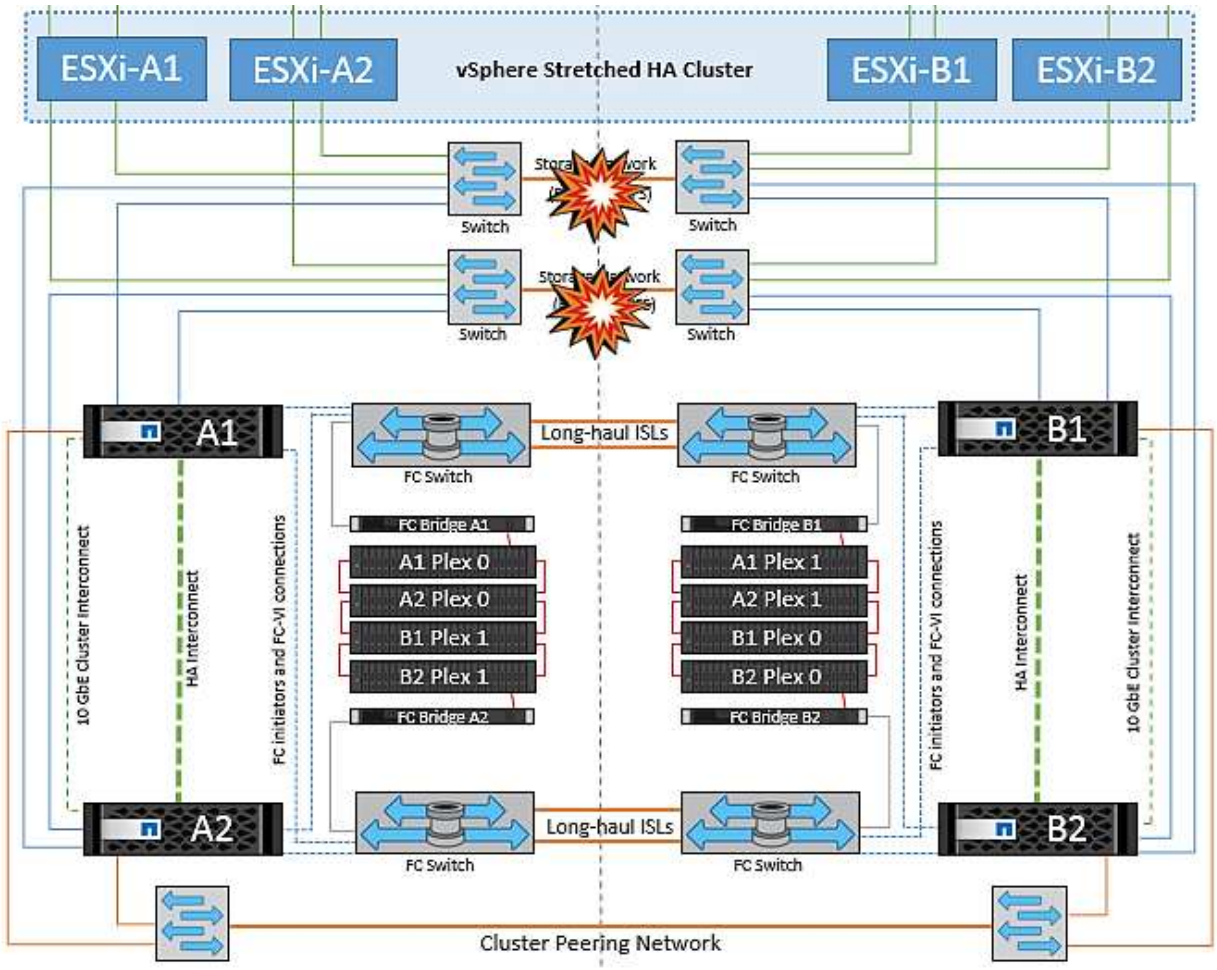
Errore collegamento interswitch sulla rete di gestione



In questo scenario, se i collegamenti ISL nella rete di gestione host front-end si guastano, gli host ESXi nel sito A non saranno in grado di comunicare con gli host ESXi nel sito B. Ciò determina una partizione di rete poiché gli host ESXi in un determinato sito non sono in grado di inviare gli heartbeat di rete al nodo master nel cluster ha. Come tale, ci saranno due segmenti di rete a causa della partizione e vi sarà un nodo master in ogni segmento che proteggerà le VM da guasti host all'interno del sito specifico.

Durante questo periodo, le macchine virtuali rimangono in esecuzione e in questo scenario non si verifica alcuna modifica nel comportamento di MetroCluster. Tutti i datastore continuano a essere intatti dai rispettivi siti.

Errore collegamento interswitch sulla rete di storage

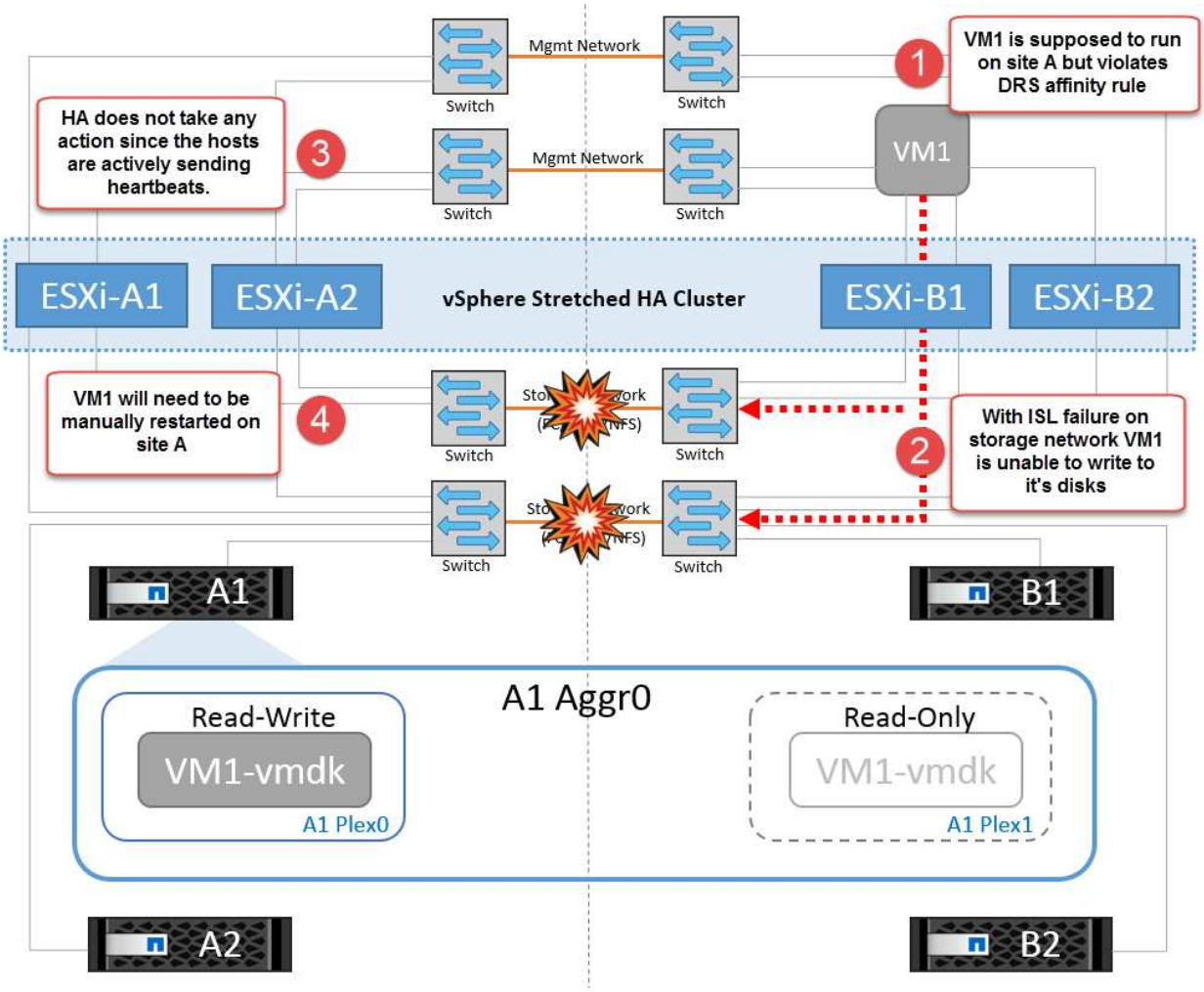


In questo scenario, se si verifica un errore nei collegamenti ISL nella rete di storage backend, gli host sul sito A perderanno l'accesso ai volumi di storage o alle LUN del cluster B nel sito B e viceversa. Le regole VMware DRS sono definite in modo che l'affinità tra il sito host e il sito di storage faciliti l'esecuzione delle macchine virtuali senza impatti all'interno del sito.

Durante questo periodo, le macchine virtuali rimangono in esecuzione nei rispettivi siti e in questo scenario non si verifica alcuna modifica nel comportamento di MetroCluster. Tutti i datastore continuano a essere intatti dai rispettivi siti.

Se per qualche motivo è stata violata la regola di affinità (ad esempio VM1, che doveva essere eseguito dal sito A in cui i dischi risiedono sui nodi del cluster locale A vengono eseguiti su un host nel sito B), il disco della macchina virtuale può essere acceduto in remoto tramite i link ISL. A causa di un errore del collegamento ISL, VM1 in esecuzione nel sito B non sarebbe in grado di scrivere sui propri dischi perché i percorsi del volume di

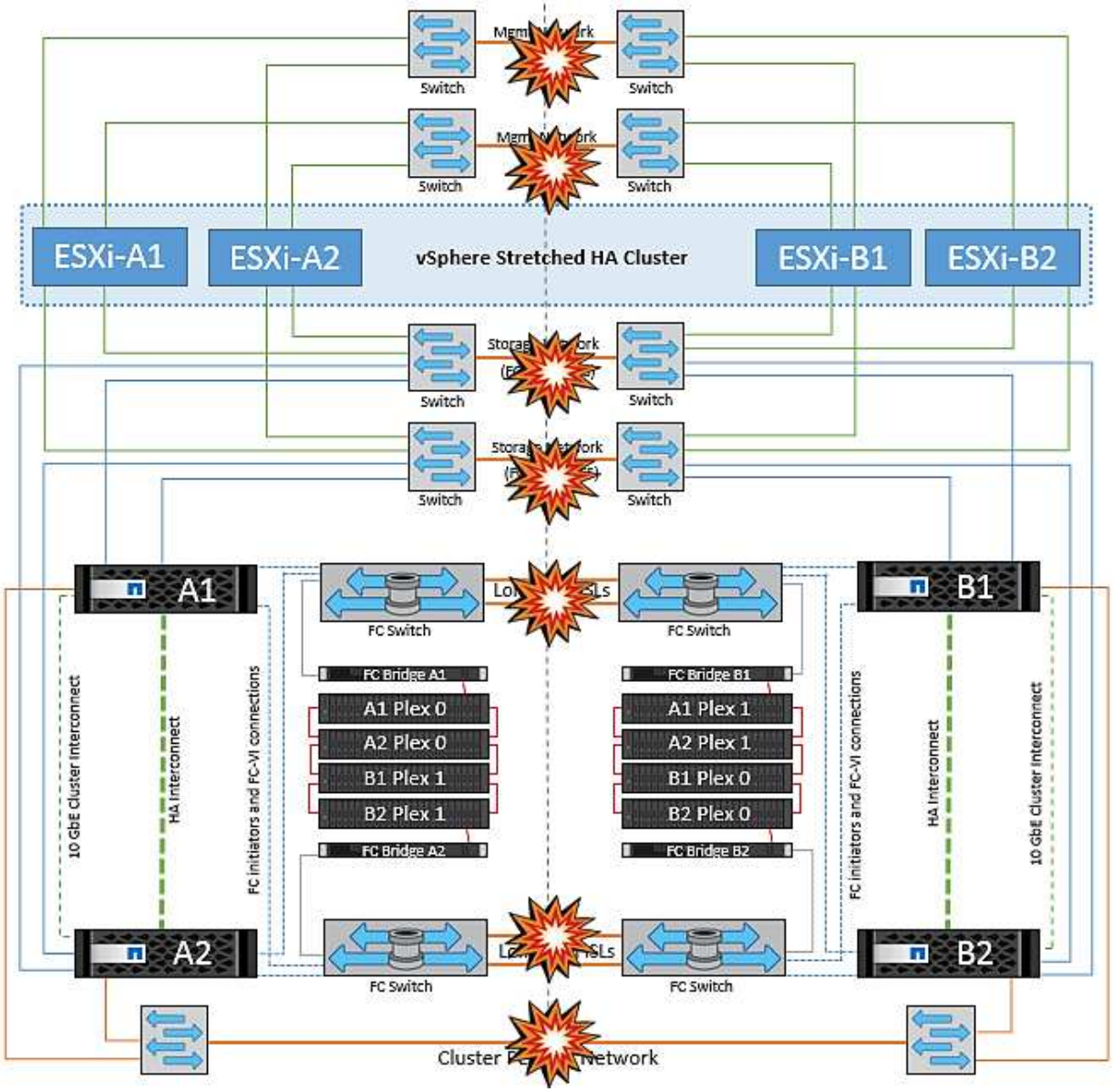
storage non sono attivi e quella particolare macchina virtuale non è attiva. In queste situazioni, VMware ha non intraprende alcuna azione poiché gli host stanno inviando heartbeat. Tali macchine virtuali devono essere spente e attivate manualmente nei rispettivi siti. La figura seguente illustra una VM che viola una regola di affinità DRS.



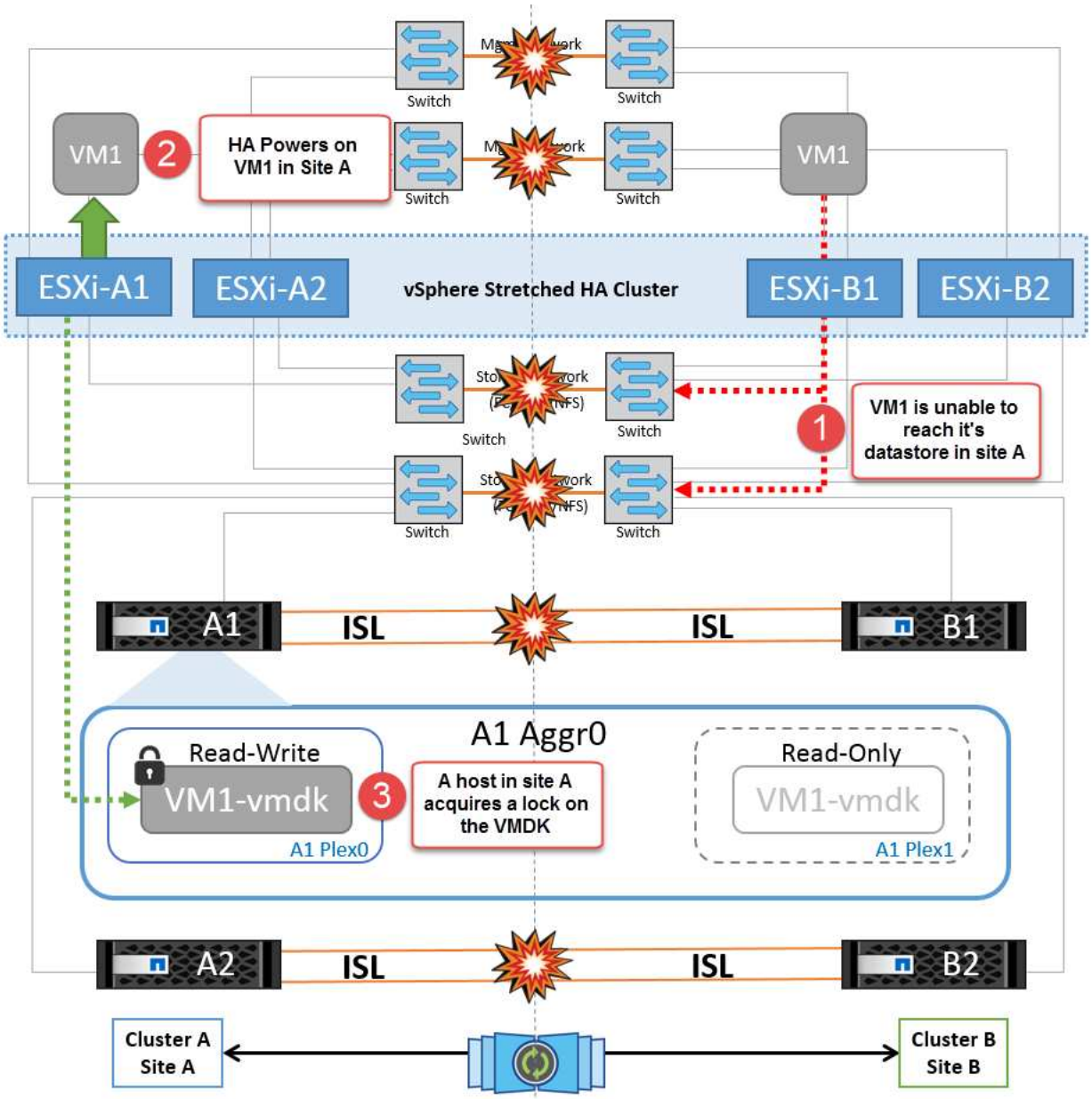
Guasto a tutti gli interswitch o partizione completa del data center

In questo scenario, tutti i collegamenti ISL tra i siti sono interrotti ed entrambi i siti sono isolati l'uno dall'altro. Come discusso in scenari precedenti, come ad esempio un errore ISL nella rete di gestione e nella rete di storage, le macchine virtuali non sono interessate da un errore ISL completo.

Dopo la partizione degli host ESXi tra i siti, l'agente vSphere ha controlla gli heartbeat del datastore e, in ciascun sito, gli host ESXi locali saranno in grado di aggiornare gli heartbeat del datastore nei rispettivi volumi/LUN di lettura/scrittura. Gli host nel sito A supporteranno che gli altri host ESXi presenti nel sito B siano guasti a causa dell'assenza di heartbeat di rete/datastore. VSphere ha nel sito A tenterà di riavviare le macchine virtuali del sito B con un errore infine dovuto al fatto che i datastore del sito B non saranno accessibili a causa di un guasto all'ISL di storage. Una situazione simile si ripete nel sito B.



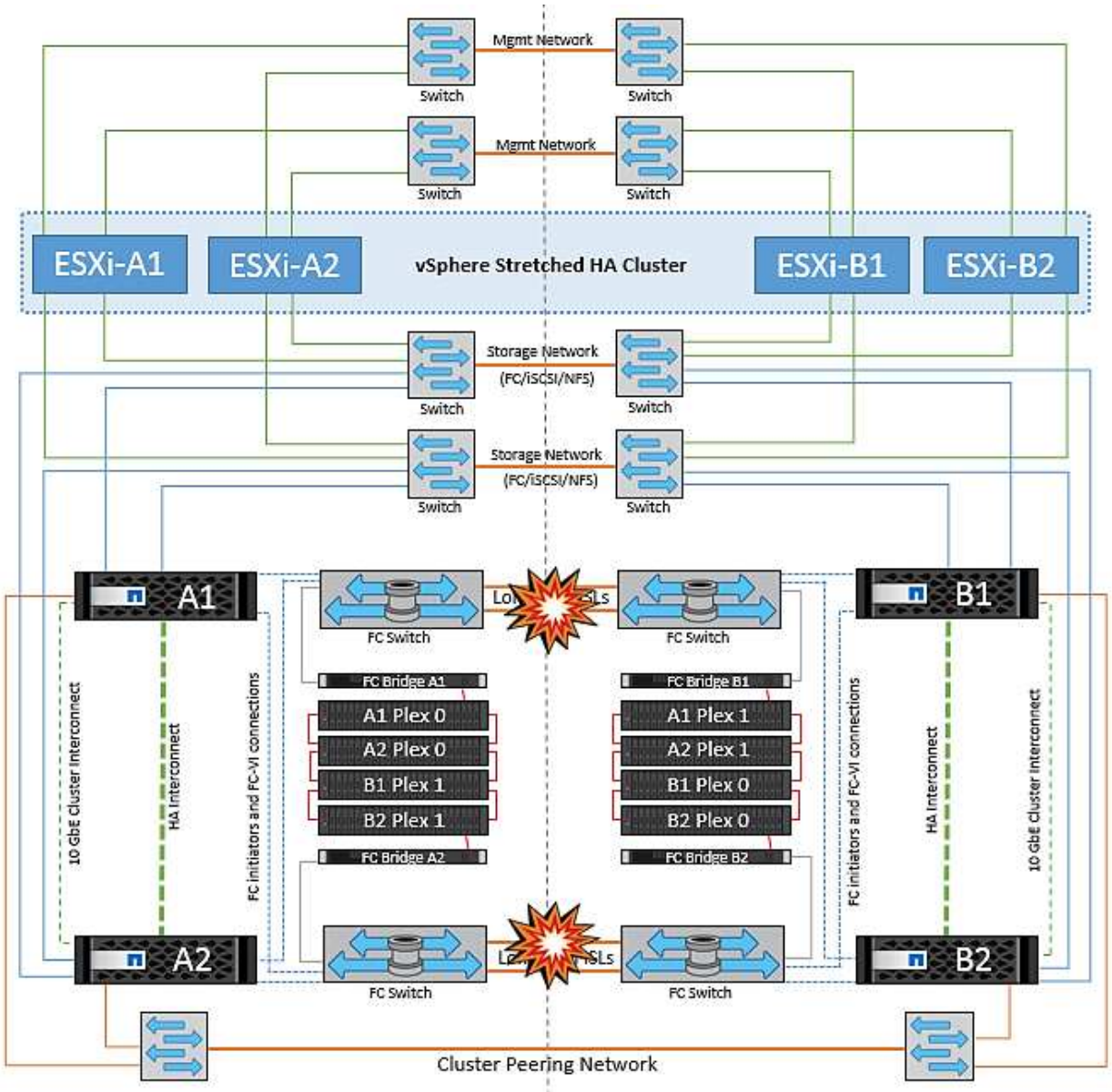
NetApp consiglia di determinare se una macchina virtuale ha violato le regole DRS. Tutte le macchine virtuali in esecuzione da un sito remoto non potranno accedere al datastore, quindi vSphere ha riavvia la macchina virtuale nel sito locale. Una volta che i collegamenti ISL sono tornati in linea, la macchina virtuale in esecuzione nel sito remoto verrà interrotta, poiché non possono esistere due istanze di macchine virtuali in esecuzione con gli stessi indirizzi MAC.



Errore collegamento interswitch su entrambi i fabric in NetApp MetroCluster

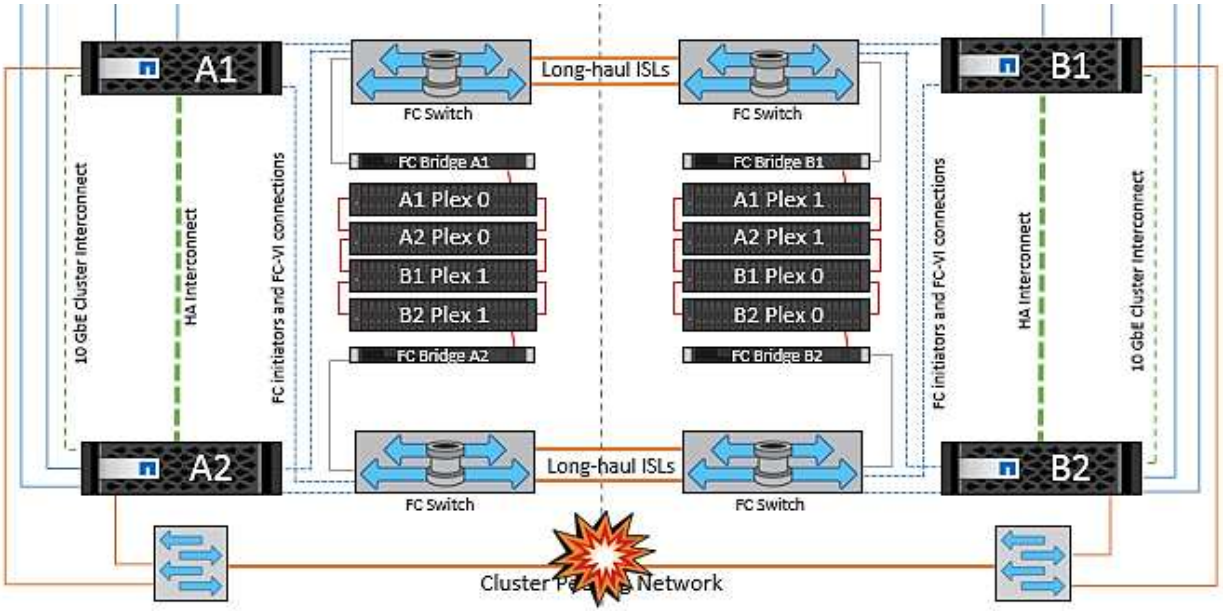
In uno scenario di errore di uno o più ISL, il traffico continua attraverso i collegamenti rimanenti. In caso di errore di tutti gli ISL su entrambi i fabric, in modo da eliminare un collegamento tra i siti per la replica di storage e NVRAM, ciascun controller continuerà a fornire i propri dati locali. Su un minimo di un ISL viene ripristinato, la risincronizzazione di tutti i plessi avviene automaticamente.

Eventuali scritture che si verificano dopo che tutti gli ISL sono inattivi non verranno mirrorate nell'altro sito. Uno switchover in caso di disastro, mentre la configurazione si trova in questo stato, causerebbe una perdita dei dati non sincronizzati. In questo caso, è necessario un intervento manuale per il ripristino dopo lo switchover. Se è probabile che non saranno disponibili ISL per un periodo prolungato, un amministratore può scegliere di arrestare tutti i servizi dati per evitare il rischio di perdita di dati se occorre eseguire uno switchover in caso di disastro. L'esecuzione di questa azione deve essere valutata rispetto alla probabilità che un evento disastroso richieda lo switchover prima che almeno un ISL diventi disponibile. In alternativa, in caso di errore degli ISL in uno scenario a cascata, un amministratore può attivare uno switchover pianificato verso uno dei siti prima che tutti i collegamenti abbiano avuto esito negativo.



Errore collegamento cluster in peering

In uno scenario di guasto al link del cluster in peering, poiché gli ISL del fabric sono ancora attivi, i servizi dati (letture e scritture) continuano in entrambi i siti verso entrambi i plessi. Eventuali modifiche alla configurazione del cluster, ad esempio l'aggiunta di una nuova SVM, il provisioning di un volume o di una LUN in una SVM esistente, non possono essere propagate all'altro sito. Questi vengono conservati nei volumi di metadati CRS locali e propagati automaticamente all'altro cluster al recupero del collegamento al cluster sottoposto a peering. Se occorre uno switchover forzato prima del ripristino del link del cluster in peering, le modifiche alla configurazione del cluster in sospeso verranno riprodotte automaticamente dalla copia replicata remota dei volumi di metadati presenti nel sito rimasto nel processo di switchover.



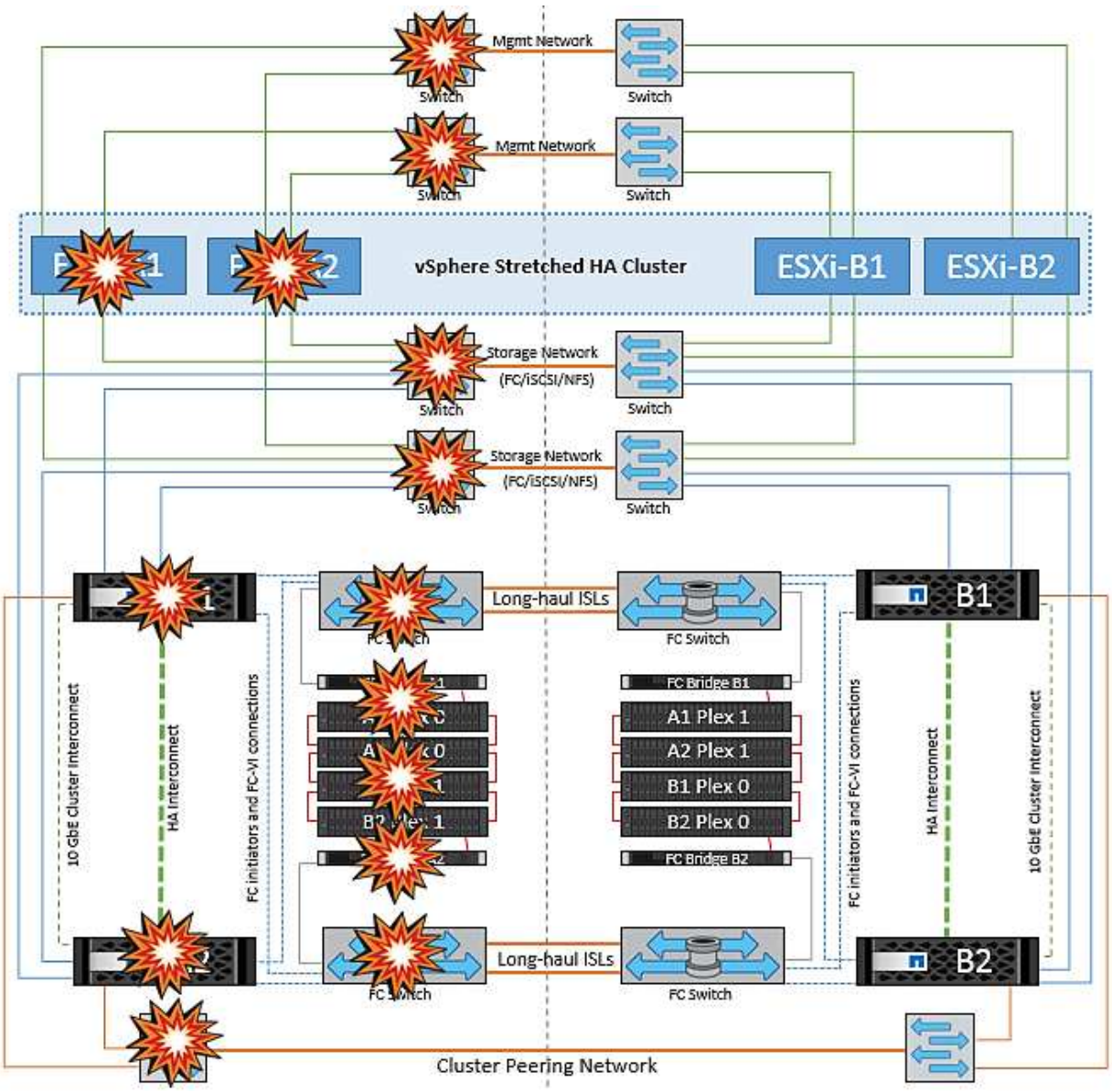
Errore completo del sito

In uno scenario di guasto completo del sito A, gli host ESXi nel sito B non otterranno l'heartbeat di rete dagli host ESXi nel sito A perché non sono attivi. Il master ha nel sito B verificherà che gli heartbeat del datastore non siano presenti, dichiarerà che gli host nel sito A non sono riusciti e tenterà di riavviare le macchine virtuali del sito A nel sito B. Durante questo periodo, l'amministratore dello storage esegue uno switchover per riprendere i servizi dei nodi guasti del sito rimasto e ripristinare i servizi di storage del sito A del sito B. Dopo che i volumi o le LUN del sito A sono disponibili nel sito B, l'agente master ha tenterà di riavviare le macchine virtuali del sito A nel sito B.

Se il tentativo dell'agente master vSphere ha di riavviare una VM (che comporta la registrazione e l'accensione) non riesce, il riavvio viene rieseguito dopo un ritardo. Il ritardo tra i riavvii può essere configurato fino a un massimo di 30 minuti. VSphere ha tenta di riavviare il sistema per un numero massimo di tentativi (sei tentativi per impostazione predefinita).

> Il master ha non avvia i tentativi di riavvio fino a quando il placement manager non trova lo storage appropriato, quindi in caso di un guasto completo del sito, ciò si verificherebbe dopo l'esecuzione dello switchover.

Se il sito A è stato sottoposto a switchover, un guasto successivo di uno dei nodi del sito B sopravvissuto può essere gestito senza problemi attraverso il failover verso il nodo rimasto. In questo caso, il lavoro di quattro nodi viene ora eseguito da un solo nodo. Il ripristino in questo caso consisterebbe nell'esecuzione di un giveback al nodo locale. Quindi, quando il sito A viene ripristinato, viene eseguita un'operazione di switchback per ripristinare il funzionamento regolare della configurazione.



Sicurezza dei prodotti

Strumenti ONTAP per VMware vSphere

La progettazione software con strumenti ONTAP per VMware vSphere si avvale delle seguenti attività di sviluppo sicure:

- **Modellazione delle minacce.** lo scopo della modellazione delle minacce è quello di individuare i difetti di sicurezza in una funzionalità, un componente o un prodotto nelle prime fasi del ciclo di vita dello sviluppo software. Un modello di minaccia è una rappresentazione strutturata di tutte le informazioni che influiscono sulla sicurezza di un'applicazione. In sostanza, si tratta di una vista dell'applicazione e del suo ambiente attraverso l'obiettivo della sicurezza.
- **Dynamic Application Security Testing (DAST).** questa tecnologia è progettata per rilevare le condizioni vulnerabili delle applicazioni in esecuzione. DAST testa le interfacce HTTP e HTML esposte delle applicazioni web-enable.
- **Valuta del codice di terze parti.** nell'ambito dello sviluppo di software con software open-source (OSS), è necessario risolvere le vulnerabilità di sicurezza che potrebbero essere associate a qualsiasi OSS

incorporato nel prodotto. Si tratta di un'operazione continua, in quanto una nuova versione di OSS potrebbe presentare una vulnerabilità scoperta di recente in qualsiasi momento.

- **Scansione delle vulnerabilità.** lo scopo della scansione delle vulnerabilità è quello di rilevare vulnerabilità di sicurezza comuni e note nei prodotti NetApp prima che vengano rilasciate ai clienti.
- Test di penetrazione.* il test di penetrazione è il processo di valutazione di un sistema, di un'applicazione Web o di una rete per individuare le vulnerabilità di sicurezza che potrebbero essere sfruttate da un utente malintenzionato. I test di penetrazione (test delle penne) di NetApp vengono condotti da un gruppo di aziende terze approvate e fidate. Il loro scopo di test include il lancio di attacchi contro un'applicazione o un software simile a intrusi o hacker ostili che utilizzano sofisticati metodi o strumenti di sfruttamento.

Funzionalità di sicurezza del prodotto

I tool ONTAP per VMware vSphere includono le seguenti funzionalità di sicurezza in ciascuna release.

- **Login banner.** SSH è disattivato per impostazione predefinita e consente l'accesso una sola volta, se abilitato dalla console della macchina virtuale. Il seguente banner di accesso viene visualizzato dopo che l'utente ha inserito un nome utente nel prompt di accesso:

  ATTENZIONE: l'accesso non autorizzato a questo sistema è vietato e sarà perseguito dalla legge. Accedendo a questo sistema, l'utente accetta che le proprie azioni possano essere monitorate in caso di sospetto di utilizzo non autorizzato.

  Dopo che l'utente ha completato l'accesso tramite il canale SSH, viene visualizzato il seguente testo:

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- **RBAC (role-based access control).** due tipi di controlli RBAC sono associati ai tool ONTAP:
  - Privilegi vCenter Server nativi
  - Privilegi specifici del plug-in vCenter. Per ulteriori informazioni, vedere "questo link".
- **Canali di comunicazione crittografati.** tutte le comunicazioni esterne avvengono su HTTPS utilizzando la versione 1.2 di TLS.
- **Esposizione minima delle porte.** solo le porte necessarie sono aperte sul firewall.

  La seguente tabella descrive i dettagli della porta aperta.

Porta TCP v4/v6 n.	Direzione	Funzione
8143	in entrata	Connessioni HTTPS per API REST
8043	in entrata	Connessioni HTTPS

Porta TCP v4/v6 n.	Direzione	Funzione
9060	in entrata	Connessioni HTTPS Utilizzato per connessioni SOAP su https Questa porta deve essere aperta per consentire a un client di connettersi al server API degli strumenti ONTAP.
22	in entrata	SSH (Disattivato per impostazione predefinita)
9080	in entrata	Connessioni HTTPS - VP e SRA - connessioni interne solo da loopback
9083	in entrata	Connessioni HTTPS - VP e SRA Utilizzato per connessioni SOAP su https
1162	in entrata	Pacchetti di trap SNMP VP
1527	solo interno	Porta del database Derby, solo tra questo computer e se stesso, connessioni esterne non accettate — solo connessioni interne
443	bidirezionale	Utilizzato per le connessioni ai cluster ONTAP

- **Supporto dei certificati firmati dall'autorità di certificazione (CA).** i tool ONTAP per VMware vSphere supportano i certificati firmati CA. Vedi questo "articolo della knowledge base" per ulteriori informazioni.
- **Registrazione audit.** i pacchetti di supporto possono essere scaricati e sono estremamente dettagliati. ONTAP Tools registra tutte le attività di login e logout degli utenti in un file di log separato. Le chiamate API VASA vengono registrate in un registro di controllo VASA dedicato (cxf.log locale).
- **Criteri per le password.** vengono seguite le seguenti policy per le password:
  - Le password non vengono registrate in alcun file di log.
  - Le password non vengono comunicate in testo normale.
  - Le password vengono configurate durante il processo di installazione.
  - La cronologia delle password è un parametro configurabile.
  - La durata minima della password è impostata su 24 ore.
  - Il completamento automatico dei campi della password è disattivato.
  - Gli strumenti ONTAP crittografano tutte le informazioni sulle credenziali memorizzate utilizzando l'hashing SHA256.

Plug-in di SnapCenter per VMware vSphere

Il plug-in NetApp SnapCenter per il software engineering VMware vSphere utilizza le seguenti attività di sviluppo sicuro:

- **Modellazione delle minacce.** lo scopo della modellazione delle minacce è quello di individuare i difetti di sicurezza in una funzionalità, un componente o un prodotto nelle prime fasi del ciclo di vita dello sviluppo software. Un modello di minaccia è una rappresentazione strutturata di tutte le informazioni che influiscono sulla sicurezza di un'applicazione. In sostanza, si tratta di una vista dell'applicazione e del suo ambiente attraverso l'obiettivo della sicurezza.
- **Dynamic Application Security testing (DAST).** tecnologie progettate per rilevare condizioni vulnerabili sulle applicazioni in esecuzione. DAST testa le interfacce HTTP e HTML esposte delle applicazioni web-enable.
- **Valuta del codice di terze parti.** come parte dello sviluppo di software e dell'utilizzo di software open-source (OSS), è importante risolvere le vulnerabilità di sicurezza che potrebbero essere associate a OSS che è stato incorporato nel prodotto. Si tratta di un impegno continuo, in quanto la versione del componente OSS potrebbe presentare una vulnerabilità scoperta di recente in qualsiasi momento.
- **Scansione delle vulnerabilità.** lo scopo della scansione delle vulnerabilità è quello di rilevare vulnerabilità di sicurezza comuni e note nei prodotti NetApp prima che vengano rilasciate ai clienti.
- Test di penetrazione.* il test di penetrazione è il processo di valutazione di un sistema, di un'applicazione Web o di una rete per individuare le vulnerabilità della sicurezza che potrebbero essere sfruttate da un utente malintenzionato. I test di penetrazione (test delle penne) di NetApp vengono condotti da un gruppo di aziende terze approvate e fidate. Il loro scopo di test include il lancio di attacchi contro un'applicazione o un software come intrusi o hacker ostili che utilizzano sofisticati metodi o strumenti di sfruttamento.
- **Attività di risposta agli incidenti di sicurezza dei prodotti.** le vulnerabilità di sicurezza sono scoperte sia internamente che esternamente all'azienda e possono rappresentare un serio rischio per la reputazione di NetApp se non vengono affrontate in modo tempestivo. Per facilitare questo processo, un Product Security Incident Response Team (PSIRT) segnala e tiene traccia delle vulnerabilità.

Funzionalità di sicurezza del prodotto

Il plug-in NetApp SnapCenter per VMware vSphere include le seguenti funzionalità di sicurezza in ciascuna release:

- **Accesso limitato alla shell.** SSH è disattivato per impostazione predefinita e gli accessi una tantum sono consentiti solo se sono abilitati dalla console della macchina virtuale.
- **Avviso di accesso nel banner di accesso.** il seguente banner di accesso viene visualizzato dopo che l'utente ha inserito un nome utente nel prompt di accesso:

  ATTENZIONE: l'accesso non autorizzato a questo sistema è vietato e sarà perseguito dalla legge. Accedendo a questo sistema, l'utente accetta che le proprie azioni possano essere monitorate in caso di sospetto di utilizzo non autorizzato.

  Una volta completato l'accesso tramite il canale SSH, viene visualizzato il seguente output:

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- **RBAC (role-based access control).** due tipi di controlli RBAC sono associati ai tool ONTAP:

- Privilegi vCenter Server nativi.
- Privilegi specifici del plug-in VMware vCenter. Per ulteriori informazioni, vedere "RBAC (Role-Based Access Control)".

- **Canali di comunicazione crittografati.** tutte le comunicazioni esterne avvengono su HTTPS utilizzando TLS.
- **Esposizione minima delle porte.** solo le porte necessarie sono aperte sul firewall.

La seguente tabella fornisce i dettagli della porta aperta.

Numero della porta TCP v4/v6	Funzione
8144	Connessioni HTTPS per API REST
8080	Connessioni HTTPS per GUI OVA
22	SSH (disattivato per impostazione predefinita)
3306	MySQL (solo connessioni interne; connessioni esterne disattivate per impostazione predefinita)
443	Nginx (servizi di protezione dei dati)

- **Supporto dei certificati firmati dall'autorità di certificazione (CA).** il plug-in SnapCenter per VMware vSphere supporta la funzione dei certificati firmati dalla CA. Vedere "Come creare e/o importare un certificato SSL nel plug-in SnapCenter per VMware vSphere (SCV)".
- **Password policy.** sono in vigore i seguenti criteri relativi alle password:
  - Le password non vengono registrate in alcun file di log.
  - Le password non vengono comunicate in testo normale.
  - Le password vengono configurate durante il processo di installazione.
  - Tutte le informazioni sulle credenziali vengono memorizzate utilizzando l'hashing SHA256.
- **Immagine del sistema operativo di base.** il prodotto viene fornito con il sistema operativo di base Debian per OVA con accesso limitato e accesso alla shell disattivato. In questo modo si riduce l'impatto degli attacchi. Ogni sistema operativo SnapCenter release base viene aggiornato con le ultime patch di sicurezza disponibili per la massima copertura di sicurezza.

NetApp sviluppa funzionalità software e patch di sicurezza per quanto riguarda il plug-in SnapCenter per l'appliance VMware vSphere e le rilascia ai clienti come piattaforma software integrata. Poiché queste appliance includono dipendenze specifiche del sistema operativo secondario Linux e il nostro software proprietario, NetApp consiglia di non apportare modifiche al sistema operativo secondario, in quanto questo potrebbe influire sull'appliance NetApp. Ciò potrebbe influire sulla capacità di NetApp di supportare l'appliance. NetApp consiglia di testare e implementare la versione più recente del codice per le appliance, perché vengono rilasciate per correggere eventuali problemi relativi alla sicurezza.

Guida alla protezione avanzata per gli strumenti ONTAP per VMware vSphere

Guida alla protezione avanzata per gli strumenti ONTAP per VMware vSphere 9,13

La guida alla protezione avanzata per gli strumenti ONTAP per VMware vSphere fornisce una serie completa di istruzioni per la configurazione delle impostazioni più sicure.

Queste guide si applicano sia alle applicazioni che al sistema operativo guest dell'appliance stessa.

Verifica dell'integrità dei tool ONTAP per i pacchetti di installazione di VMware vSphere 9,13

Sono disponibili due metodi per verificare l'integrità dei pacchetti di installazione degli strumenti ONTAP.

1. Verifica dei checksum
2. Verifica della firma

I checksum sono disponibili nelle pagine di download dei pacchetti di installazione di OTV. Gli utenti devono verificare i checksum dei pacchetti scaricati in base al checksum fornito nella pagina di download.

Verifica della firma degli strumenti ONTAP OVA

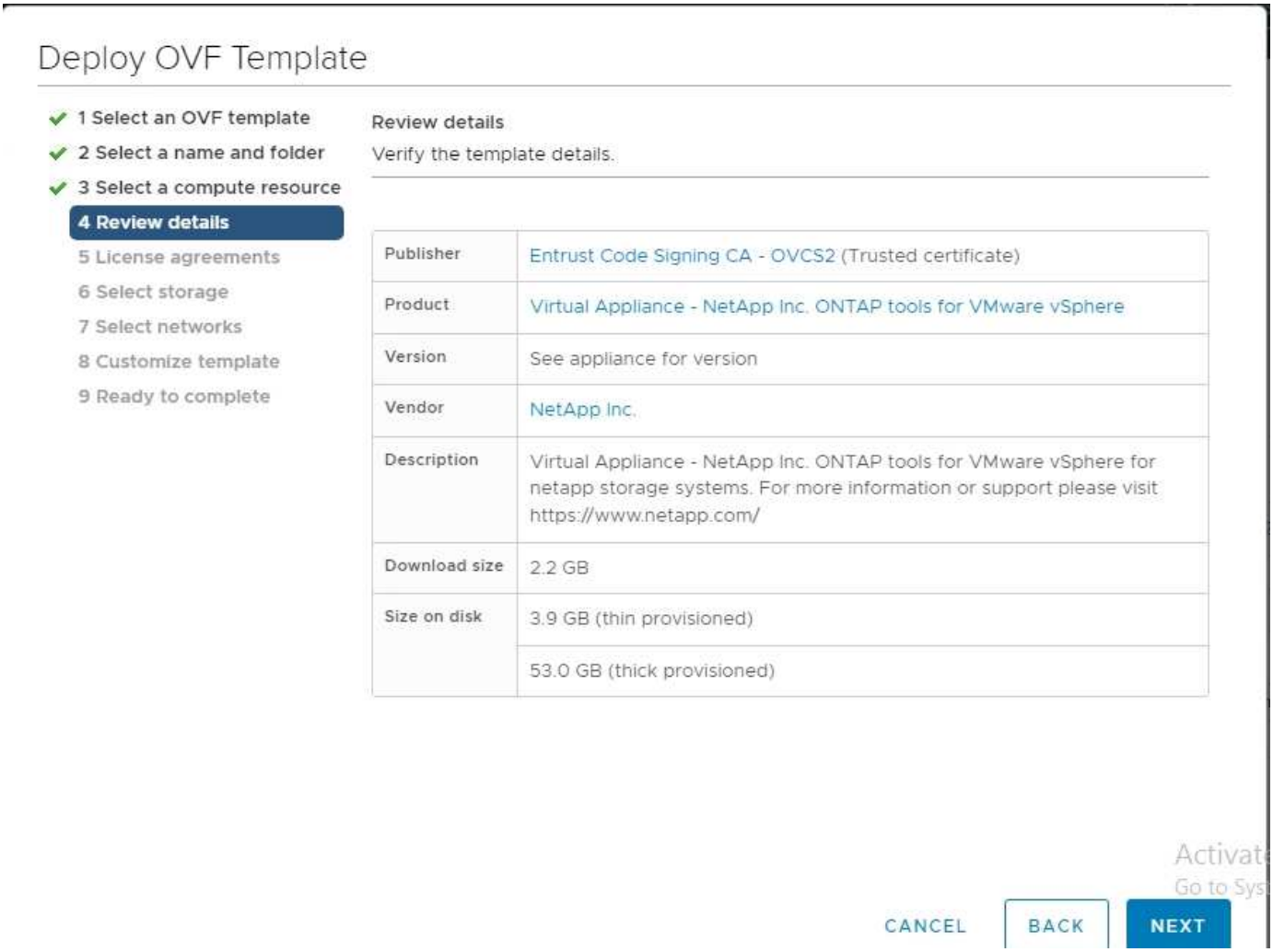
Il pacchetto di installazione vApp viene fornito sotto forma di tarball. Questo tarball contiene certificati intermedi e root per l'appliance virtuale insieme a un file README e un pacchetto OVA. Il file README guida gli utenti su come verificare l'integrità del pacchetto vApp OVA.

I clienti devono inoltre caricare il certificato root e intermedio fornito su vCenter versione 7.0U3E e successive. Per le versioni vCenter comprese tra 7.0.1 e 7,0.U3E, la funzionalità di verifica del certificato non è supportata da VMware. I clienti non devono caricare alcun certificato per le versioni 6.x. di vCenter

Caricamento del certificato root attendibile in vCenter

1. Accedere con il client VMware vSphere a vCenter Server.
2. Specificare il nome utente e la password per adminutator@vsphere.local o un altro membro del gruppo vCenter Single Sign-on Administrators. Se durante l'installazione è stato specificato un dominio diverso, accedere come Administrator@mydomain.
3. Accedere all'interfaccia utente di Gestione certificati: a. dal menu iniziale, selezionare Amministrazione. b. in certificati, fare clic su Gestione certificati.
4. Se richiesto dal sistema, immettere le credenziali di vCenter Server.
5. In certificati principali attendibili, fare clic su Aggiungi.
6. Fare clic su Sfoglia e selezionare la posizione del file .pem del certificato (OTV_OVA_INTER_ROOT_CERT_CHAIN.pem).
7. Fare clic su Aggiungi. Il certificato viene aggiunto al negozio.

Fare riferimento a. "Aggiungere un certificato radice attendibile all'archivio certificati" per ulteriori informazioni. Durante la distribuzione di una vApp (utilizzando il file OVA), la firma digitale per il pacchetto vApp può essere verificata nella pagina "Dettagli revisione". Se il pacchetto vApp scaricato è originale, nella colonna 'Publisher' viene visualizzato 'Trusted Certificate' (certificato attendibile) (come nella seguente schermata).



Verifica della firma degli attrezzi ONTAP ISO e SRA tar.gz

NetApp condivide il proprio certificato di firma del codice con i clienti nella pagina di download del prodotto, insieme ai file zip del prodotto per OTV-ISO e SRA.tgz.

Dal certificato di firma del codice, gli utenti possono estrarre la chiave pubblica nel modo seguente:

```
 #> openssl x509 -in <code-sign-cert, pem file> -pubkey -noout > <public-
key name>
Quindi, utilizzare la chiave pubblica per verificare la firma per il
prodotto zip iso e tgz come indicato di seguito:
```

```
 #> openssl dgst -sha256 -verify <public-key> -signature <signature-file>
<binary-name>
Esempio:
```

```
#> openssl x509 -in OTV_ISO_CERT.pem -pubkey -noout > OTV_ISO.pub
#> openssl dgst -sha256 -verify OTV_ISO.pub -signature netapp-ontap-tools-
for-vmware-vsphere-9.12-upgrade-iso.sig netapp-ontap-tools-for-vmware-
vsphere-9.12-upgrade.iso
Verified OK  => response
```

Porte e protocolli per gli strumenti ONTAP 9,13

Di seguito sono elencate le porte e i protocolli necessari per consentire la comunicazione tra gli strumenti ONTAP per il server VMware vSphere e altre entità come i sistemi di storage gestito, i server e altri componenti.

Porte in entrata e in uscita richieste per OTV

Annotare la tabella riportata di seguito che elenca le porte in entrata e in uscita necessarie per il corretto funzionamento degli strumenti ONTAP. È importante assicurarsi che solo le porte menzionate nella tabella siano aperte per i collegamenti da macchine remote, mentre tutte le altre porte devono essere bloccate per i collegamenti da macchine remote. In questo modo si garantisce la sicurezza e la sicurezza del sistema.

La seguente tabella descrive i dettagli della porta aperta.

Porta TCP v4/V6 #	Direzione	Funzione
8143	in entrata	Connessioni HTTPS per API REST
8043	in entrata	Connessioni HTTPS
9060	in entrata	Connessioni HTTPS Utilizzato per connessioni SOAP su HTTPS Questa porta deve essere aperta per consentire a un client di connettersi al server API degli strumenti ONTAP.
22	in entrata	SSH (Disattivato per impostazione predefinita)
9080	in entrata	Connessioni HTTPS - VP e SRA - connessioni interne solo da loopback
9083	in entrata	Connessioni HTTPS - VP e SRA Utilizzato per le connessioni SOAP su HTTPS
1162	in entrata	Pacchetti di trap SNMP VP
8443	in entrata	Plugin remoto
1527	solo interno	Porta del database Derby, solo tra questo computer e se stesso, connessioni esterne non accettate — solo connessioni interne
8150	solo interno	Il servizio integrità registro viene eseguito sulla porta
443	bidirezionale	Utilizzato per le connessioni ai cluster ONTAP

Controllo dell'accesso remoto al database Derby

Gli amministratori possono accedere al database derby con i seguenti comandi. È possibile accedervi tramite la VM locale degli strumenti ONTAP e un server remoto con i seguenti passaggi:

```
java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
connect 'jdbc:derby://<OTV-
IP>:1527//opt/netapp/vpserver/vvoldb;user=<user>;password=<password>';
```

esempio:

```
root@UnifiedVSC:~# java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
ij version 10.15
ij> connect 'jdbc:derby://localhost:1527//opt/netapp/vpserver/vvoldb;user=app;password=[redacted]';
ij> show tables;
TABLE_SCHEM         |TABLE_NAME                    |REMARKS
------------------------------------------------------------------------
SYS                 |SYSALIASES                    |
SYS                 |SYSCHECKS                     |
SYS                 |SYSCOLPERMS                   |
SYS                 |SYSCOLUMNS                    |
SYS                 |SYSCONGLOMERATES              |
SYS                 |SYSCONSTRAINTS                |
SYS                 |SYSDEPENDS                    |
SYS                 |SYSFILES                      |
SYS                 |SYSFOREIGNKEYS                |
SYS                 |SYSKEYS                       |
SYS                 |SYSPERMS                      |
```

Tool ONTAP per access point VMware vSphere 9,13 (utenti)

L'installazione di ONTAP Tools per VMware vSphere consente di creare e utilizzare tre tipi di utenti:

1. System User (utente di sistema): L'account utente root
2. Utente dell'applicazione: Gli account utente amministratore, utente principale e utente di database
3. Support user: L'account utente diag

1. Utente di sistema

L'utente System(root) viene creato dall'installazione degli strumenti ONTAP sul sistema operativo sottostante (Debian).

- Un utente di sistema predefinito "root" viene creato su Debian tramite l'installazione degli strumenti ONTAP. L'impostazione predefinita è disattivata e può essere attivata ad hoc tramite la console 'Maint'.

2. Utente dell'applicazione

L'utente dell'applicazione viene denominato come utente locale negli strumenti di ONTAP. Si tratta di utenti creati nell'applicazione ONTAP Tools. Nella tabella seguente sono elencati i tipi di utenti dell'applicazione:

Utente	Descrizione
Administrator User (utente amministratore)	Viene creato durante l'installazione degli strumenti di ONTAP e l'utente fornisce le credenziali durante la distribuzione degli strumenti di ONTAP. Gli utenti hanno la possibilità di modificare la 'password' nella console 'Mainta'. La password scadrà tra 90 giorni e gli utenti saranno tenuti a cambiarla.
Utente manutenzione	Viene creato durante l'installazione degli strumenti di ONTAP e l'utente fornisce le credenziali durante la distribuzione degli strumenti di ONTAP. Gli utenti hanno la possibilità di modificare la 'password' nella console 'Mainta'. Si tratta di un utente addetto alla manutenzione che viene creato per eseguire le operazioni della console di manutenzione.
Utente database	Viene creato durante l'installazione degli strumenti di ONTAP e l'utente fornisce le credenziali durante la distribuzione degli strumenti di ONTAP. Gli utenti hanno la possibilità di modificare la 'password' nella console 'Mainta'. La password scadrà tra 90 giorni e gli utenti saranno tenuti a cambiarla.

3. Utente di assistenza (utente diag)

Durante l'installazione di ONTAP Tools, viene creato un utente di supporto. Questo utente può essere utilizzato per accedere agli strumenti ONTAP in caso di problemi o interruzioni del server e per raccogliere i registri. Per impostazione predefinita, questo utente è disattivato, ma può essere attivato su base adhoc tramite la console 'Maint'. È importante notare che l'utente verrà disattivato automaticamente dopo un determinato periodo di tempo.

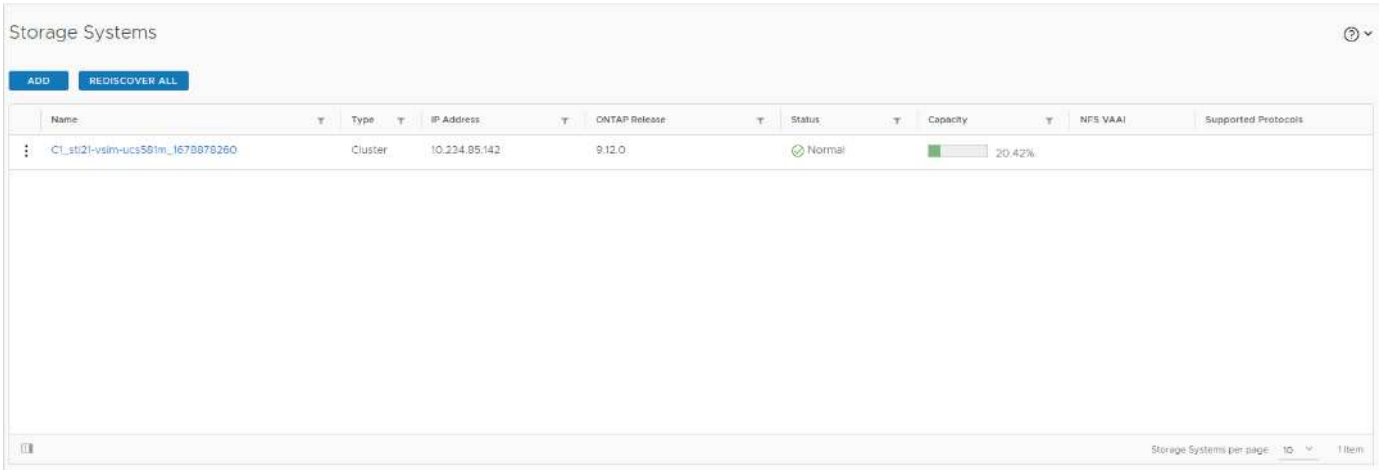
ONTAP tools 9,13 Mutual TLS (autenticazione basata su certificato)

Le versioni ONTAP 9,7 e successive supportano la comunicazione mutua TLS. A partire dai tool ONTAP per VMware e vSphere 9,12, il TLS reciproco viene utilizzato per la comunicazione con i cluster appena aggiunti (in base alla versione di ONTAP).

ONTAP

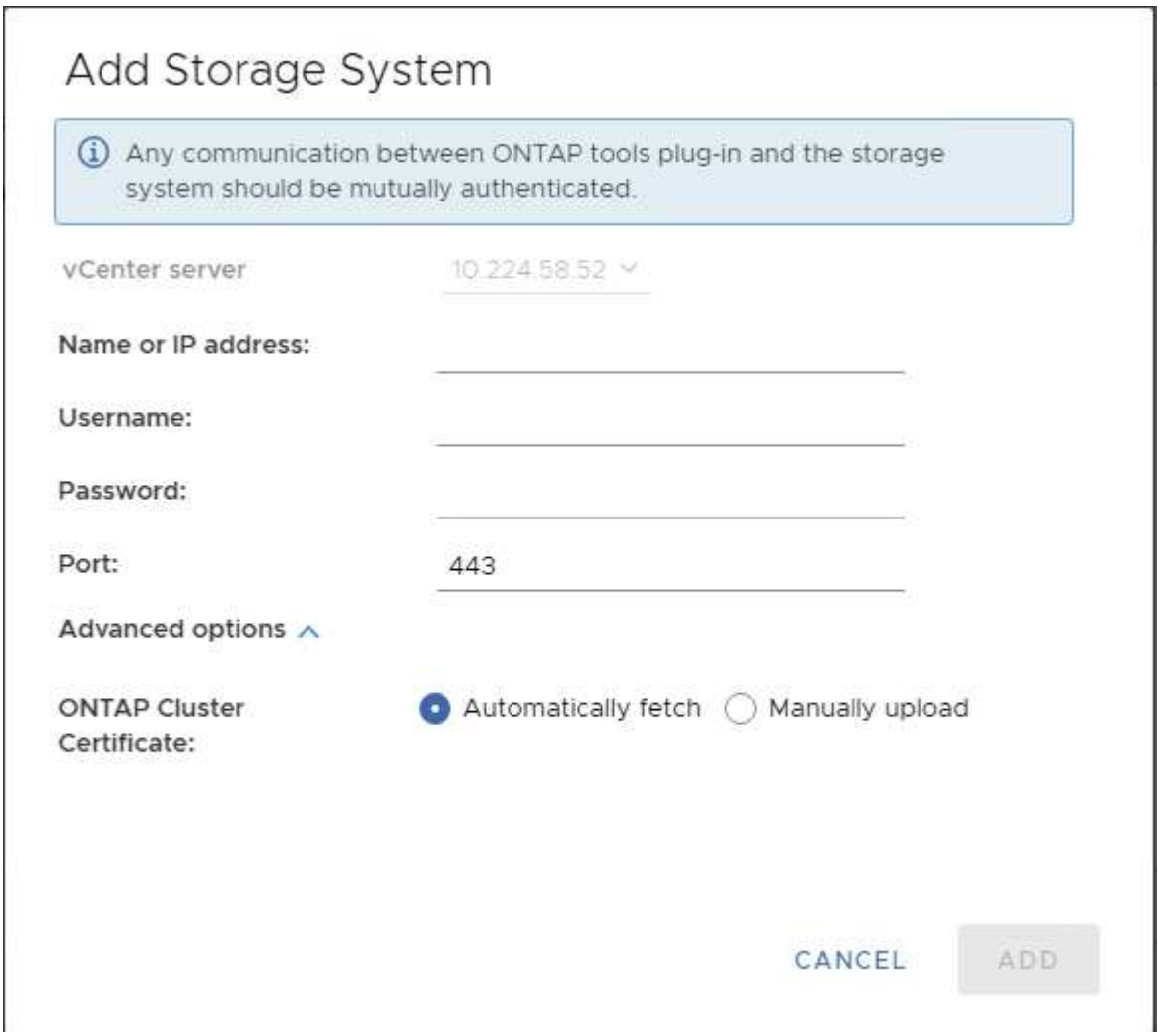
Per tutti i sistemi storage aggiunti in precedenza: Durante un aggiornamento, tutti i sistemi storage aggiunti diventeranno automaticamente attendibili e verranno configurati i meccanismi di autenticazione basati su certificato.

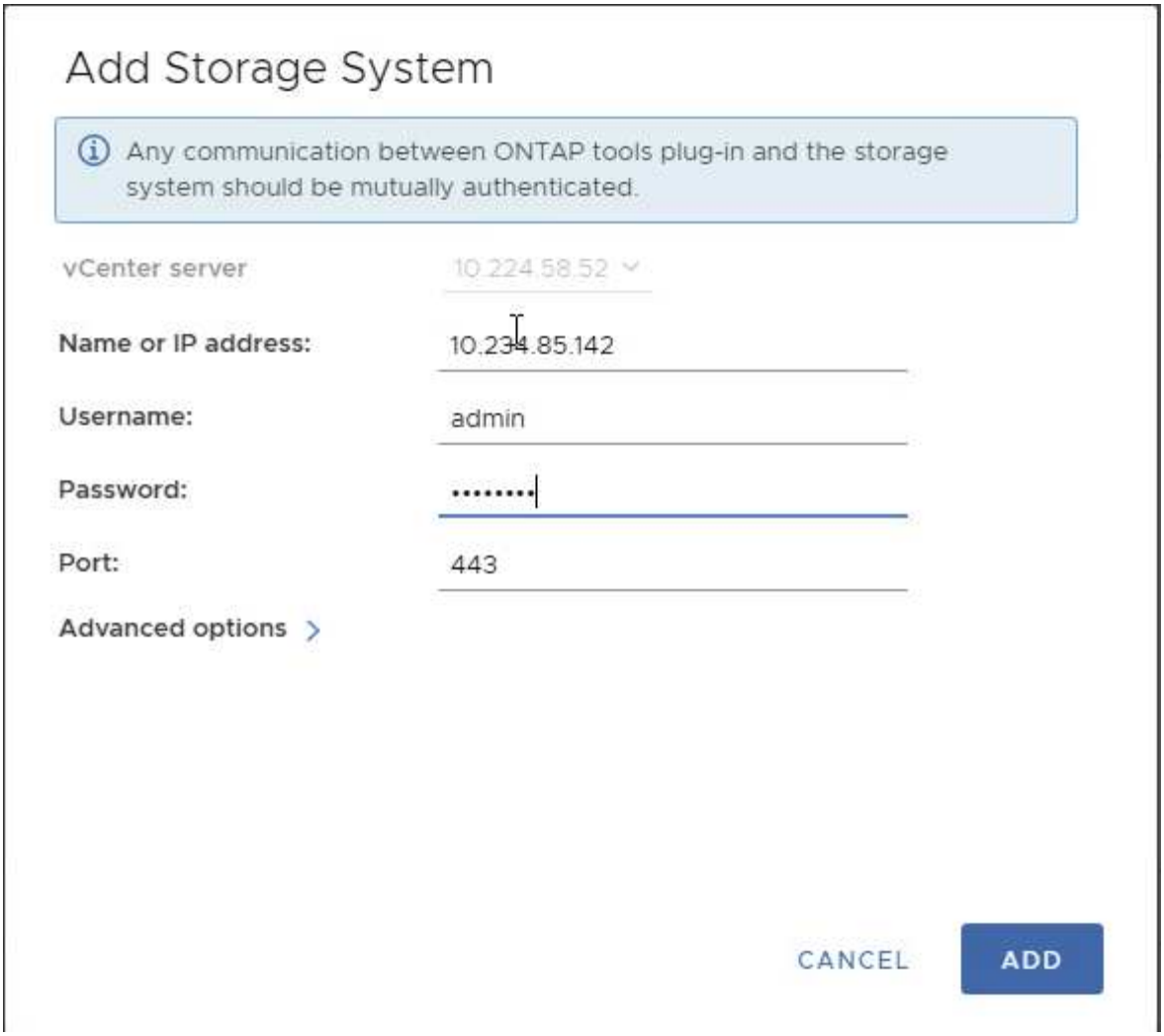
Come nella schermata riportata di seguito, nella pagina di configurazione del cluster viene visualizzato lo stato di Mutual TLS (autenticazione basata su certificato), configurato per ciascun cluster.

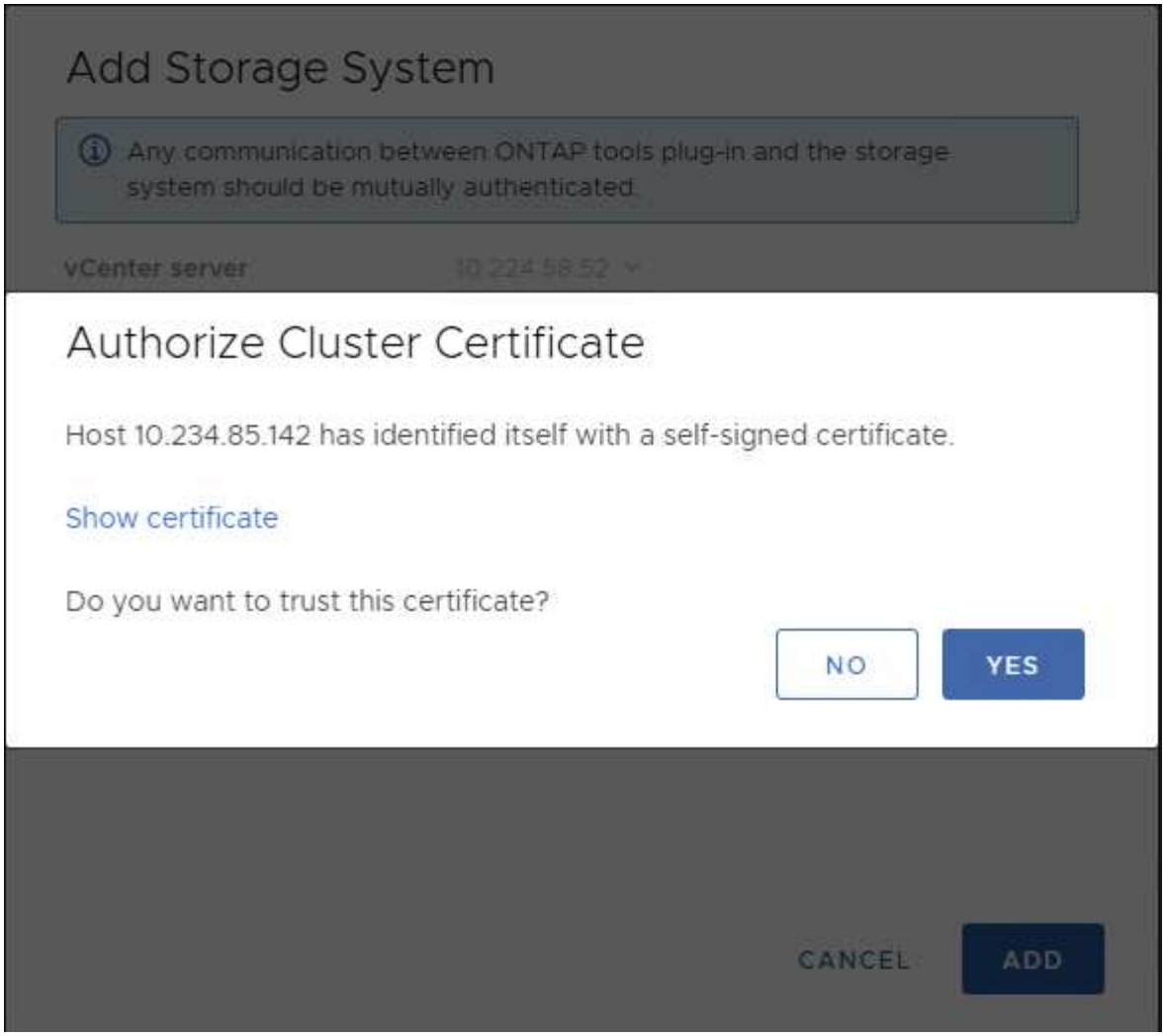


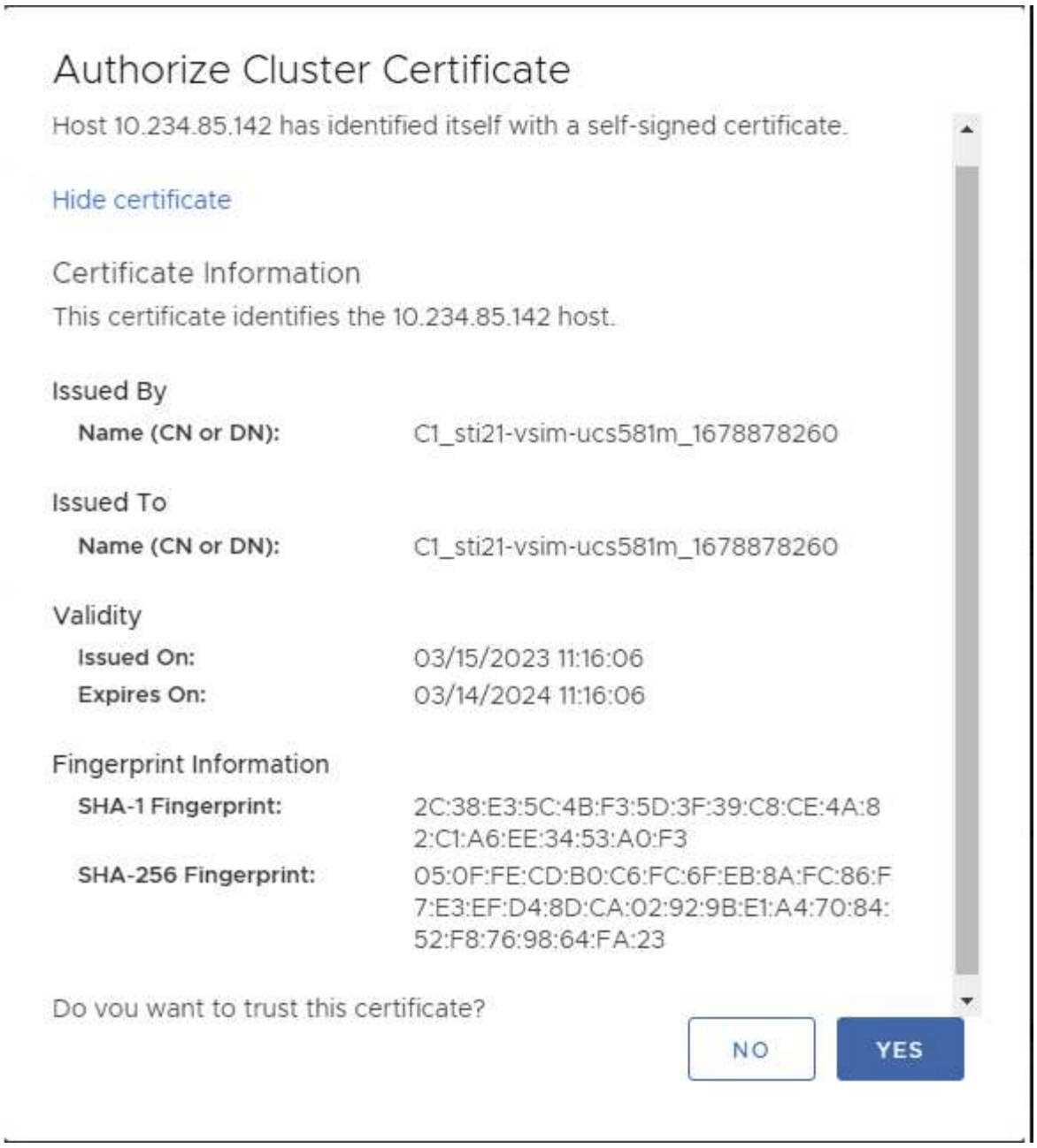
Aggiunta cluster

Durante il flusso di lavoro di aggiunta del cluster, se il cluster che viene aggiunto supporta MTLS, MTLS verrà configurato per impostazione predefinita. L'utente non deve eseguire alcuna configurazione per questo. La schermata riportata di seguito mostra la schermata presentata all'utente durante l'aggiunta del cluster.




Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
10.224.58.52
Name or IP address:
10.234.85.142
Username:
admin
Password:
Port:
443
Advanced options
CANCEL
ADD

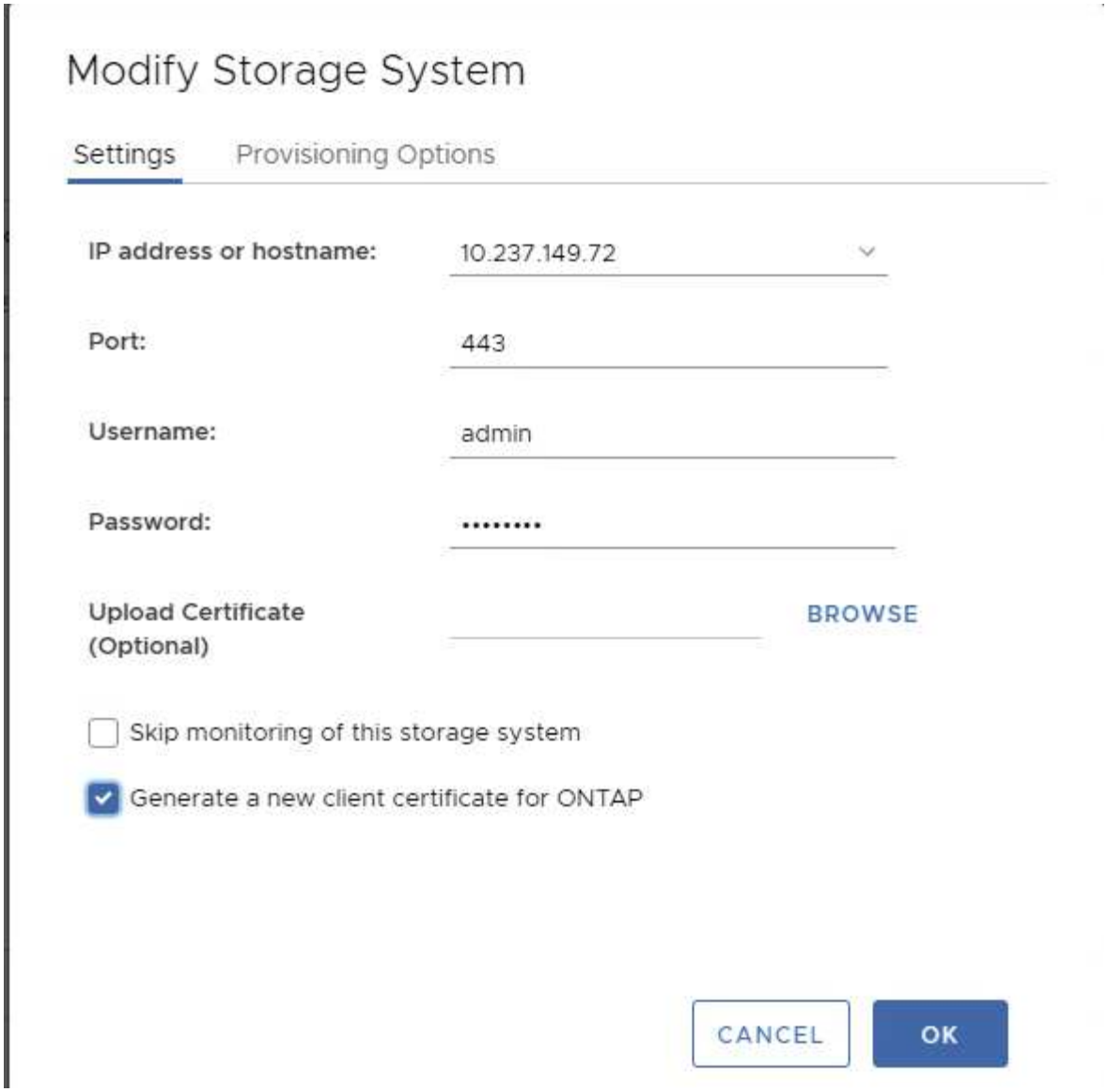

Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
Authorize Cluster Certificate
Host 10.234.85.142 has identified itself with a self-signed certificate.
Show certificate
Do you want to trust this certificate?
NO
YES
CANCEL
ADD

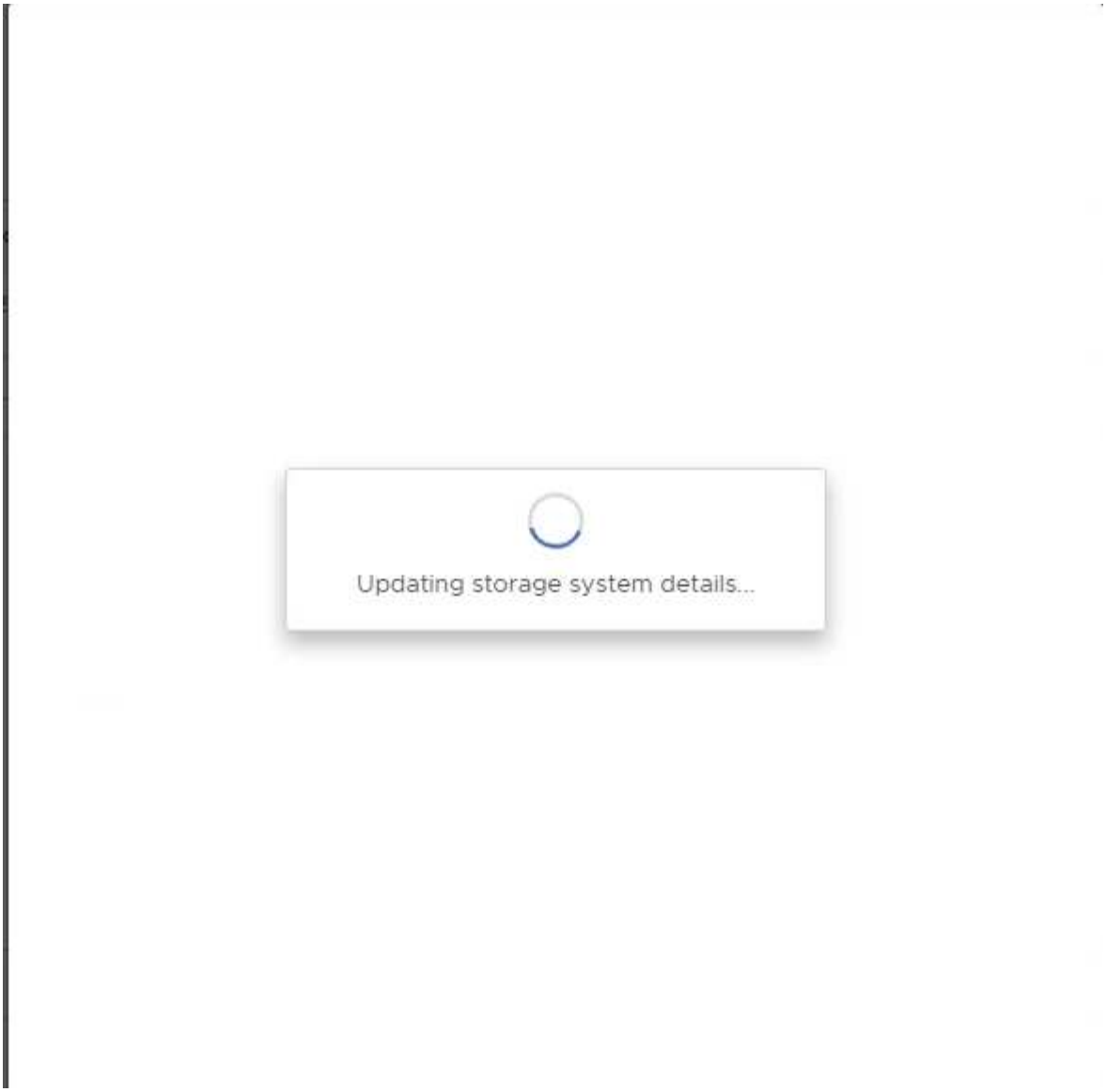


Cluster Edit (Modifica cluster)

Durante l'operazione di modifica del cluster, esistono due scenari:

- Se il certificato ONTAP scade, l'utente dovrà ottenere il nuovo certificato e caricarlo.
- Se il certificato OTV scade, l'utente può rigenerarlo selezionando la casella di controllo.
  - *Genera un nuovo certificato client per ONTAP.*


Modify Storage System
Settings
Provisioning Options
IP address or hostname:
10.237.149.72
Port:
443
Username:
admin
Password:
Upload Certificate (Optional)
BROWSE
Skip monitoring of this storage system
Generate a new client certificate for ONTAP
CANCEL
OK



ONTAP tools 9,13 certificato HTTPS

Per impostazione predefinita, gli strumenti ONTAP utilizzano un certificato autofirmato creato automaticamente durante l'installazione per proteggere l'accesso HTTPS all'interfaccia utente Web. Gli strumenti ONTAP offrono le seguenti funzionalità:

1. Rigenerare il certificato HTTPS

Durante l'installazione degli strumenti ONTAP, viene installato un certificato CA HTTPS e il certificato viene memorizzato nell'archivio chiavi. L'utente può rigenerare il certificato HTTPS tramite la console principale.

È possibile accedere alle opzioni sopra riportate nella console *maint* accedendo a *'Configurazione applicazione' → 'rigenerare certificati'.*

Banner di accesso di ONTAP Tools 9,13

Il seguente banner di accesso viene visualizzato dopo che l'utente ha immesso un nome utente nel prompt di accesso. Tenere presente che SSH è disattivato per impostazione

predefinita e consente l'accesso una tantum solo se attivato dalla console VM.

```
WARNING: Unauthorized access to this system is forbidden and will be
prosecuted by law. By accessing this system, you agree that your actions
may be monitored if unauthorized usage is suspected.
```

Una volta completato l'accesso tramite il canale SSH, viene visualizzato il seguente testo:

```
Linux UnifiedVSC 5.10.0-21-amd64 #1 SMP Debian 5.10.162-1 (2023-01-21)
x86_64
```

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
```

```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

Timeout di inattività per gli strumenti ONTAP 9,13

Per impedire l'accesso non autorizzato, viene impostato un timeout di inattività che disconnette automaticamente gli utenti inattivi per un determinato periodo di tempo durante l'utilizzo di risorse autorizzate. In questo modo, solo gli utenti autorizzati possono accedere alle risorse e mantenere la sicurezza.

- Per impostazione predefinita, le sessioni del client vSphere si chiudono dopo 120 minuti di inattività, richiedendo all'utente di accedere nuovamente per riprendere a utilizzare il client. È possibile modificare il valore di timeout modificando il file webclient.properties. È possibile configurare il timeout del client vSphere "Configurare il valore di timeout del client vSphere"
- Gli strumenti ONTAP hanno un tempo di disconnessione della sessione Web-cli di 30 minuti.

Numero massimo di richieste simultanee per utente (protezione di rete/attacco DOS) Strumenti ONTAP per VMware vSphere 9,13

Per impostazione predefinita, il numero massimo di richieste simultanee per utente è 48. L'utente root negli strumenti ONTAP può modificare questo valore in base ai requisiti del proprio ambiente. **Questo valore non deve essere impostato su un valore molto alto in quanto fornisce un meccanismo contro gli attacchi DOS (Denial of Service).**

Gli utenti possono modificare il numero massimo di sessioni simultanee e altri parametri supportati nel file ***/opt/netapp/vscserver/etc/dosfilterParams.json***.

Possiamo configurare il filtro con i seguenti parametri :

- ***delayMS***: Il ritardo in millisecondi dato a tutte le richieste oltre il limite di velocità prima che vengano prese in considerazione. Dare -1 per respingere la richiesta.
- ***throttleMS***: Per quanto tempo attendere il semaforo in modalità asincrona.
- ***maxRequestMS***: Per quanto tempo consentire l'esecuzione di questa richiesta.
- ***ipWhitelist***: Un elenco separato da virgole di indirizzi IP che non saranno limitati dalla velocità. (Possono essere indirizzi IP vCenter, ESXi e SRA)
- ***maxRequestsPerSec***: Il numero massimo di richieste da una connessione al secondo.

Valori predefiniti nel file *dosfilterParams*:

```
{"delayMs":"-1",
"throttleMs":"1800000",
"maxRequestMs":"300000",
"ipWhitelist":"10.224.58.52",
"maxRequestsPerSec":"48"}
```

Configurazione del protocollo NTP (Network Time Protocol) per gli strumenti ONTAP 9,13

A volte, possono verificarsi problemi di protezione dovuti a discrepanze nelle configurazioni dell'ora di rete. È importante assicurarsi che tutti i dispositivi all'interno di una rete dispongano di impostazioni dell'ora precise per evitare tali problemi.

Virtual appliance

È possibile configurare i server NTP dalla console di manutenzione dell'appliance virtuale. Gli utenti possono aggiungere i dettagli del server NTP in *System Configuration* ⇒ *Add new NTP Server* option

Per impostazione predefinita, il servizio per NTP è ntpd. Si tratta di un servizio legacy che in alcuni casi non funziona bene per le macchine virtuali.

Debian

Su Debian, l'utente può accedere al file /etc/ntp.conf per i dettagli del server ntp.

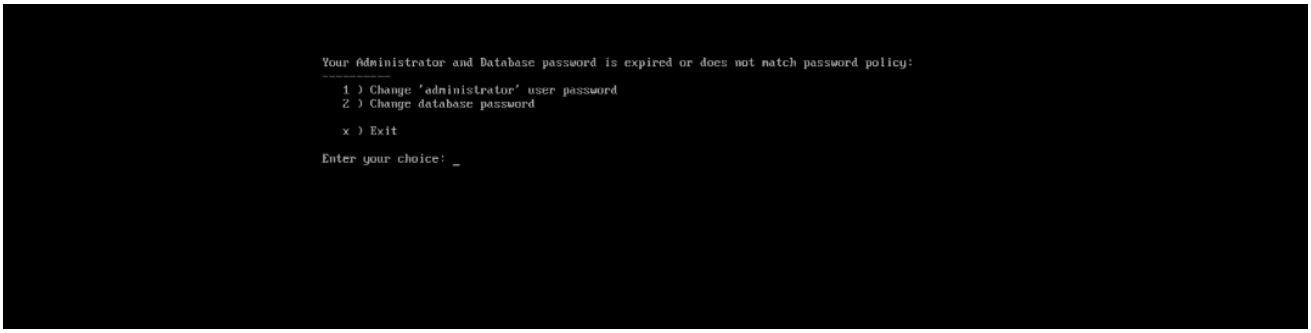
Criteri delle password per gli strumenti ONTAP 9,13

Gli utenti che distribuiscono gli strumenti ONTAP per la prima volta o che eseguono l'aggiornamento alla versione 9,12 o successiva dovranno seguire il criterio password complessa sia per gli utenti dell'amministratore che per quelli del database. Durante il processo di distribuzione, ai nuovi utenti verrà richiesto di immettere le password. Per gli utenti di brownfield che effettuano l'aggiornamento alla versione 9,12 o successiva, l'opzione per seguire il criterio password complessa sarà disponibile nella console di manutenzione.

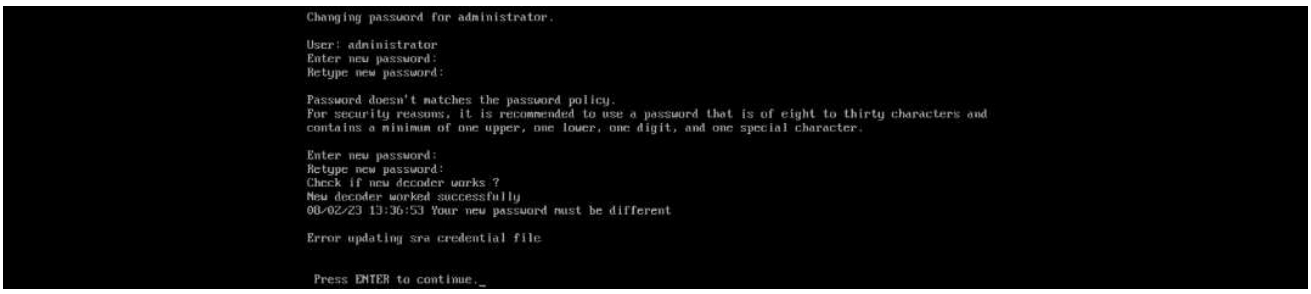
- Una volta che l'utente accede alla console principale, le password verranno controllate in base al set di regole complesso e, se risulta non essere seguite, all'utente verrà chiesto di reimpostare lo stesso.

- La validità predefinita della password è di 90 giorni e dopo 75 giorni l'utente inizierà a ricevere la notifica di modifica della password.
- È necessario impostare una nuova password ad ogni ciclo; il sistema non utilizzerà l'ultima password come nuova password.
- Ogni volta che un utente accede alla console principale, prima di caricare il menu principale controlla i criteri delle password, come le schermate seguenti:

- Se non viene rilevato seguendo il criterio password o la relativa configurazione di aggiornamento da ONTAP Tools 9,11 o precedenti. L'utente visualizzerà quindi la seguente schermata per reimpostare la password:



- Se l'utente tenta di impostare una password debole o restituisce l'ultima password, viene visualizzato il seguente errore:



Note legali

Le note legali forniscono l'accesso a dichiarazioni di copyright, marchi, brevetti e altro ancora.

Copyright

"https://www.netapp.com/company/legal/copyright/"

Marchi

NETAPP, il logo NETAPP e i marchi elencati nella pagina dei marchi NetApp sono marchi di NetApp, Inc. Altri nomi di società e prodotti potrebbero essere marchi dei rispettivi proprietari.

"https://www.netapp.com/company/legal/trademarks/"

Brevetti

Un elenco aggiornato dei brevetti di proprietà di NetApp è disponibile all'indirizzo:

https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf

Direttiva sulla privacy

"https://www.netapp.com/company/legal/privacy-policy/"

Open source

I file di avviso forniscono informazioni sul copyright e sulle licenze di terze parti utilizzate nel software NetApp.

ONTAP

"Avviso per ONTAP 9.16.1" "Avviso per ONTAP 9.16.0" "Avviso per ONTAP 9.15.1" "Avviso per ONTAP 9.15.0" "Avviso per ONTAP 9.14.1" "Avviso per ONTAP 9.14.0" "Avviso per ONTAP 9.13.1" "Avviso per ONTAP 9.12.1" "Avviso per ONTAP 9.12.0" "Avviso per ONTAP 9.11.1" "Avviso per ONTAP 9.10.1" "Avviso per ONTAP 9.10.0" "Avviso per ONTAP 9.9.1" "Avviso per ONTAP 9.8" "Avviso per ONTAP 9,7" "Avviso per ONTAP 9,6" "Avviso per ONTAP 9,5" "Avviso per ONTAP 9,4" "Avviso per ONTAP 9,3" "Avviso per ONTAP 9,2" "Avviso per ONTAP 9,1"

ONTAP Mediator per configurazioni IP MetroCluster

"9.9.1 Avviso per ONTAP Mediator per le configurazioni IP di MetroCluster" "9,8 Avviso per ONTAP Mediator per le configurazioni IP di MetroCluster" "9,7 Avviso per ONTAP Mediator per le configurazioni IP di MetroCluster"

Informazioni sul copyright

Copyright © 2025 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEQUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina http://www.netapp.com/TM sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.