



ONTAP e aplicações empresariais

Enterprise applications

NetApp

December 17, 2024

Índice

ONTAP e aplicações empresariais	1
Épico	2
Épico em ONTAP	2
Épico em ONTAP	3
Arquitetura e design épicos	9
Configuração e práticas recomendadas	17
Informações adicionais sobre a Epic no ONTAP	23
Hyper-V	24
Diretrizes de implantação e práticas recomendadas de storage	24
Microsoft SQL Server	66
Visão geral	66
Configuração do banco de dados	67
Configuração de armazenamento	74
Proteção de dados	87
Recuperação de desastres	89
Segurança dos dados	111
MySQL	114
Visão geral	114
Configuração do banco de dados	114
Configuração de host	121
Configuração de armazenamento	123
Banco de dados Oracle	127
Bancos de dados Oracle no ONTAP	127
Configuração ONTAP	127
Configuração do banco de dados	139
Configuração de host	143
Configuração de rede	158
Configuração de armazenamento	166
Virtualização	185
Disposição em camadas	188
Proteção de dados Oracle	197
Recuperação de desastres Oracle	221
Migração de banco de dados Oracle	276
Notas adicionais	399
PostgreSQL	409
Visão geral	409
Configuração do banco de dados	409
Configuração de armazenamento	413
Proteção de dados	417
SAP	420
VMware	421
VMware vSphere com ONTAP	421
Volumes virtuais (vVols) com ONTAP	465

VMware Site Recovery Manager com ONTAP	491
Cluster de armazenamento do vSphere Metro com o ONTAP	511
Segurança do produto	542
Guia de fortalecimento da segurança para as ferramentas do ONTAP para VMware vSphere	546
Avisos legais	561
Direitos de autor	561
Marcas comerciais	561
Patentes	561
Política de privacidade	561
Código aberto	561
ONTAP	561
Mediador ONTAP para IP MCC	561

ONTAP e aplicações empresariais

Épico

Épico em ONTAP

A chave da transformação digital é fazer mais com seus dados.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-3923: Melhores práticas da NetApp para a Epic*.

Os hospitais exigem grandes quantidades de dados para iniciar a jornada da transformação digital. Parte do processo de tratamento de pacientes, gerenciamento de horários de funcionários e recursos médicos é que as informações são coletadas e processadas. No entanto, muitas ações ainda são realizadas manualmente ou através de sistemas desatualizados. A única constante é que a quantidade de dados continua crescendo exponencialmente e, assim, se torna cada vez mais difícil de gerenciar.

A principal causa desse problema é que os dados hospitalares geralmente são armazenados em silos de dados. Muito tempo é gasto em entradas manuais e atualizações que levam ao esgotamento e erros. Este documento é sobre uma parte dos dados de saúde, a Epic Electronic Health Records (EHR). No entanto, a estratégia de gerenciamento de dados aqui abordada pode e deve ser aplicada a todos os dados de saúde. A NetApp tem uma história comprovada de modernização e simplificação da infraestrutura digital. Nossa infraestrutura de dados inteligente forma a base da transformação digital.

A NetApp oferece uma solução única de gerenciamento de dados para todas as necessidades de saúde. Além disso, podemos orientar os hospitais em sua jornada rumo à transformação digital. Construindo uma base com estrutura e soluções inteligentes, a saúde pode extrair todo o valor dessas informações preciosas. Essa estrutura pode ajudar os provedores médicos a diagnosticar doenças mais rapidamente e desenvolver planos de tratamento individualizados para melhor apoiar os processos de tomada de decisão em situações de emergência. Você também poderá criar sua própria infraestrutura de dados inteligente e permitir que seu hospital desbloqueie silos de dados, facilite a interoperabilidade dos dados e proteja informações confidenciais dos pacientes.

Use este documento como um guia para construir e implementar com sucesso EPIC EHR. Em vez de criar vários silos da Epic, crie uma única infraestrutura de dados da Epic e transforme seu hospital.

Finalidade

Este documento descreve as práticas recomendadas para a integração do storage do NetApp em um ambiente de software da Epic. Ele contém as seguintes seções:

- Um entendimento técnico do ambiente de software da Epic e seus requisitos de storage em várias configurações.
- Considerações épicas de storage, que descrevem fatores importantes para a tomada de decisões das soluções da Epic.
- Recomendações de storage da NetApp, que descrevem as práticas recomendadas de configuração de storage da NetApp para atender aos requisitos de storage da Epic.

Âmbito de aplicação

Este documento não abrange os seguintes assuntos:

- Requisitos quantitativos de desempenho e orientação de dimensionamento, que são abordados em "[TR-](#)

Público-alvo

O NetApp assume que o leitor tem os seguintes conhecimentos:

- Uma sólida compreensão dos conceitos de SAN e nas
- Familiaridade técnica com os sistemas de storage da ONTAP
- Familiaridade técnica com a configuração e administração do ONTAP

Épico em ONTAP

ÉPICO em ONTAP

O Epic é mais fácil com o ONTAP.

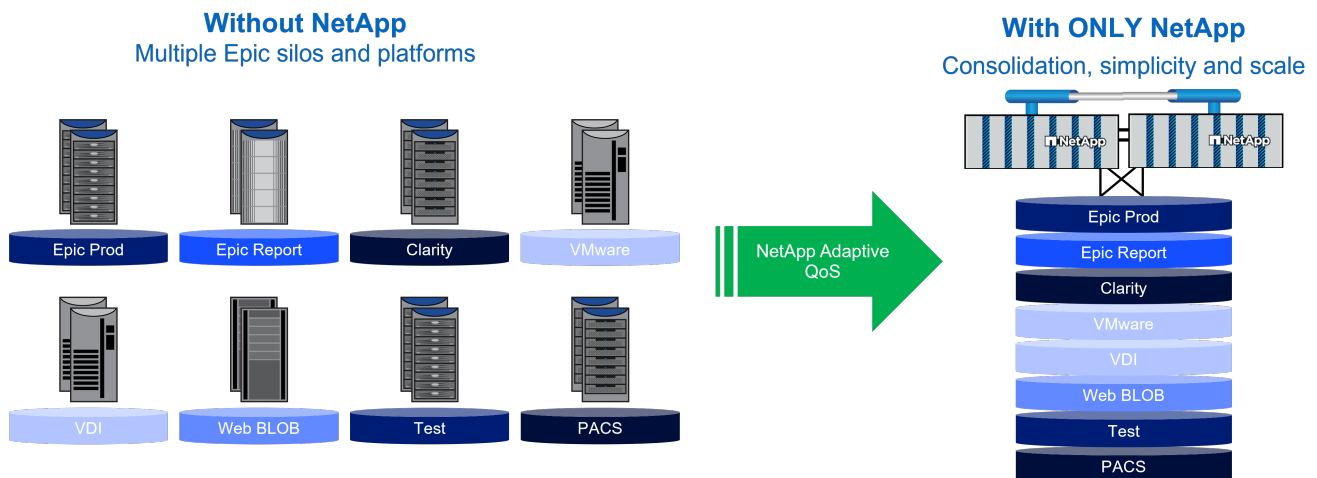
O ONTAP é uma plataforma de gerenciamento de dados que permite consolidar workloads da Epic enquanto atende a todos os requisitos de performance, proteção de dados e gerenciamento de dados.

Somente na NetApp você padroniza todos os workloads do setor de saúde para SAN, nas e objetos em uma única plataforma de gerenciamento de dados de alta disponibilidade. O ONTAP é a plataforma de software de storage mais implantada no mundo e vem com quase 30 anos de inovação constante. Você pode enfrentar todos os desafios da Epic com ferramentas nativas de gerenciamento de dados da ONTAP e integração de aplicações. Não há necessidade de comprar uma infinidade de ferramentas de terceiros para preencher lacunas na solução.

Muitos fornecedores de storage oferecem storage de bloco tradicional, confiável e rápido. Eles funcionam bem, mas normalmente são implantados em silos para executar um único workload, como produção, relatório, Clarity, VDI, VMware e nas. Cada um desses silos tem hardware diferente e ferramentas de gerenciamento diferentes, e eles geralmente são gerenciados por diferentes grupos DE TI. Esta abordagem tradicional adiciona ao maior problema com a saúde hoje em dia - complexidade.

Com o NetApp, o gerenciamento de dados fica mais fácil e eficiente. Em vez de gastar dinheiro com o problema com silos superdimensionados, a ONTAP usa inovação e tecnologia para permitir um SLA consistente e garantido para cada workload em uma única plataforma em qualquer protocolo com proteção de dados integrada. Esses recursos e ferramentas também se estendem para a nuvem de sua escolha, conforme ilustrado abaixo.

Scale and simplicity for Healthcare



Disponibilidade épica em ONTAP

No centro da ONTAP estão operações ininterruptas que permitem a você evitar interrupções dispendiosas nas operações de negócios.

O NetApp oferece disponibilidade de mais de 99,99,999999% com base em dados de produção, que são chamados de casa por meio do NetApp Active IQ. Cada par de HA no cluster não tem um ponto único de falha. O ONTAP data de 1992 e é o software de gerenciamento de dados mais implantado do mundo, com um histórico excepcional de fornecimento de storage confiável. Agora, com o monitoramento proativo do Active IQ e a resolução automática de 97% dos problemas, a disponibilidade aumenta com bastante menos casos de suporte.

A Epic recomenda o uso de sistemas de storage de HA para reduzir a falha de componentes de hardware. Esta recomendação se estende do hardware básico (como fontes de alimentação redundantes) à rede (como rede multipath).

Quando você precisa atualizar o storage, fazer escalabilidade vertical, fazer escalabilidade horizontal ou rebalancear workloads no cluster, não há efeito no atendimento aos pacientes. Você pode mover dados, mas nunca mais precisa interromper o atendimento ao paciente com migrações de dados ou atualizações em massa. Passe para a tecnologia de nova geração, pronta para o futuro e evite o aprisionamento tecnológico. O NetApp ainda oferece uma garantia de disponibilidade por escrito de 100%.

Mais informações sobre confiabilidade, disponibilidade, capacidade de manutenção e recursos de segurança da NetApp podem ser encontradas no ["Confiabilidade, disponibilidade, capacidade de manutenção e segurança da NetApp ONTAP"](#) white paper.

Épico na consolidação de ONTAP

Um dos principais desafios da saúde é a ineficiência dos ambientes em silos.

Soluções de múltiplos pontos são criadas por vários grupos que impedem o progresso. Ter uma estratégia unificada para o gerenciamento de dados traz eficiência para acelerar a transformação. Tecnologia disruptiva, como a digitalização de Registros de pacientes, ransomware e IA generativa, impulsionam a necessidade de consolidação.

Com o ONTAP, você consolida arquivo/bloco/objeto e cada um dos workloads de categoria 0/1/2/3, no local e na nuvem, todos executados no ONTAP.

Épica em eficiência de ONTAP

A Epic é executada em all-flash arrays em que a maior parte do custo é o disco. Portanto, a eficiência de storage é essencial para a economia de custos.

A eficiência de storage in-line da NetApp alcança economias líderes do setor em storage sem afetar a performance. Além disso, oferecemos uma garantia de eficiência por escrito com os all-flash arrays.

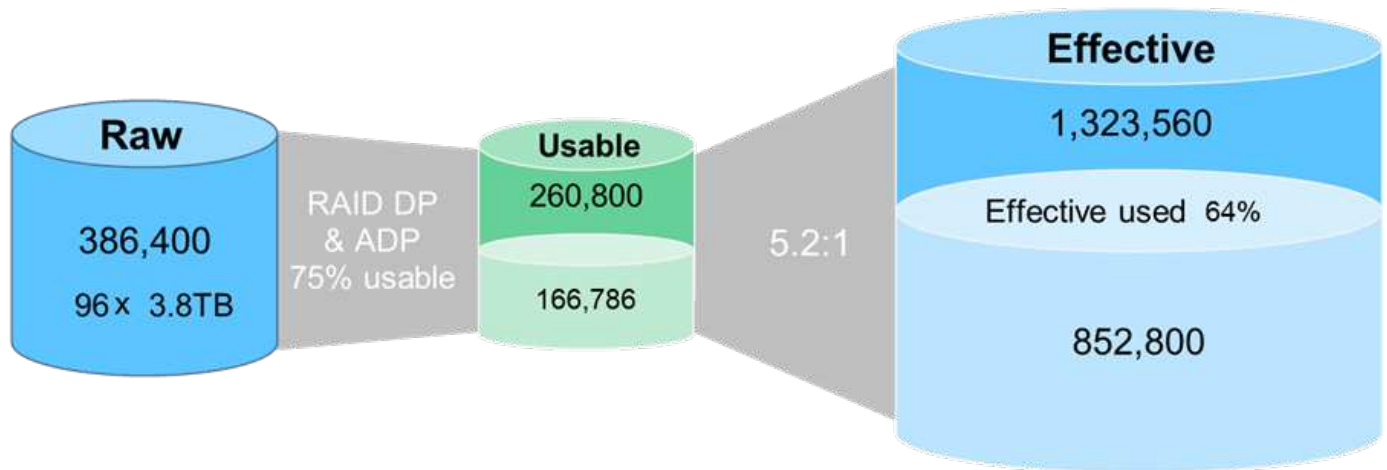
Ao calcular a eficiência de storage, é importante medir o bruto para utilizável para capacidade efetiva.

- * Capacidade bruta* antes de qualquer RAID ser aplicado, tamanho do disco pelo número de discos.
- **Capacidade utilizável** após a aplicação do RAID, quanto armazenamento utilizável está disponível.
- **Capacidade efetiva** quanto armazenamento é provisionado e apresentado ao host ou cliente.

A figura abaixo mostra um exemplo de cálculo de eficiência de uma implantação típica da Epic, incluindo todos os workloads que exigem 852TB TB de storage efetivo e com eficiência de 5,2:1, fornecendo 1,32PB TB de dados efetivos totais.



Com base no número de discos, a capacidade bruta para utilizável varia ligeiramente.



A NetApp não usa a tecnologia NetApp Snapshot nem o thin Provisioning para calcular a eficiência no programa de garantia. Fazer isso mostraria eficiências irrealistas de 30 a 100:1, o que não significa nada ao dimensionar a capacidade de storage do mundo real.

Desempenho épico em ONTAP

A ONTAP apresentou as tecnologias flash em 2009 e oferece suporte para SSDs desde 2010. Essa longa experiência com storage flash permite que a NetApp ajuste os recursos do ONTAP para otimizar o desempenho da SSD e aprimorar a resistência da Mídia flash, mantendo os recursos avançados do ONTAP.

Desde o ano 2020, todos os workloads da Epic ODB precisam estar no storage all-flash. As cargas de trabalho épicas normalmente operam em aproximadamente 1.000 a 2.000 IOPs por terabyte de storage (8k

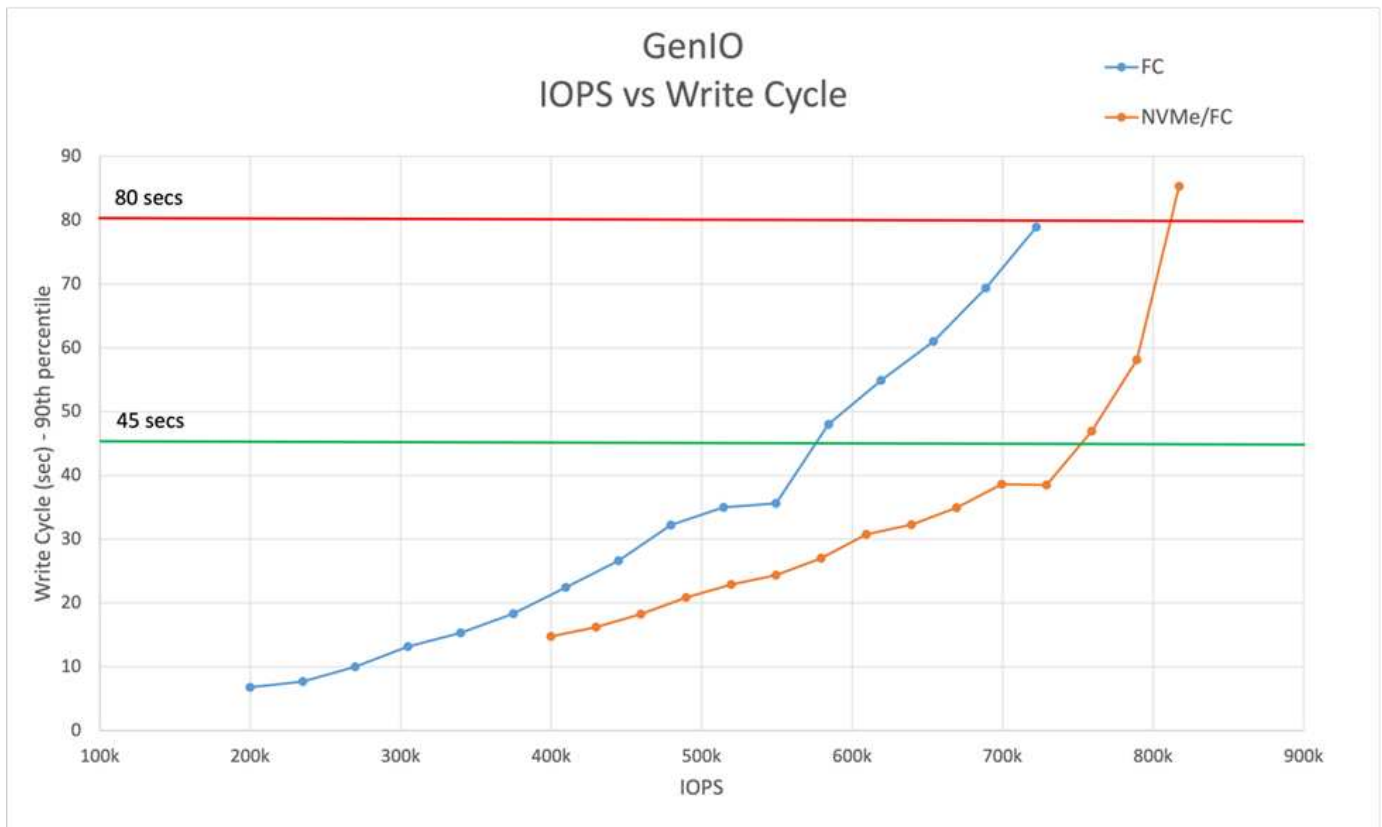
bloco, 75%/25% de taxa de leitura e gravação e 100% de forma aleatória). O Epic é muito sensível à latência, e a alta latência tem um efeito visível na experiência do usuário final, bem como em tarefas operacionais, como execução de relatórios, backup, verificações de integridade e tempos de atualização do ambiente.

- O fator limitante para all-flash arrays não são as unidades, mas sim a utilização nas controladoras.
- O ONTAP usa uma arquitetura ativo-ativo. Para obter performance, os dois nós do par de HA gravam nas unidades.
- Isso resulta na maximização da utilização da CPU, o fator mais importante que permite que a NetApp publique a melhor performance da Epic no setor.
- As tecnologias NetApp RAID DP, Advanced Disk Partitioning (ADP) e WAFL atendem a todos os requisitos da Epic. Todos os workloads distribuem e/S por todos os discos. Sem gargalos.
- O ONTAP é otimizado para gravação; as gravações são confirmadas depois de gravadas no NVRAM espelhado antes de serem gravadas no disco na velocidade de memória inline.
- O WAFL, o NVRAM e a arquitetura modular permitem que o NetApp use o software para inovar com eficiências in-line, criptografia e performance. Eles também permitem que o NetApp apresente novos recursos e funcionalidades sem afetar o desempenho.
- Historicamente, com cada nova versão do ONTAP há um aumento no desempenho e eficiência na faixa de 30-50%. O desempenho é ideal quando você se mantém atualizado com o ONTAP.

NVMe

Quando a performance é fundamental, o NetApp também oferece suporte a NVMe/FC, o protocolo FC SAN de última geração.

Como pode ser visto na figura abaixo, nossos testes do Genio alcançaram um número muito maior de IOPS usando o protocolo NVMe/FC em comparação com o protocolo FC. A solução conectada NVMe/FC alcançou mais de 700k IOPS antes de superar o limite do ciclo de gravação de 45 segundos. Com a substituição de comandos SCSI pelo NVMe, você também reduz significativamente a utilização no host.



Escalabilidade épica para ONTAP

O Guia de configuração de hardware da Epic representa um crescimento de cerca de 20% por ano por 3 anos. No entanto, os ambientes também podem crescer inesperadamente.

O NetApp pode dimensionar a performance e a capacidade de forma otimizada até 12 nós para clusters nas, SAN e objetos. Como resultado, você pode fazer escalabilidade vertical e horizontal sem interrupções à medida que sua empresa cresce.

O Epic Iris oferece recursos adicionais de dimensionamento. Com ele, clientes maiores que têm várias instâncias da Epic podem se consolidar em uma única instância. ["Arquitetura verificada da NetApp na SAN moderna"](#) O documento demonstra que a Epic pode dimensionar workloads consolidados de forma otimizada para 720K IOPS em um único HA e fazer escalabilidade horizontal para mais de 4M IOPS em um cluster. Faça a escalabilidade vertical sem interrupções atualizando controladores ou adicionando discos a clusters existentes.

Os dados nas, SAN e de objetos também podem ser movidos entre nós no cluster sem interrupções. Cada par de HA no cluster pode ser qualquer combinação de tipos e tamanhos de sistemas ONTAP FAS e AFF. Você pode equilibrar seus workloads em um único cluster para maximizar seu investimento em storage.

O ONTAP também oferece a opção de usar o storage de objetos no StorageGRID ou na nuvem como destino de backup e/ou destino automático de disposição em camadas de storage frio. Essa funcionalidade permite que você libere automaticamente discos all-flash caros, snapshots em camadas e dados inativos no Object.

Como resultado, a Epic simplesmente funciona melhor com o portfólio de produtos da NetApp, utilizando ONTAP, vários protocolos, StorageGRID e a nuvem de sua preferência. Esses produtos oferecem opções para recuperação de desastres, arquivamento, análises, disposição em camadas e muito mais.

Configuração de eficiência de storage épica

Um snapshot é uma cópia pontual de um volume que é somente leitura.

Um snapshot coloca um bloqueio lógico em todos os blocos no sistema de arquivos ativo. As cópias Snapshot do NetApp ONTAP são quase instantâneas e não usam storage adicional.

Write Anywhere File Layout, ou WAFL, é um sistema de arquivos somente de gravação; ele não executa IO adicional, como copiar os dados em um bloco protegido por snapshot antes de ser substituído. Nenhum dado é movido. Portanto, os snapshots não afetam a capacidade ou a performance do storage. Os snapshots oferecem uma grande economia de storage enquanto aumentam a solução de backup.

FlexClone

Um volume NetApp ONTAP FlexClone é um clone de um volume existente ou um snapshot de um volume existente. De outra forma, ele é um volume ONTAP como qualquer outro e pode ser clonado, protegido com snapshots e configurado com uma política de QoS.

Tal como acontece com snapshots, um volume FlexClone não requer nenhum espaço adicional no momento da criação. Somente as alterações no clone exigem capacidade adicional.

A Epic exige de 10 a 30 cópias dos bancos de dados de produção para vários requisitos operacionais, como backups de streaming, verificações de integridade e ambientes de atualização de preparação. A necessidade de uma solução baseada no FlexClone volumes aumentou com a migração para atualizações mais frequentes.



O NetApp fornece uma solução de backup Epic totalmente automatizada e uma solução de atualização Epic como parte da solução usando o Ansible e ferramentas nativas do NetApp.

Épico em segurança ONTAP

A segurança é a preocupação número um para as organizações e os executivos de saúde hoje. Nunca foi tão difícil gerenciar, e as organizações desafiam a conformidade, a governança de dados, a proteção antivírus e o ransomware.

Um guia completo para a segurança de storage e da Epic está além do escopo deste documento. No entanto "[Guia de endurecimento de segurança para ONTAP](#)", detalha todos os recursos de segurança avançados e abrangentes disponíveis com o ONTAP.

O NetApp Active IQ Unified Manager monitora violações de segurança com base nas informações incluídas "[TR-4569](#)" e as informa no painel para simplificar o gerenciamento de segurança. Essas ferramentas podem ajudar sua organização a atingir suas metas de segurança para proteger, detectar e corrigir ataques.

A NetApp também fez uma parceria com fornecedores de segurança para fornecer integração por meio "[Política de NetApp](#)" de software para aprimorar sua oferta de segurança. Além disso, "[Autenticação de vários fatores \(MFA\)](#)" é possível adicionar para proteger seu ambiente Epic contra acesso não autorizado com credenciais vazadas.

Por fim, as cópias Snapshot nativas da ONTAP e as tecnologias SnapLock imutáveis com "[Ciber Vault da ONTAP](#)" , oferecem uma capacidade única de lacuna aérea para proteger seus Registros de pacientes contra ransomware. Consulte a documentação do NetApp em "[A solução NetApp para ransomware](#)". Para obter uma abordagem mais estratégica para a segurança, "[NetApp e confiança zero](#)" consulte .

Arquitetura e design épicos

Arquitetura épica

Esta seção descreve o ambiente de software da Epic e os principais componentes que exigem storage. Ele fornece considerações essenciais para ajudar a orientar o design do armazenamento.

A EPIC, com sede em Verona, Wisconsin, faz software para grupos médicos de médio a grande porte, hospitais e organizações de saúde integradas. Os clientes também incluem hospitais comunitários, instalações acadêmicas, organizações infantis, provedores de redes de segurança e sistemas multi-hospitalares. O software integrado ao EPIC abrange funções clínicas, de acesso e de receita e se estende para casa.

Está além do escopo deste documento cobrir a ampla gama de funções suportadas pelo software Epic. No entanto, do ponto de vista do sistema de storage, todos os softwares da Epic compartilham um único banco de dados centrado em pacientes para cada implantação. A Epic está migrando do banco de dados InterSystems Caché para o novo banco de dados InterSystems Iris. Como os requisitos de armazenamento são os mesmos para Caché e Iris, vamos nos referir ao banco de dados como Iris em todo o resto deste documento. O Iris está disponível para os sistemas operacionais AIX e Linux.

Intersistemas Iris

InterSystems Iris é o banco de dados usado pelo aplicativo Epic. Neste banco de dados, o servidor de dados é o ponto de acesso para dados armazenados persistentemente. O servidor de aplicativos gerencia consultas de banco de dados e faz solicitações de dados para o servidor de dados. Para a maioria dos ambientes de software da Epic, o uso da arquitetura de multiprocessadores simétricos (SMP) em um único servidor de banco de dados é suficiente para atender às solicitações de banco de dados dos aplicativos da Epic. Em grandes implantações, um modelo distribuído pode ser suportado usando o Enterprise Caché Protocol (ECP) da InterSystems.

O uso de hardware em cluster habilitado para failover permite que um servidor de dados em espera acesse o mesmo armazenamento que o servidor de dados primário. Ele também permite que o servidor de dados em espera assuma responsabilidades de processamento durante uma falha de hardware.

A InterSystems também fornece tecnologias para atender aos requisitos de replicação de dados, recuperação de desastres e alta disponibilidade (HA). A tecnologia de replicação da InterSystems é usada para replicar um banco de dados Iris de forma síncrona ou assíncrona de um servidor de dados primário para um ou mais servidores de dados secundários. O NetApp SnapMirror é usado para replicar o armazenamento WebBLOB ou para backup e recuperação de desastres.

O banco de dados atualizado do Iris tem muitas vantagens:

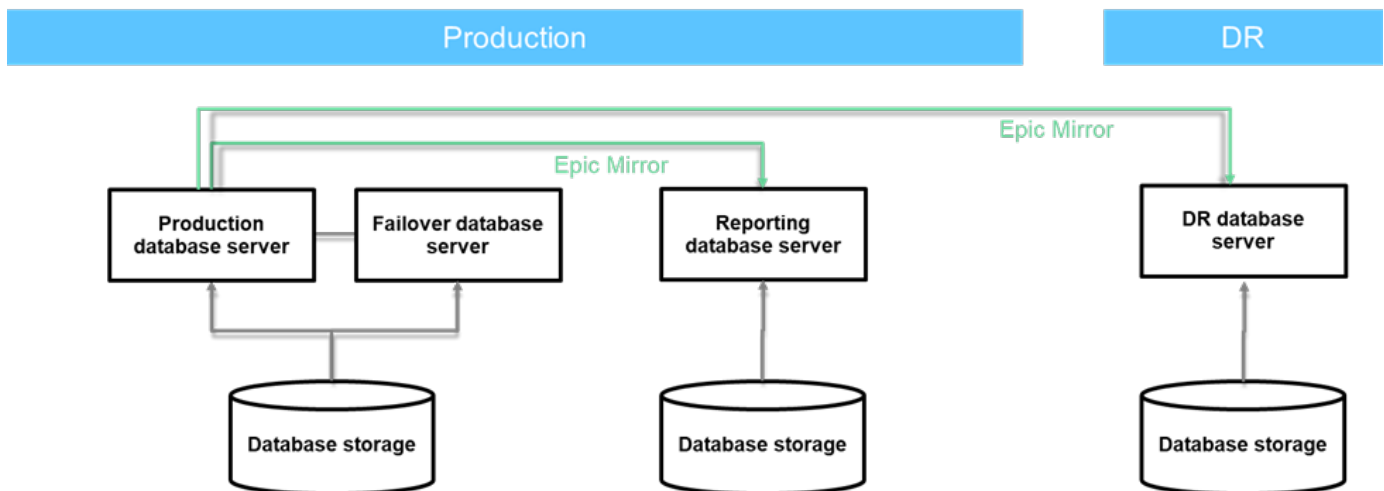
- Aumentou a escala e permite que organizações maiores, com várias instâncias da Epic, se consolidem em uma instância maior.
- Um feriado de licenciamento onde os clientes podem agora se mover entre AIX e Red Hat Enterprise Linux (RHEL) sem pagar por uma nova licença de plataforma.

Servidores de banco de dados Caché e uso de armazenamento

- **Produção** em ambientes de software da Epic, um único banco de dados centrado no paciente é implantado. Nos requisitos de hardware da Epic, o servidor físico que hospeda o servidor de dados Iris primário de leitura/gravação é chamado de servidor de banco de dados de produção. Este servidor requer

armazenamento all-flash de alto desempenho para arquivos pertencentes à instância de banco de dados principal. Para alta disponibilidade, a Epic oferece suporte ao uso de um servidor de banco de dados de failover que tenha acesso aos mesmos arquivos. O Iris usa o Epic Mirror para replicação para relatórios somente leitura, recuperação de desastres e suporte a cópias somente leitura. Cada tipo de servidor de banco de dados pode ser comutado para o modo de leitura/gravação por motivos de continuidade de negócios.

- **Report** Um servidor de banco de dados espelho de relatório fornece acesso somente leitura aos dados de produção. Ele hospeda um servidor de dados Iris configurado como um espelho de backup do servidor de dados Iris de produção. O servidor de banco de dados de relatórios tem os mesmos requisitos de capacidade de armazenamento que o servidor de banco de dados de produção. A performance de gravação do relatório é a mesma que a produção, mas as características do workload de leitura são diferentes e dimensionadas de maneira diferente.
- **Suporta somente leitura** este servidor de banco de dados é opcional e não é mostrado a figura abaixo. Um servidor de banco de dados espelhado também pode ser implantado para dar suporte à funcionalidade Epic somente leitura, na qual o acesso é fornecido a uma cópia de produção no modo somente leitura. Este tipo de servidor de banco de dados pode ser comutado para o modo de leitura/gravação por motivos de continuidade de negócios.
- **Recuperação de desastres** para atender aos objetivos de continuidade de negócios e recuperação de desastres, um servidor de banco de dados espelhado de recuperação de desastres é comumente implantado em um local geograficamente separado dos servidores de banco de dados espelhados de produção e/ou relatório. Um servidor de banco de dados espelhado de recuperação de desastres também hospeda um servidor de dados Iris configurado como um espelho de backup do servidor de dados Iris de produção. Se o local de produção ficar indisponível por um longo período de tempo, este servidor de banco de dados espelho de backup pode ser configurado para atuar como uma instância de leitura/gravação espelhada (SRW). O servidor de banco de dados espelho de backup tem os mesmos requisitos de armazenamento de arquivos que o servidor de banco de dados de produção. Em contraste, o armazenamento do banco de dados espelhado de backup é dimensionado da mesma forma que o storage de produção da perspectiva de desempenho para a continuidade dos negócios.



- **Test** as organizações de saúde costumam implantar ambientes de desenvolvimento, teste e teste. Servidores de dados Iris adicionais para esses ambientes também exigem armazenamento, que pode ser acomodado pelo mesmo sistema de armazenamento. A Epic tem requisitos e restrições específicos para fornecer storage adicional a partir de um sistema de storage compartilhado. Esses requisitos específicos são tratados genericamente pelas melhores práticas neste documento.

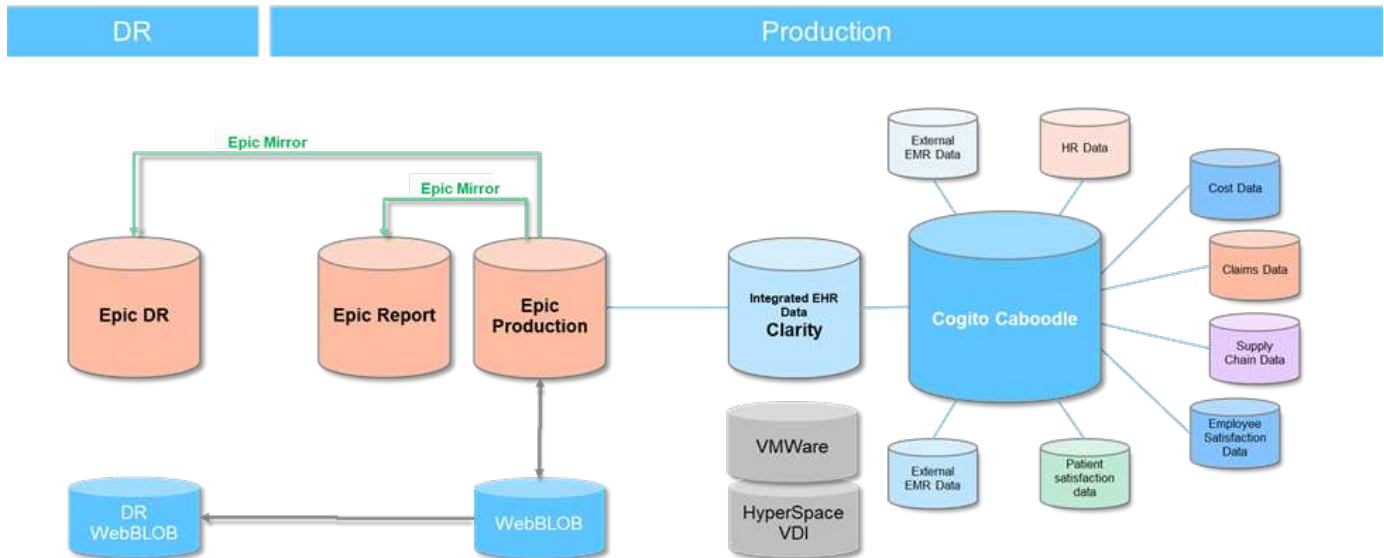
Além dos servidores de dados do Iris ODB, os ambientes de software da Epic geralmente incluem outros componentes, como o seguinte e como mostrado na figura abaixo:

- Um servidor de banco de dados Oracle ou Microsoft SQL Server como um back-end para as ferramentas de relatório de negócios Clarity da Epic



O Clarity é usado para relatar dados extraídos diariamente do banco de dados Iris de relatório.

- Servidor WebBLOB (SMB)
- Servidor de banco de dados multiusuário
- Máquinas virtuais multiusuário (VMs)
- Hiperespaço para acesso ao cliente



Os requisitos de storage de todos esses vários workloads, pools, protocolos nas e SAN podem ser consolidados e hospedados por um único cluster ONTAP. Com essa consolidação, as organizações de saúde têm uma estratégia única de gerenciamento de dados para todos os workloads da Epic e não da Epic.

Workloads de banco de dados operacional

Cada servidor de banco de dados Epic executa e/S nos seguintes tipos de arquivos:

- Ficheiros de base de dados
- Arquivos de diário
- Ficheiros de aplicação

O workload de um servidor de banco de dados individual depende de sua função no ambiente de software da Epic. Por exemplo, arquivos de banco de dados de produção geralmente incorrem na carga de trabalho mais exigente, consistindo de 100% de solicitações de e/S aleatórias. O workload de qualquer banco de dados espelhado geralmente é menos exigente e tem menos solicitações de leitura. As cargas de trabalho de arquivo diário são principalmente sequenciais.

A Epic mantém um modelo de workload para benchmark de desempenho de storage e workload de clientes. Para obter mais informações sobre o modelo de workload da Epic, os resultados de benchmark e as orientações sobre o uso de ferramentas de dimensionamento do NetApp para dimensionar o storage corretamente para ambientes da Epic, consulte "[TR-3930i: Diretrizes de dimensionamento de NetApp para a Epic](#)" (login no NetApp necessário).

A Epic também fornece a cada cliente um guia de configuração de hardware personalizado que contém projeções de e/S e requisitos de capacidade de storage. Os requisitos de storage finais podem incluir ambientes de desenvolvimento, teste e/ou preparo e quaisquer outros workloads auxiliares que possam ser consolidados. Os clientes podem usar o guia de configuração de hardware para comunicar os requisitos totais de armazenamento ao NetApp. Este guia contém todos os dados necessários para dimensionar a implantação da Epic.

Durante a fase de implantação, a Epic oferece um Guia de layout de storage de banco de dados, que fornece detalhes mais granulares em nível de LUN que podem ser usados para um design avançado de storage. Observe que o Guia de layout de armazenamento de banco de dados é recomendações gerais de armazenamento e não específico do NetApp. Use este guia para determinar o melhor layout de armazenamento no NetApp.

Dimensionamento épico

Uma das principais considerações da arquitetura ao dimensionar um ambiente de storage Epic é o tamanho do banco de dados ODB.

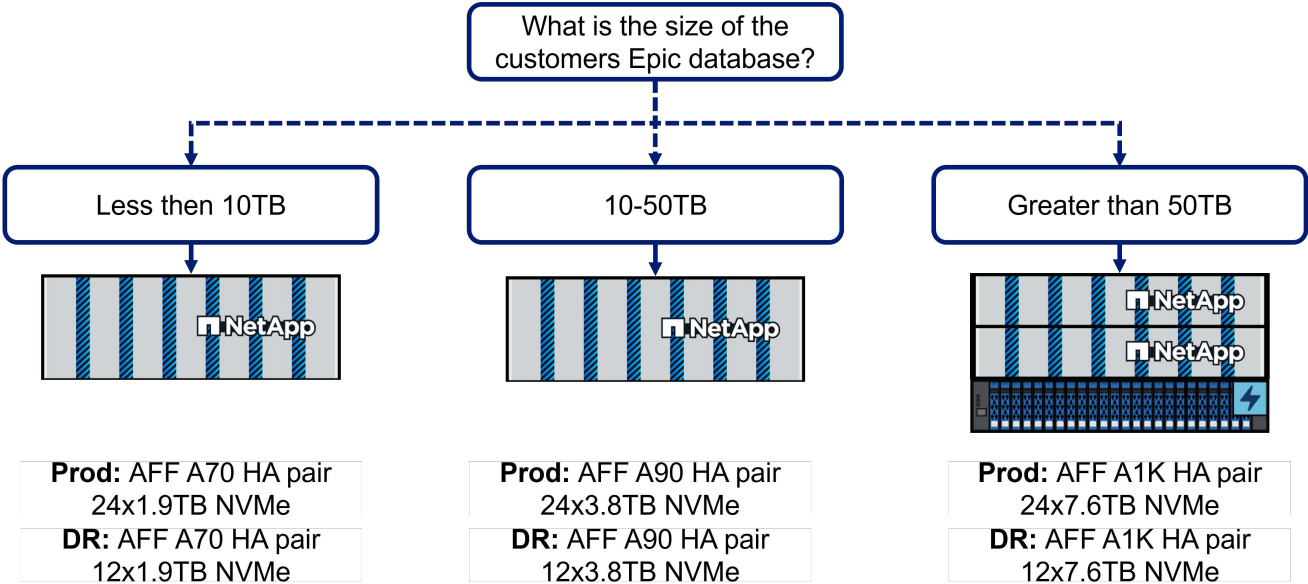
Use o diagrama abaixo para selecionar uma arquitetura de storage da Epic para pequeno e médio porte. Esses designs incluem a execução de todas as cargas de trabalho listadas no Guia de configuração de hardware. A árvore de dimensionamento é baseada em dados de mais de 100 guias de configuração de hardware e deve ser principalmente estimativa precisa.

É importante notar que este é apenas um ponto de partida. Você deve trabalhar com nossa equipe de aliança Epic para confirmar todos os designs da Epic. A equipe pode ser alcançada em NetApp.com. Todas as implantações precisam atender às solicitações dos clientes, aderindo às práticas recomendadas pela Epic e pela NetApp.

- Arquitetura pequena da Epic com um banco de dados da Epic menos de 10TB TB
- Arquitetura média da Epic com um banco de dados da Epic de 10TB a 50TB
- Grande arquitetura da Epic com um banco de dados da Epic de mais de 50TB TB

Epic sizing decision tree

Work with the NetApp Epic alliance team to validate designs



Requisitos de storage épicos

Em geral, recursos de storage dedicados são fornecidos para o banco de dados de produção, enquanto as instâncias espelhadas compartilham recursos de storage secundário com outros componentes relacionados ao software da Epic, como as ferramentas de geração de relatórios Clarity.

Outros ambientes de storage de software, como os usados para arquivos de sistema e aplicações, também são fornecidos pelos recursos de storage secundário.

Além do dimensionamento, a Epic tem as seguintes regras de layout de storage adicionais e principais considerações:

- Desde 2020, todos os workloads de banco de dados operacional (ODB) precisam estar em all-flash arrays.
- A Epic recomenda que cada pool de storage esteja em hardware físico separado, incluindo pool1, pool2, pool3, NAS1 e NAS2.



Um nó em um cluster pode ser considerado como um pool de storage. Com o ONTAP 9.4 ou posterior e o AQuoS, você pode criar pools protegidos usando políticas.

- Nova recomendação de backup da Epic 3-2-1.
 - a. Cópia localizada no local remoto (recuperação de desastres)
 - b. Uma das cópias deve estar em uma plataforma de storage diferente da cópia primária
 - c. Cópias dos dados



Os clientes que usam o NetApp SnapMirror para fazer backup do NetApp não atendem às recomendações do 3-2-1. O motivo é que o ONTAP do ONTAP não satisfaz o segundo requisito listado acima. Você pode usar o SnapMirror diretamente do ONTAP para o storage de objetos no local (por meio da StorageGRID, por exemplo) ou para a nuvem para atender aos requisitos da Epic.

Para obter mais informações sobre mandatos de storage, consulte os seguintes guias da Epic disponíveis no Galaxy:

- Considerações sobre SAN
- Produtos de armazenamento e Status da tecnologia (spats)
- Guia de configuração de hardware

Arquitetura épica de quatro nós

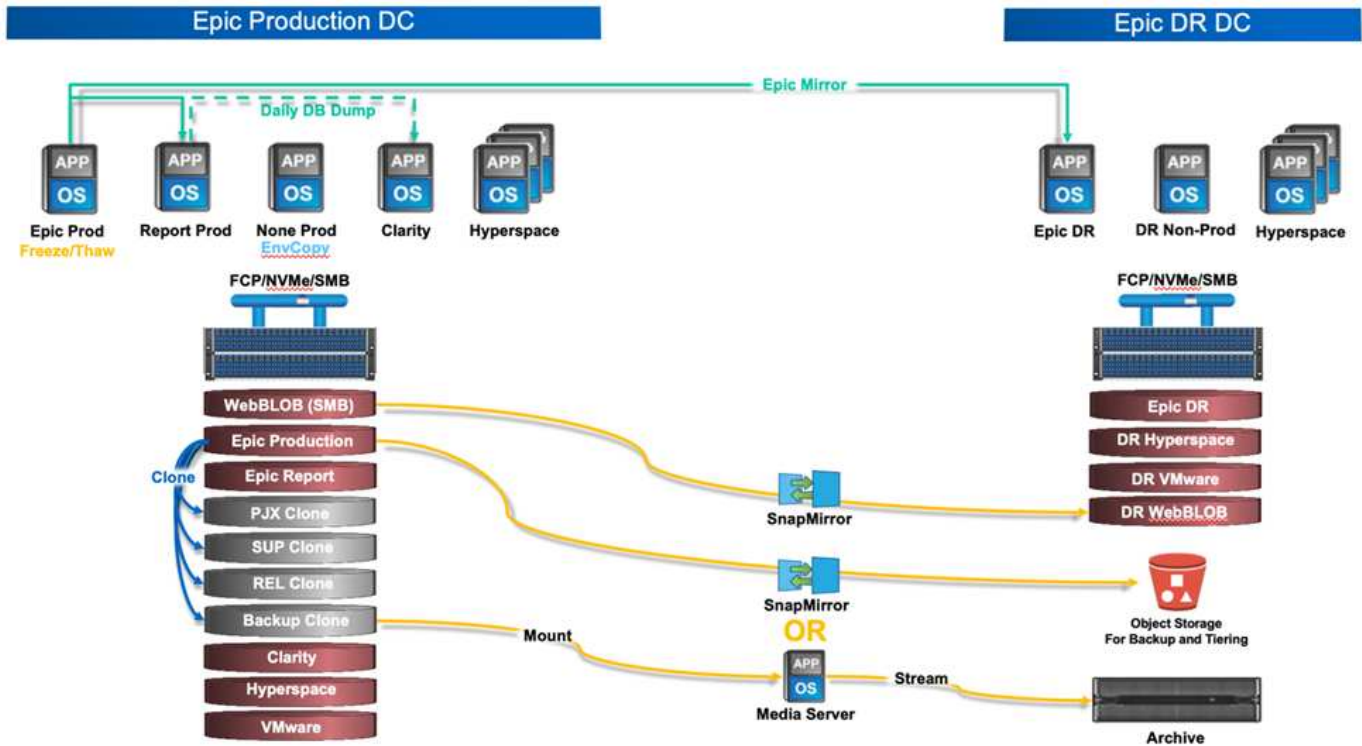
As figuras abaixo mostram o layout de storage para uma arquitetura de quatro nós: Um par de HA em produção e um par de HA na recuperação de desastres. O tamanho dos controladores e o número de discos baseiam-se na última imagem de dimensionamento.

A NetApp garante um desempenho mínimo ao nível do piso, aceitando as políticas recomendadas pela SLM. A Epic é compatível com a consolidação de pools de storage no ONTAP em muito menos hardware. Para obter mais informações, consulte o documento RELATÓRIOS trimestrais da Epic. Basicamente, pool1, pool2 e NAS1 (listados no Guia de configuração de hardware da Epic) podem ser executados em um único par de HA

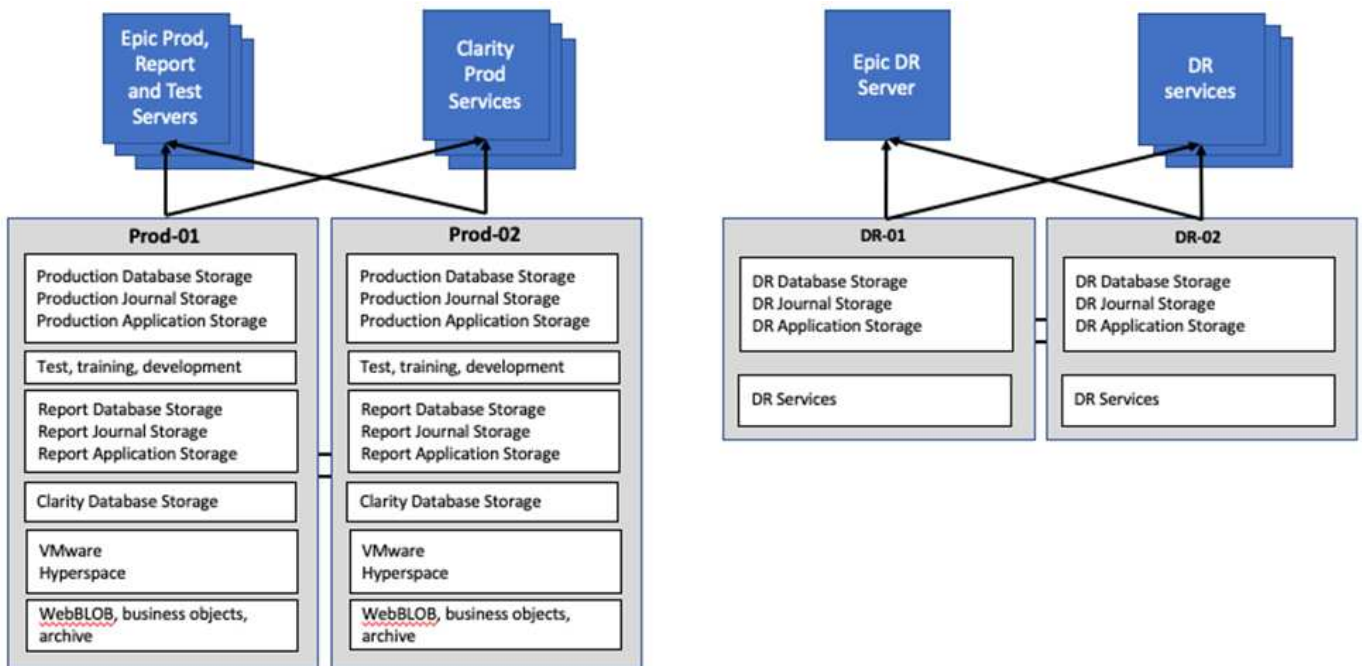
com workloads distribuídos uniformemente entre as duas controladoras. Na recuperação de desastres, o pool 3 e as 3 da Epic também são divididos entre as duas controladoras no par de HA.

Ambientes de cópia completa de teste (como SUP, REL e PJX) são clonados da Epic Production, Epic Report ou Epic Disaster Recovery. Para obter informações sobre backup e atualização da Epic, consulte a seção intitulada "Gerenciamento de dados".

Arquitetura de quatro nós



Disposição do workload de quatro nós

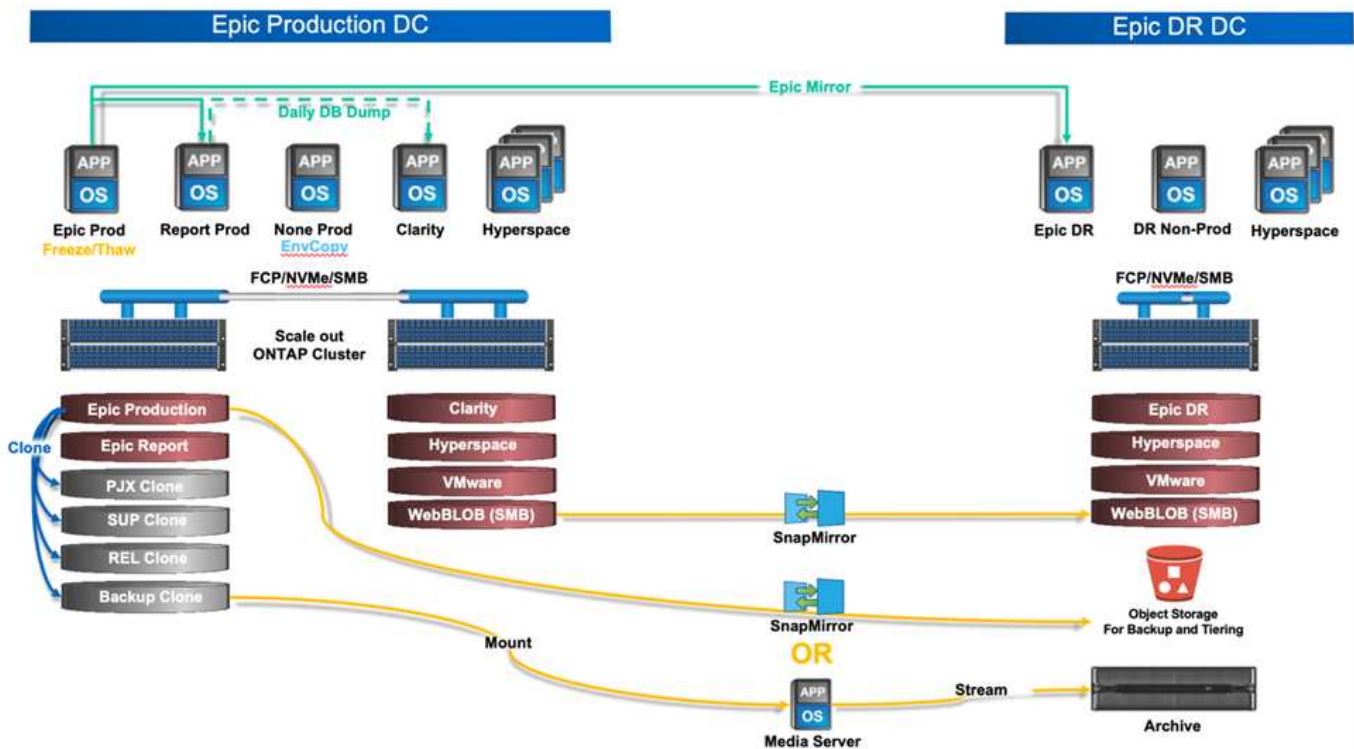


Arquitetura épica de seis nós

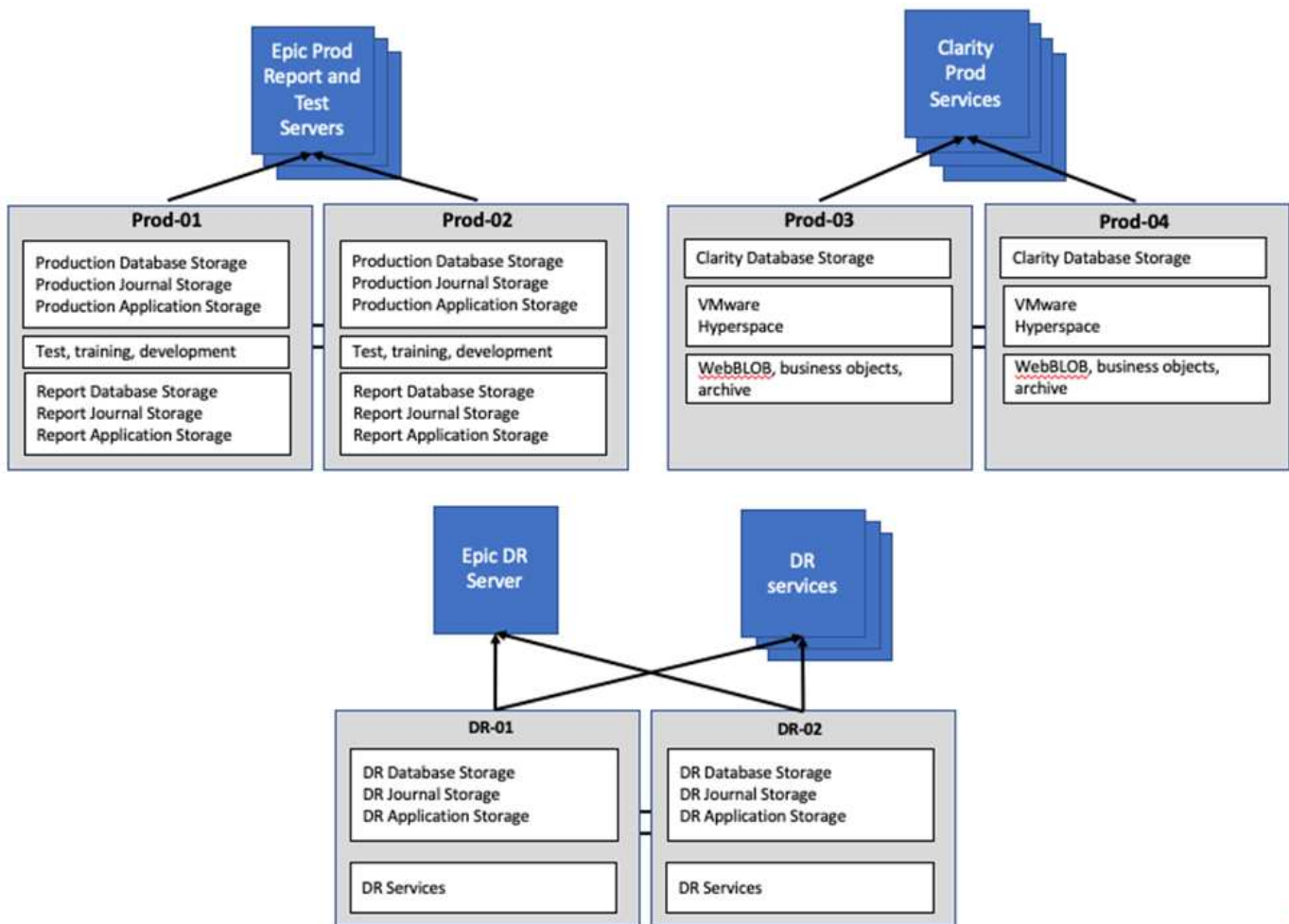
Os clientes podem querer começar com um design de seis nós ou fazer escalabilidade horizontal de quatro a seis nós de forma otimizada, com demanda crescente. Com a escalabilidade horizontal, você pode mover workloads entre nós e rebalancear de modo ininterrupto no cluster.

Essa arquitetura oferece o melhor desempenho e equilíbrio de capacidade no cluster. Produção épica, Epic Report e Epic Test são executados no primeiro par de HA. O segundo par de HA é usado para Clarity, Hyperspace, VMware, NAS1 e os demais workloads da Epic. A recuperação de desastres é a mesma que a arquitetura de quatro nós na seção anterior.

Arquitetura de seis nós



Disposição do workload de seis nós

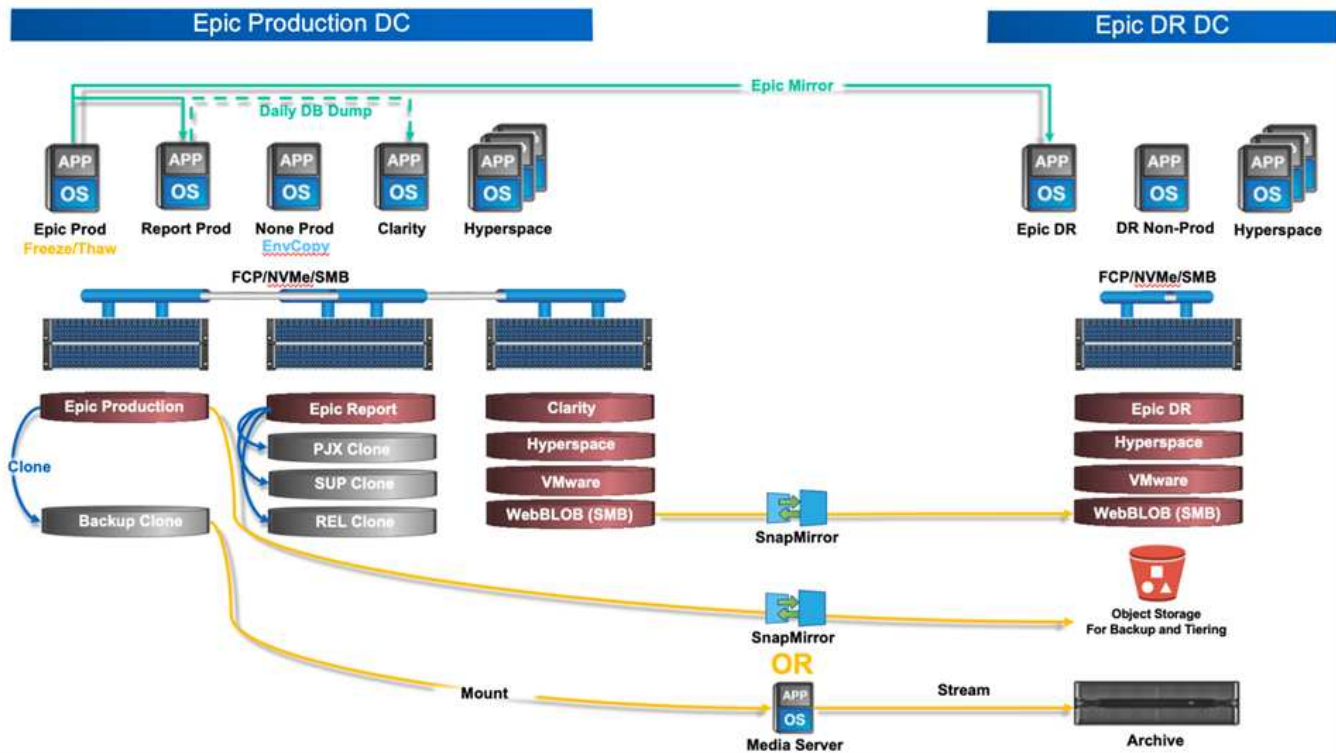


Arquitetura épica de oito nós

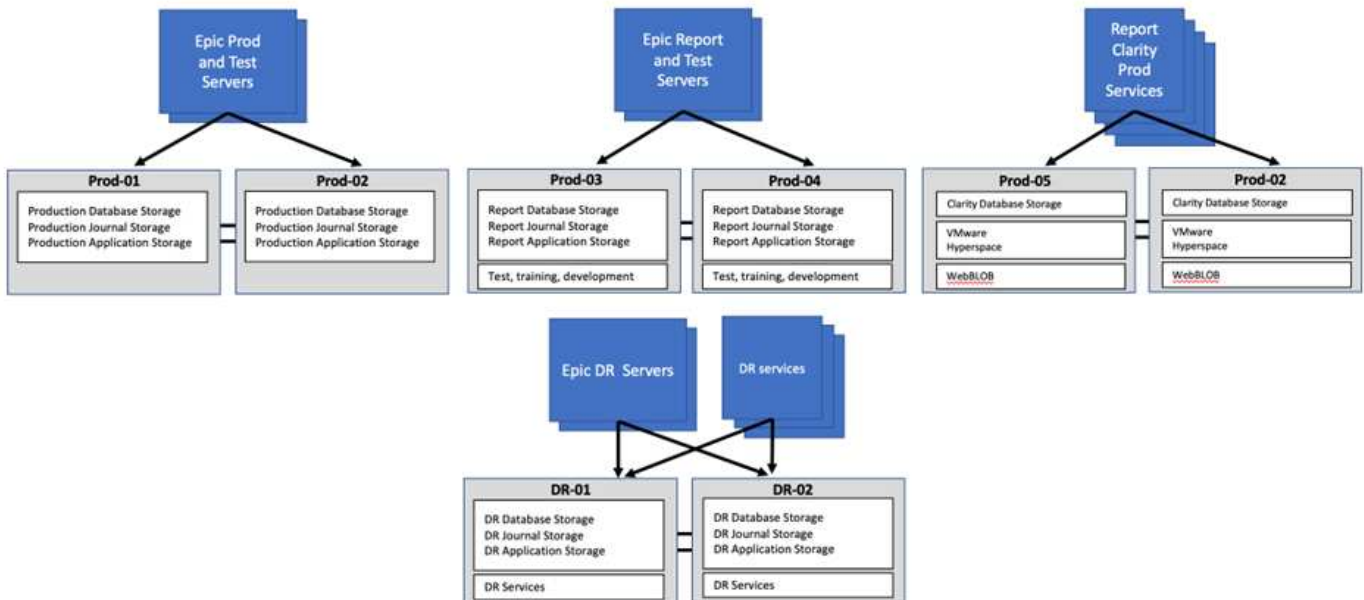
As figuras abaixo mostram a arquitetura de oito nós com escalabilidade horizontal. Novamente, você pode começar com quatro nós e escalar para seis nós e continuar a escalar para oito nós ou mais. Essa arquitetura oferece o melhor equilíbrio de performance e capacidade entre os seis nós em produção.

Os ambientes de teste são clonados do Report em vez da produção nesse projeto. Isso descarrega os ambientes de teste e as verificações de integridade da produção.

Arquitetura de oito nós



Posicionamento do workload de oito nós



Configuração e práticas recomendadas

EPIC em ONTAP - Utilitários de host

Os utilitários de host NetApp são pacotes de software para vários sistemas operacionais que contêm utilitários de gerenciamento, como o `sanlun` binário CLI, drivers multipath e outros arquivos importantes necessários para operações de SAN adequadas.



A NetApp recomenda a instalação dos Utilitários de host do NetApp em hosts que estejam conectados e acessando sistemas de armazenamento do NetApp. Para obter mais informações, consulte "[Ferramenta de Matriz de interoperabilidade](#)" e "[Anfitriões de SAN](#)" documentação.



Com o AIX, é especialmente importante que os Utilitários de host sejam instalados antes de descobrir LUNs. Isso garante que o comportamento de multipathing LUN esteja configurado corretamente. Se a descoberta foi realizada sem os Utilitários do host, os LUNs precisarão ser desfigurados do sistema usando o `rmdev -dl` comando e, em seguida, redescobertos via `cfgmgr` ou uma reinicialização.

Configuração épica de volume e LUN

O documento de recomendações de layout de storage do banco de dados Epic fornece orientações sobre o tamanho e o número de LUNs para cada banco de dados.

É importante analisar este documento com o suporte do Epic DBA e da Epic, além de finalizar o número de LUNs e tamanhos de LUN de acordo com o ajuste. Essas recomendações de armazenamento são importantes para a profundidade da fila HBA, o desempenho do armazenamento, a facilidade de operação e a facilidade de expansão.

Para a consideração da profundidade da fila do SO do servidor, use um mínimo de oito LUNs (um LUN por volume) para um banco de dados. Aumente o número de LUNs pelo número de nós no cluster do ONTAP. Por exemplo, adicione 4 LUNs ao usar um cluster de 4 nós (2 par de HA). Para ambientes maiores, mais LUNs podem ser necessários; use o mesmo número de volumes (oito no total, distribuídos entre nós de storage) e adicione LUNs em várias de dois nós e volumes do cluster. Com essa abordagem, você escala facilmente seu ambiente Epic.

Exemplo 1: Cluster ONTAP de 2 nós

2 nós, 1 pares de HA, 8 volumes, 4 volumes por nó 8 LUNs, um LUN por volume adicionando 2 LUNs adicionais, um em node01 em volume01, um em Node02 em volume02.

Exemplo 2: Cluster ONTAP de 4 nós

4 nós, 2 pares de HA, 8 volumes, 2 volumes por nó 8 LUNs, um LUN por volume adicionando 4 LUNs adicionais, um em node01 em volume01, um em Node02 em volume02, um em node03 em volume03 e um em node04 em volume04.

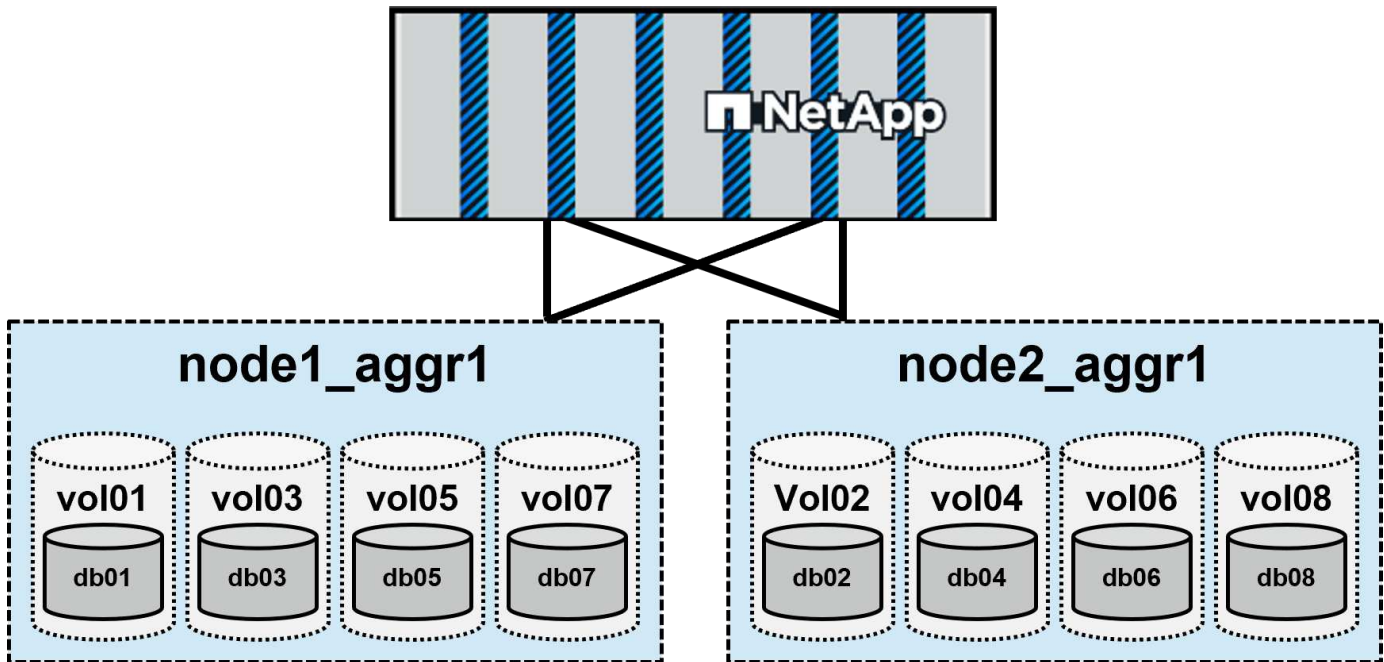
Para maximizar a performance de um workload, como Epic ODB ou Clarity, cada layout funciona melhor também para o storage NetApp. Com o uso de oito volumes, a e/S de gravação é distribuída uniformemente entre as controladoras, maximizando a utilização da CPU. Para replicação e backup, é melhor limitar o número de volumes a oito para simplificar as operações.

Opções de dimensionamento

Se o servidor exigir mais armazenamento, a opção mais fácil é aumentar os LUNs que contêm volumes. A segunda opção é adicionar LUNs aos grupos de volume em múltiplos de dois de cada vez (um por volume por nó).

Exemplo:

Volume e layout de 8 LUN



Se estiver em um ambiente grande que exija mais de 4 nós ou 8 LUNs, consulte nossa equipe de aliança Epic para confirmar os designs de LUN. A equipe pode ser alcançada em NetApp.com.

Práticas recomendadas

- Use 8 LUNs em volumes 8 para começar, adicionando 2 LUNs de cada vez, em todos os nós do cluster.
- Equilibre os workloads entre o par de HA para maximizar a performance e a eficiência.
- Criar LUNs no tamanho esperado para um crescimento de 3 anos. (Consulte o "[Documentação do ONTAP](#)" para obter os tamanhos máximos de LUN.)
- Usar LUNs e volumes provisionados com thin Provisioning.
- Use no mínimo oito LUNs de banco de dados, dois LUNs de diário e dois LUNs de aplicativos. Esta configuração maximiza o desempenho do armazenamento e a profundidade da fila do SO. Mais pode ser usado se necessário por capacidade ou outros motivos.
- Se você precisar adicionar LUNs a grupos de volumes, adicione oito LUNs de cada vez.
- Os grupos de consistência (CGS) são necessários para que o grupo de volumes e LUNs sejam copiados juntos.
- Não use QoS durante o Genio ou qualquer desempenho de e/S.
- Após o teste Genio ou Clarity, a NetApp recomenda excluir o armazenamento e o reprovisionamento antes de carregar os dados de produção.
- É importante que `-space-allocation` ativado esteja definido nos LUNs. Caso contrário, os dados excluídos nos LUNs não serão vistos pelo ONTAP e poderão causar problemas de capacidade. Para obter mais informações, consulte o Guia de Referência rápida de Configuração do Epic Storage.

Protocolos épicos e de arquivos

A combinação de nas e SAN no mesmo array all-flash é compatível.



A NetApp recomenda usar volumes FlexGroup para compartilhamentos nas, como WebBLOB (quando disponível).

WebBLOB é até 95% de dados inativos. Como opção, você pode liberar espaço em seu array all-flash, em categorias e backups e dados inativos no storage de objetos no local ou na nuvem, usando o "FabricPool" recurso do ONTAP. Tudo o que pode ser realizado sem qualquer efeito de desempenho perceptível. FabricPool é um recurso incluído do ONTAP. Os clientes podem gerar um relatório de dados inativos (ou inativos) para analisar quanto benefício poderia ser obtido com a ativação do FabricPool. Você pode definir a idade dos dados para categorizar por meio da política. Com esse recurso, os clientes da Epic conseguiram economias significativas.

Gerenciamento de performance épico

A maioria dos all-flash arrays garante a performance necessária para os workloads da Epic. O diferencial do NetApp é a capacidade de definir políticas de performance no nível básico e garantir uma performance consistente para cada aplicação.

Qualidade do serviço (QoS)

A NetApp recomenda o uso de QoS. O benefício da QoS é a capacidade de consolidar todos os workloads da Epic. Todos os protocolos e pools de storage podem residir em menos hardware. Não é necessário separar pools de storage.

- A NetApp recomenda que todos os workloads no cluster sejam atribuídos a uma política de QoS para melhor gerenciar o espaço livre no cluster.
- A NetApp recomenda o balanceamento de todos os workloads uniformemente entre o par de HA.
- Não use políticas de QoS ao executar qualquer teste de e/S; caso contrário, o teste do Genio falhará. Analise os diferentes workloads de produção por 2-4 semanas antes de atribuir políticas de QoS.

Épico em ONTAP - protocolos

FCP é o protocolo preferido para a apresentação de LUNs.



A NetApp recomenda zoneamento de iniciador único: Um iniciador por zona com todas as portas de destino necessárias no armazenamento usando nomes de portas mundiais (WWPNs). A presença de mais de um iniciador em uma única zona é suscetível de levar a interferência cruzada HBA intermitente, o que causa uma interrupção significativa.

Depois que o LUN é criado, mapeie o LUN para o grupo de iniciadores (igroup) que contém as WWPNs do host para habilitar o acesso.

O NetApp também dá suporte ao uso de NVMe/FC (se você tiver versões dos sistemas operacionais AIX e RHEL capazes) e melhora a performance. FCP e NVMe/FC podem coexistir na mesma malha.

Configuração de eficiência de storage épica

As eficiências de linha do ONTAP são ativadas por padrão e funcionam independentemente do protocolo de storage, aplicação ou camada de storage.

A eficiência reduz a quantidade de dados gravados em um storage flash caro e reduz o número de unidades

necessárias. O ONTAP preserva a eficiência com a replicação. Cada uma das eficiências tem pouco ou nenhum efeito na performance, mesmo para uma aplicação sensível à latência, como a Epic.



A NetApp recomenda ativar todas as configurações de eficiência para maximizar a utilização do disco. Estas definições estão ativadas por predefinição nos sistemas baseados em AFF e ASA.

Os recursos a seguir possibilitam essa eficiência de storage:

- A deduplicação economiza espaço no storage primário ao remover cópias redundantes de blocos em um volume que hospeda LUNs. Esta opção recomendada está ativada por predefinição.
- A compactação in-line reduz a quantidade de dados a serem gravados em disco, e uma economia considerável de espaço é obtida com os workloads da Epic. Esta opção recomendada está ativada por predefinição.
- A compactação in-line leva 4K blocos menos da metade cheios e os combina em um único bloco. Esta opção recomendada está ativada por predefinição.
- A thin replication está no centro do portfólio de software de proteção de dados da NetApp, que inclui o software NetApp SnapMirror. O thin replication SnapMirror protege os dados essenciais aos negócios, ao mesmo tempo em que minimiza os requisitos de capacidade de storage. **A NetApp recomenda** a ativação desta opção.
- Deduplicação de agregado. A deduplicação sempre esteve no nível do volume. Com o ONTAP 9.2, a deduplicação agregada tornou-se disponível, proporcionando economia adicional de redução de disco. A deduplicação agregada pós-processo foi adicionada com o ONTAP 9.3. **A NetApp recomenda** a ativação desta opção.

Configuração de eficiência de storage épica

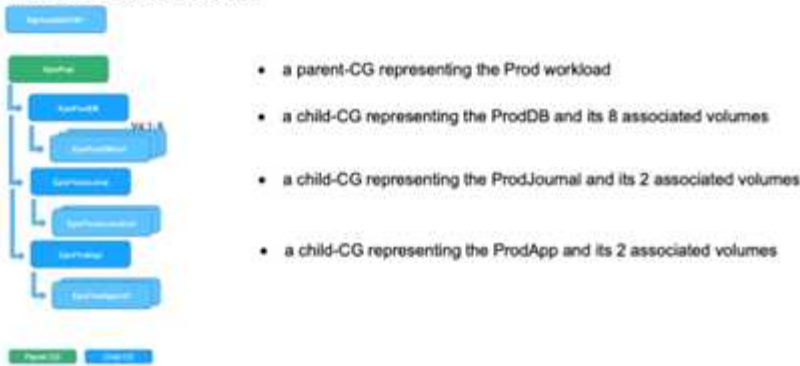
Os aplicativos com armazenamento distribuído em mais de um volume com um ou mais LUNs de quantidades apropriadas para a carga de trabalho precisam do backup do conteúdo, garantindo que a proteção de dados consistente exija CGS.

Grupos de consistência (CGS para abreviação) fornecem essa capacidade e muito mais. Eles podem ser usados todas as noites para criar snapshots consistentes sob demanda ou programados usando uma política. Use isso para restaurar, clonar e até mesmo replicar dados.

Para obter informações adicionais sobre a CGS, consulte a ["Visão geral dos grupos de consistência"](#)

Assim que os volumes e LUNs forem provisionados conforme detalhado nas seções anteriores deste documento, eles poderão ser configurados em um conjunto de CGS. A melhor prática recomendada é configurá-los como descrito na imagem abaixo:

**CG/EPIC EHR
Prod: CG Storage Layout**



Instantâneos do grupo de consistência

Um cronograma de instantâneos CG noturno deve ser definido em cada um dos CGS-filho associado aos volumes que fornecem armazenamento para o banco de dados de produção. Isso resultará em um novo conjunto de backups consistentes desses CGS todas as noites. Eles podem ser usados para clonar o banco de dados de produção para uso em ambientes que não sejam de produção, como desenvolvimento e teste. A NetApp desenvolveu fluxos de trabalho Ansible automatizados baseados em CG para a Epic a fim de automatizar o backup dos bancos de dados de produção, os ambientes de atualização e teste também.

Os snapshots CG podem ser usados para dar suporte às operações de restauração do banco de dados de produção da Epic.

Para volumes SAN, desative a política de instantâneos padrão em cada volume usado para CGS. Esses snapshots geralmente são gerenciados pela aplicação de backup usada ou pelo serviço de automação do NetApp Epic Ansible.

Para volumes SAN, desative a política de instantâneos padrão em cada volume. Esses snapshots geralmente são gerenciados por uma aplicação de backup ou pela automação do Epic Ansible.[NS2]

Os conjuntos de dados WebBLOB e VMware devem ser configurados como apenas volumes, não associados ao CGS. Você pode usar o SnapMirror para manter snapshots em sistemas de storage separados da produção.

Quando concluída, a configuração ficaria da seguinte forma:



Dimensionamento de storage para Epic

Você deve trabalhar com nossa equipe de aliança Epic para confirmar todos os designs

da Epic. A equipe pode ser alcançada em NetApp.com. Todas as implantações precisam atender às solicitações dos clientes, aderindo às práticas recomendadas pela Epic e pela NetApp.

Para obter informações sobre como usar as ferramentas de dimensionamento do NetApp para determinar o tamanho e o número corretos de grupos RAID para as necessidades de storage do ambiente de software da Epic, consulte "[TR-3930i: Diretrizes de dimensionamento de NetApp para a Epic](#)" (login no NetApp necessário).



O acesso ao portal de campo do NetApp é necessário.

Informações adicionais sobre a Epic no ONTAP

Para saber mais sobre as informações descritas neste documento, consulte os seguintes documentos e/ou sites:

- "[Documentação do produto NetApp](#)"
- "[Documentação do ONTAP 9](#)"
- "[Grupos de consistência](#)"
- "[Recursos de documentação do Gerenciador de sistemas ONTAP e ONTAP](#)"
- "[TR-3930i: Diretrizes de dimensionamento de NetApp para a Epic](#)" (É necessário iniciar sessão no NetApp)

Documentos épicos de orientação para o cliente

A Epic fornece aos clientes os seguintes documentos para orientação sobre servidor, armazenamento e rede. Estes documentos são referenciados neste relatório técnico.

- Considerações sobre a rede da área de armazenamento
- Guia de soluções técnicas de continuidade de negócios
- Manual de estratégia de arquitetura de referência all-flash
- Produtos de armazenamento e Status da tecnologia
- Considerações sobre a nuvem épica
- Guia de configuração de hardware (específico do cliente)
- Recomendações de layout de armazenamento de banco de dados (específicas do cliente)

Hyper-V

Diretrizes de implantação e práticas recomendadas de storage

Visão geral

O Microsoft Windows Server é um sistema operacional (SO) de classe empresarial que abrange rede, segurança, virtualização, nuvem privada, nuvem híbrida, infraestrutura de desktop virtual, proteção de acesso, proteção de informações, serviços da Web, infraestrutura de plataforma de aplicativos e muito mais.



Esta documentação substitui relatórios técnicos publicados anteriormente *TR-4568: Diretrizes de implantação do NetApp e práticas recomendadas de armazenamento para o Windows Server*

O ONTAP é executado em controladores de storage NetApp. Está disponível em vários formatos.

- Uma arquitetura unificada que dá suporte a protocolos de arquivos, objetos e blocos. Isso permite que os controladores de storage atuem como dispositivos nas e SAN, bem como armazenamentos de objetos
- Um All SAN Array (ASA) que se concentra apenas em protocolos de bloco e otimiza o tempo de retorno de I/O (IORT), adicionando multipathing ativo-ativo simétrico para hosts conectados
- Uma arquitetura unificada definida por software
 - ONTAP Select em execução no VMware vSphere ou KVM
 - Cloud Volumes ONTAP executado como uma instância nativa da nuvem
- Primeiras ofertas de fornecedores de nuvem em hiperescala
 - Amazon FSX para NetApp ONTAP
 - Azure NetApp Files
 - Google Cloud NetApp volumes

O ONTAP fornece recursos de eficiência de storage da NetApp, como a tecnologia NetApp Snapshot®, clonagem, deduplicação, thin Provisioning, thin replication, compactação, disposição em camadas de storage virtual e muito mais, com performance e eficiência aprimoradas.

Juntos, o Windows Server e o ONTAP podem operar em ambientes grandes e agregar imenso valor à consolidação do data center e às implantações de nuvem privada ou híbrida. Essa combinação também fornece workloads sem interrupções com eficiência e dá suporte à escalabilidade otimizada.

Público-alvo

Este documento destina-se a arquitetos de sistemas e armazenamento que projetam soluções de armazenamento NetApp para o Windows Server.

Fazemos as seguintes suposições neste documento:

- O leitor tem conhecimento geral das soluções de hardware e software da NetApp. Consulte "[Guia de administração do sistema para administradores de cluster](#)" para obter detalhes.

- O leitor tem conhecimento geral de protocolos de acesso a bloco, como iSCSI, FC e o protocolo de acesso a arquivos SMB/CIFS. Consulte "[Gerenciamento de SAN Clustered Data ONTAP](#)" para obter informações relacionadas com SAN. Consulte a "[Gerenciamento nas](#)" para obter informações relacionadas com CIFS/SMB.
- O leitor tem conhecimento geral do sistema operacional Windows Server e Hyper-V.

Para obter uma matriz completa e atualizada regularmente de configurações SAN e nas testadas e com suporte, consulte "[Ferramenta de Matriz de interoperabilidade \(IMT\)](#)" no site de suporte da NetApp. Com o IMT, você pode determinar as versões exatas do produto e do recurso que são suportadas para seu ambiente específico. O NetApp IMT define os componentes e versões do produto compatíveis com as configurações suportadas pelo NetApp. Os resultados específicos dependem da instalação de cada cliente de acordo com as especificações publicadas.

Armazenamento NetApp e ambiente Windows Server

Como mencionado no "[Visão geral](#)", as controladoras de storage NetApp fornecem uma arquitetura realmente unificada que dá suporte aos protocolos de arquivo, bloco e objeto. Isso inclui SMB/CIFS, NFS, NVMe/TCP, NVMe/FC, iSCSI, FC (FCP) e S3, além de criar acesso unificado de cliente e host. O mesmo controlador de storage pode, ao mesmo tempo, fornecer serviço de storage de bloco na forma de LUNs SAN e serviço de arquivos como NFS e SMB/CIFS. O ONTAP também está disponível como um ASA (All SAN Array) que otimiza o acesso ao host por meio de multipathing ativo-ativo simétrico com iSCSI e FCP, enquanto os sistemas ONTAP unificados usam multipathing ativo-ativo assimétrico. Em ambos os modos, o ONTAP usa o ANA para gerenciamento de multipath NVMe sobre Fabrics (NVMe-of).

Uma controladora de storage NetApp executando o ONTAP pode dar suporte aos seguintes workloads em um ambiente Windows Server:

- VMs hospedadas em compartilhamentos SMB 3,0 continuamente disponíveis
- VMs hospedadas em LUNs de volume compartilhado (CSV) de cluster executadas em iSCSI ou FC
- Bancos de dados SQL Server em compartilhamentos SMB 3,0
- Bancos de dados SQL Server em NVMe-of, iSCSI ou FC
- Outros workloads de aplicações

Além disso, os recursos de eficiência de storage da NetApp, como deduplicação, cópias NetApp FlexClone®, tecnologia NetApp Snapshot, thin Provisioning, compactação e disposição em camadas de storage, agregam valor significativo para workloads executados em Windows Server.

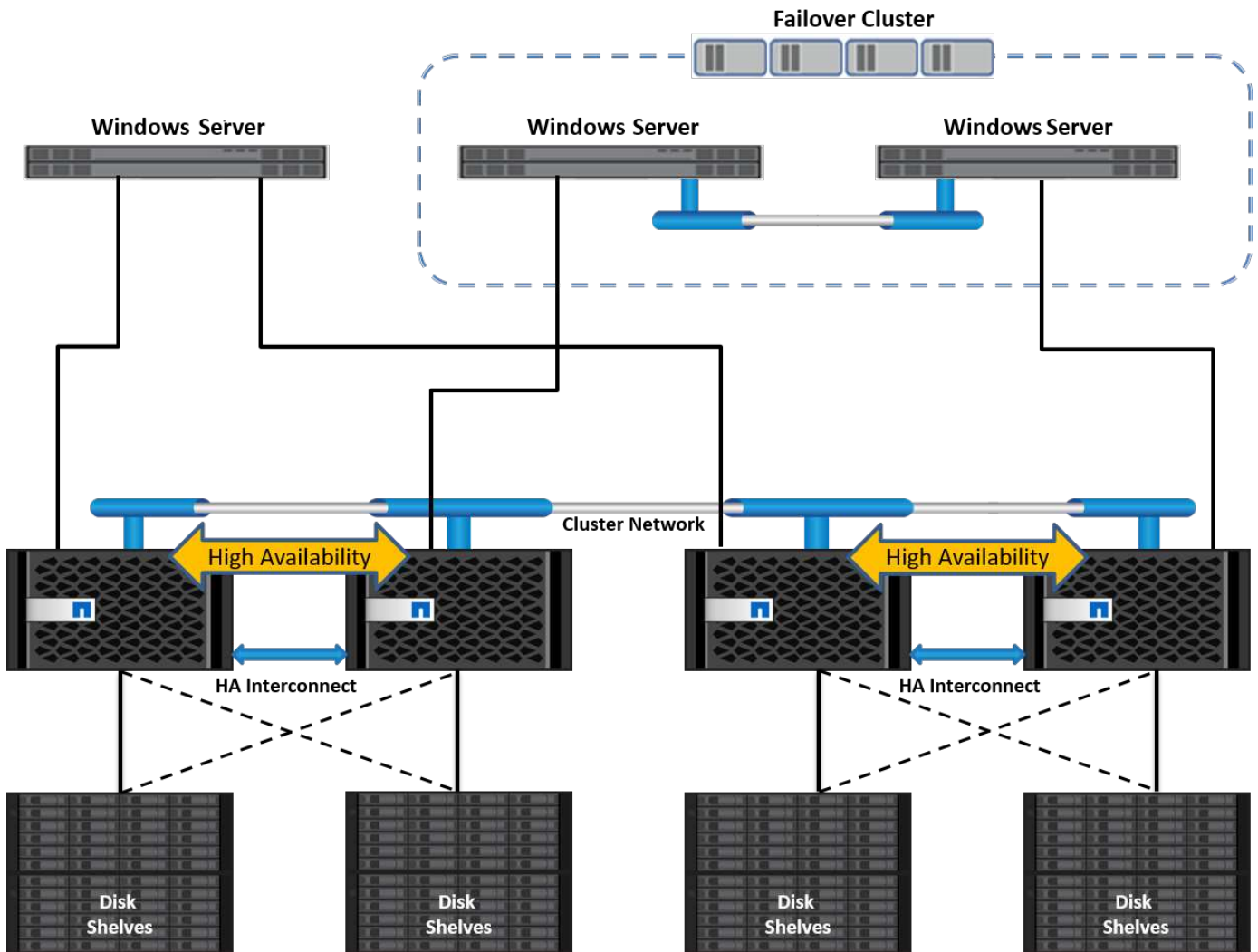
ONTAP data Management

O ONTAP é um software de gerenciamento executado em um controlador de storage da NetApp. Referido como um nó, um controlador de armazenamento NetApp é um dispositivo de hardware com um processador, RAM e NVRAM. O nó pode ser conectado a unidades de disco SATA, SAS ou SSD ou a uma combinação dessas unidades.

Vários nós são agregados em um sistema em cluster. Os nós no cluster comunicam-se continuamente entre si para coordenar as atividades do cluster. Os nós também podem mover dados de forma transparente de nó para nó usando caminhos redundantes para uma rede de cluster dedicada que consiste em dois switches Ethernet 10Gb. Os nós no cluster podem assumir o controle uns dos outros para fornecer alta disponibilidade

durante quaisquer cenários de failover. Os clusters são administrados em um cluster inteiro, em vez de por nó, e os dados são fornecidos a partir de uma ou mais máquinas virtuais de storage (SVMs). Para fornecer dados, um cluster precisa ter pelo menos um SVM.

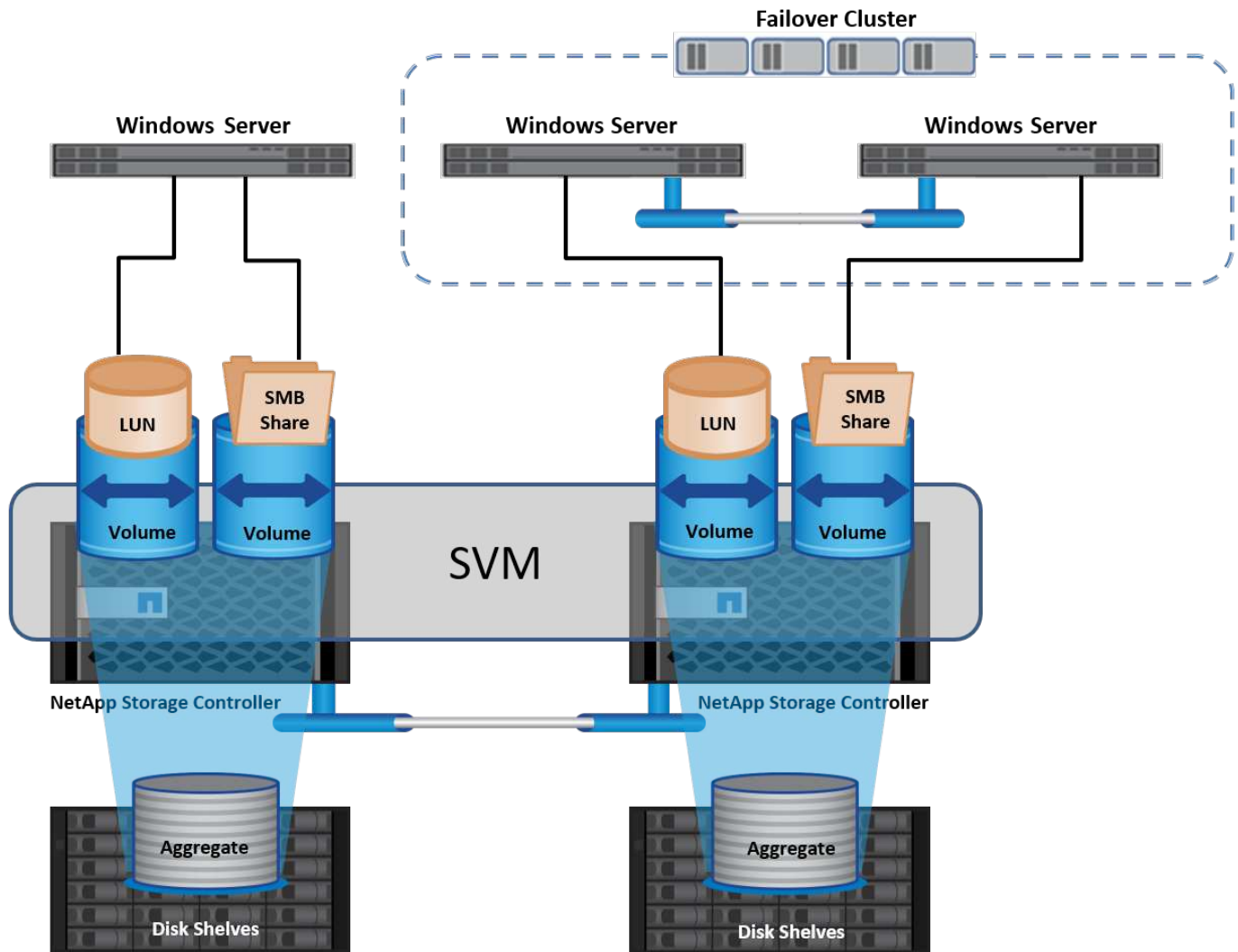
A unidade básica de um cluster é o nó e os nós são adicionados ao cluster como parte de um par de alta disponibilidade (HA). Os pares DE HA permitem alta disponibilidade comunicando-se entre si por meio de uma interconexão de HA (separada da rede de cluster dedicada) e mantendo conexões redundantes com os discos do par de HA. Os discos não são compartilhados entre pares de HA, embora as gavetas possam conter discos que pertencem a um dos membros de um par de HA. A figura a seguir mostra uma implantação de storage do NetApp em um ambiente do Windows Server.



Máquinas virtuais de armazenamento

O ONTAP SVM é um servidor de storage lógico que fornece acesso a dados a LUNs e/ou namespace nas a partir de uma ou mais interfaces lógicas (LIFs). O SVM é, portanto, a unidade básica de segmentação de storage que permite a alocação segura a vários clientes no ONTAP. Cada SVM é configurado para possuir volumes de storage provisionados de um agregado físico e interfaces lógicas (LIFs) atribuídas a uma rede Ethernet física ou a portas de destino FC.

Discos lógicos (LUNs) ou compartilhamentos CIFS são criados dentro dos volumes de uma SVM e são mapeados para hosts e clusters do Windows para fornecer espaço de storage, conforme mostrado na figura a seguir. Os SVMs são independentes de nós e baseados em cluster; eles podem usar recursos físicos, como volumes ou portas de rede em qualquer lugar do cluster.



Provisionamento de armazenamento NetApp para Windows Server

O storage pode ser provisionado para o Windows Server em ambientes SAN e nas. Em um ambiente SAN, o storage é fornecido como discos de LUNs no volume NetApp como storage de bloco. Em um ambiente nas, o storage é fornecido como compartilhamentos CIFS/SMB em volumes NetApp como storage de arquivos. Esses discos e compartilhamentos podem ser aplicados no Windows Server da seguinte forma:

- Storage para hosts do Windows Server para workloads de aplicativos
- Armazenamento para Nano Server e contentores
- Armazenamento para hosts Hyper-V individuais para armazenar VMs
- Armazenamento compartilhado para clusters Hyper-V na forma de CSVs para armazenar VMs
- Armazenamento para bancos de dados SQL Server

Gerenciamento do storage do NetApp

Para conectar, configurar e gerenciar o armazenamento do NetApp a partir do Windows Server 2016, use um dos seguintes métodos:

- **Secure Shell (SSH).** Use qualquer cliente SSH no Windows Server para executar comandos NetApp CLI.
- **System Manager.** Este é o produto de gerenciamento baseado em GUI da NetApp.

- **Kit de ferramentas do NetApp PowerShell.** Este é o Toolkit do NetApp PowerShell para automatizar e implementar scripts e fluxos de trabalho personalizados.

Kit de ferramentas do NetApp

O NetApp PowerShell Toolkit (PSTK) é um módulo do PowerShell que fornece automação de ponta a ponta e permite a administração de storage do ONTAP. O módulo ONTAP contém mais de 2.000 cmdlets e ajuda na administração de FAS, NetApp All Flash FAS (AFF), hardware comum e recursos de nuvem.

Coisas para se lembrar

- O NetApp não suporta espaços de armazenamento do Windows Server. Os espaços de armazenamento são usados apenas para JBOD (apenas um monte de discos) e não funciona com qualquer tipo de RAID (armazenamento de conexão direta [DAS] ou SAN).
- Os pools de armazenamento em cluster no Windows Server não são suportados pelo ONTAP.
- O NetApp suporta o formato de disco rígido virtual compartilhado (VHDX) para cluster de convidados em ambientes SAN do Windows.
- O Windows Server não suporta a criação de pools de armazenamento usando iSCSI ou FC LUNs.

Leitura adicional

- Para obter mais informações sobre o Kit de ferramentas do NetApp PowerShell, visite o ["Site de suporte da NetApp"](#).
- Para obter informações sobre as práticas recomendadas do NetApp PowerShell Toolkit, ["TR-4475: Guia de práticas recomendadas do kit de ferramentas do NetApp PowerShell"](#) consulte .

Melhores práticas de rede

As redes Ethernet podem ser amplamente segregadas nos seguintes grupos:

- Uma rede cliente para as VMs
- Mais uma rede de armazenamento (iSCSI ou SMB conectando-se aos sistemas de armazenamento)
- Uma rede de comunicação de cluster (heartbeat e outra comunicação entre os nós do cluster)
- Uma rede de gerenciamento (para monitorar e solucionar problemas do sistema)
- Uma rede de migração (para migração ao vivo do host)
- Replicação de VM (uma réplica do Hyper-V)

Práticas recomendadas

- A NetApp recomenda ter portas físicas dedicadas para cada uma das funcionalidades anteriores para isolamento e desempenho da rede.
- Para cada um dos requisitos de rede anteriores (exceto para os requisitos de armazenamento), várias portas de rede físicas podem ser agregadas para distribuir carga ou fornecer tolerância a falhas.
- A NetApp recomenda ter um switch virtual dedicado criado no host Hyper-V para conexão de armazenamento convidado dentro da VM.
- Certifique-se de que o host Hyper-V e os caminhos de dados iSCSI Guest usam diferentes portas físicas e switches virtuais para isolamento seguro entre o convidado e o host.
- A NetApp recomenda evitar o agrupamento de NIC para NICs iSCSI.

- A NetApp recomenda o uso de ONTAP multipath input/output (MPIO) configurado no host para fins de armazenamento.
- A NetApp recomenda o uso de MPIO em uma VM convidada se estiver usando iniciadores iSCSI convidados. O uso do MPIO deve ser evitado dentro do convidado se você usar discos pass-through. Neste caso, a instalação do MPIO no host deve ser suficiente.
- A NetApp recomenda não aplicar políticas de QoS ao switch virtual atribuído à rede de storage.
- A NetApp recomenda não usar o endereçamento IP privado automático (APIPA) em NICs físicos porque o APIPA não é roteável e não está registrado no DNS.
- A NetApp recomenda a ativação de quadros jumbo para redes de migração CSV, iSCSI e Live para aumentar a taxa de transferência e reduzir os ciclos da CPU.
- A NetApp recomenda desmarcar a opção permitir que o sistema operacional de Gerenciamento compartilhe esta placa de rede para que o switch virtual Hyper-V crie uma rede dedicada para as VMs.
- A NetApp recomenda a criação de caminhos de rede redundantes (vários switches) para migração em tempo real e para a rede iSCSI para fornecer resiliência e QoS.

Provisionamento em ambientes SAN

As SVMs ONTAP são compatíveis com os protocolos de bloco iSCSI e FC. Quando um SVM é criado com protocolo de bloco iSCSI ou FC, o SVM obtém um nome qualificado iSCSI (IQN) ou um nome mundial FC (WWN), respectivamente. Esse identificador apresenta um destino SCSI para hosts que acessam o storage de bloco do NetApp.

Provisionamento de LUN NetApp no Windows Server

Pré-requisitos

O uso do armazenamento NetApp em ambientes SAN no Windows Server tem os seguintes requisitos:

- Um cluster NetApp é configurado com uma ou mais controladores de storage NetApp.
- O cluster NetApp ou os controladores de armazenamento têm uma licença iSCSI válida.
- Estão disponíveis portas configuradas iSCSI e/ou FC.
- O zoneamento FC é realizado em um switch FC para FC.
- Pelo menos um agregado é criado.
- Um SVM deve ter um LIF por rede Ethernet ou malha Fibre Channel em todas as controladoras de storage que fornecerão dados usando iSCSI ou Fibre Channel.

Implantação

1. Criar um novo SVM com protocolo de bloco iSCSI e/ou FC habilitado. É possível criar um novo SVM com qualquer um dos seguintes métodos:
 - Comandos CLI no storage NetApp
 - Gerente do sistema da ONTAP
 - Kit de ferramentas do NetApp
2. Configurar o protocolo iSCSI e/ou FC.
3. Atribua o SVM com LIFs em cada nó de cluster.

4. Inicie o serviço iSCSI e/ou FC no SVM.
5. Crie conjuntos de portas iSCSI e/ou FC usando os LIFs SVM.
6. Crie um grupo de iniciadores iSCSI e/ou FC para Windows usando o conjunto de portas criado.
7. Adicione um iniciador ao grupo de iniciadores. O iniciador é o IQN para iSCSI e WWPN para FC. Eles podem ser consultados no Windows Server executando o cmdlet Get-InitiatorPort do 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
```

O IQN para iSCSI no Windows Server também pode ser verificado na configuração das propriedades do iniciador iSCSI.

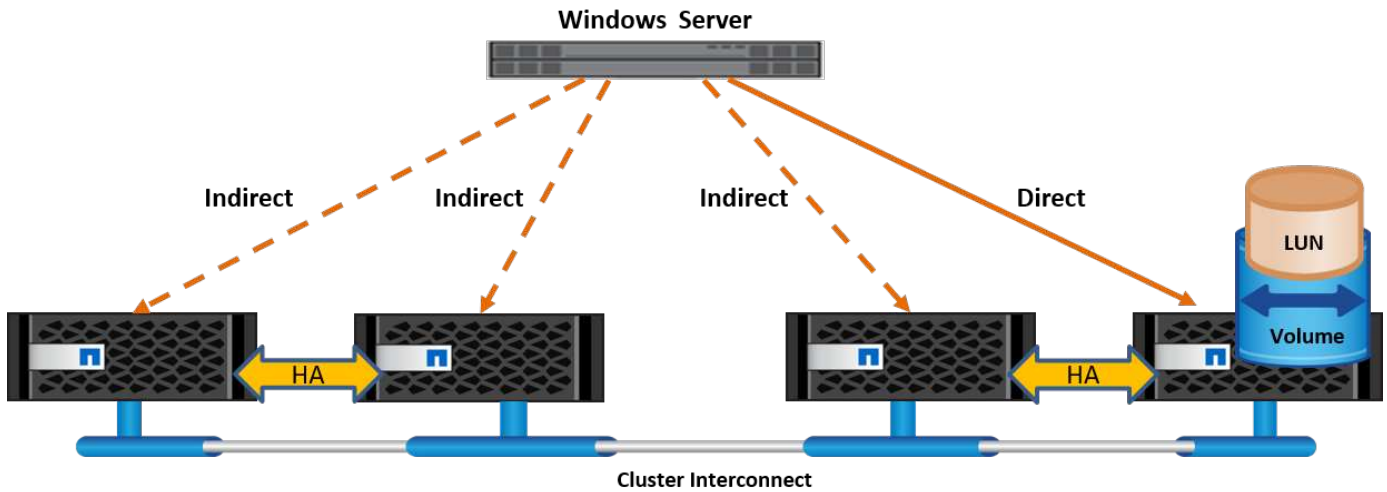
- Crie um LUN usando o assistente criar LUN e associe-o ao grupo de iniciadores criado.

Integração de host

O Windows Server usa a extensão MPIO (ALUA) de acesso lógico assimétrico para determinar caminhos diretos e indiretos para LUNs. Mesmo que cada LIF de uma SVM aceite solicitações de leitura/gravação de LUNs, apenas um dos nós do cluster realmente é proprietário dos discos que suportam esse LUN em um dado momento. Isso divide os caminhos disponíveis para um LUN em dois tipos, diretos ou indiretos, como mostrado na figura a seguir.

Um caminho direto para um LUN é um caminho no qual as LIFs de uma SVM e o LUN que estão sendo acessados residem no mesmo nó. Para ir de uma porta de destino física para o disco, não é necessário atravessar a rede do cluster.

Caminhos indiretos são caminhos de dados nos quais as LIFs de uma SVM e o LUN que estão sendo acessados residem em nós diferentes. Os dados devem atravessar a rede do cluster para ir de uma porta de destino físico para o disco.



MPIO

O ONTAP fornece armazenamento altamente disponível no qual podem existir vários caminhos do controlador de armazenamento para o servidor Windows. Multipathing é a capacidade de ter vários caminhos de dados de um servidor para um storage array. O multipathing protege contra falhas de hardware (cortes de cabos, falha do adaptador de barramento de host e switch [HBA], etc.), e pode fornecer limites de desempenho mais altos usando o desempenho agregado de várias conexões. Quando um caminho ou conexão se torna indisponível, o software multipathing muda automaticamente a carga para um dos outros caminhos disponíveis. O recurso MPIO combina os vários caminhos físicos para o storage como um único caminho lógico usado para acesso aos dados para fornecer resiliência de storage e balanceamento de carga. Para usar esse recurso, o recurso MPIO deve estar habilitado no Windows Server.

Ativar MPIO

Para ativar o MPIO no Windows Server, execute as seguintes etapas:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Inicie o Server Manager.
3. Na seção Gerenciar, clique em Adicionar funções e recursos.
4. Na página Selecionar recursos, selecione Multipath I/O.

Configure o MPIO

Ao usar o protocolo iSCSI, você deve informar ao Windows Server para aplicar suporte de multipath a dispositivos iSCSI nas propriedades MPIO.

Para configurar o MPIO no Windows Server, execute as seguintes etapas:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Inicie o Server Manager.
3. Na seção Ferramentas, clique em MPIO.
4. Em Propriedades de MPIO em Discover Multi-Paths, selecione Adicionar suporte para dispositivos iSCSI e clique em Adicionar. Em seguida, um prompt solicita que você reinicie o computador.
5. Reinicie o Windows Server para ver o dispositivo MPIO listado na seção dispositivos MPIO das Propriedades MPIO.

Configurar iSCSI

Para detetar armazenamento de bloco iSCSI no Windows Server, execute as seguintes etapas:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Inicie o Server Manager.
3. Na seção Ferramentas, clique em Iniciador iSCSI.
4. Na guia descoberta, clique em descobrir Portal.
5. Forneça o endereço IP das LIFs associadas ao SVM criado para o protocolo NetApp storage para SAN. Clique em Avançado, configure as informações na guia Geral e clique em OK.
6. O iniciador iSCSI deteta automaticamente o destino iSCSI e o lista no separador alvos.
7. Selecione o destino iSCSI em destinos descobertos. Clique em Connect (ligar) para abrir a janela Connect to Target (ligar ao destino).
8. Você deve criar várias sessões do host do Windows Server para os LIFs iSCSI de destino no cluster de armazenamento do NetApp. Para fazer isso, execute as seguintes etapas:
9. Na janela conectar ao destino, selecione Ativar MPIO e clique em Avançado.
10. Em Configurações avançadas na guia Geral, selecione o adaptador local como o iniciador iSCSI da Microsoft e selecione o IP do Iniciador e o IP do Portal de destino.
11. Você também deve se conectar usando o segundo caminho. Portanto, repita os passos 5 a 8, mas desta vez selecione o IP do Iniciador e o IP do Portal de destino para o segundo caminho.
12. Selecione o destino iSCSI em destinos descobertos na janela principal Propriedades iSCSI e clique em Propriedades.
13. A janela Propriedades mostra que várias sessões foram detetadas. Selecione a sessão, clique em dispositivos e, em seguida, clique no MPIO para configurar a política de balanceamento de carga. Todos os caminhos configurados para o dispositivo são exibidos e todas as políticas de balanceamento de carga são suportadas. O NetApp geralmente recomenda round robin com subconjunto, e essa configuração é o padrão para arrays com ALUA habilitado. Round robin é o padrão para arrays ativo-ativo que não suportam ALUA.

Detectar storage de bloco

Para detetar armazenamento de blocos iSCSI ou FC no Windows Server, execute as seguintes etapas:

1. Clique em Gerenciamento de computador na seção Ferramentas do Gerenciador de servidores.
2. Em Gerenciamento de computador, clique na seção Gerenciamento de disco no armazenamento e clique em mais ações e Rescan Disks. Ao fazê-lo, apresenta os iSCSI LUNs brutos.
3. Clique no LUN descoberto e faça-o online. Em seguida, selecione Inicializar disco usando a partição MBR ou GPT. Crie um novo volume simples fornecendo o tamanho do volume e a letra da unidade e formate-o usando FAT, FAT32, NTFS ou o Resilient File System (refs).

Práticas recomendadas

- A NetApp recomenda a ativação do thin Provisioning nos volumes que hospedam os LUNs.
- Para evitar problemas de multipathing, o NetApp recomenda usar todas as sessões 10Gb ou todas as sessões 1GB para um determinado LUN.
- A NetApp recomenda que você confirme se o ALUA está ativado no sistema de storage. O ALUA é ativado por padrão no ONTAP.

- No host do servidor Windows para onde o LUN NetApp é mapeado, ative o Serviço iSCSI (TCP-in) para Serviço de entrada e iSCSI (TCP-out) para saída nas configurações do firewall. Essas configurações permitem que o tráfego iSCSI passe de e para o host Hyper-V e o controlador NetApp.

Provisionamento de LUNs NetApp em Nano Server

Pré-requisitos

Além dos pré-requisitos mencionados na seção anterior, a função de armazenamento deve ser ativada a partir do lado do Nano Server. Por exemplo, o Nano Server deve ser implantado usando a opção `-Storage`. Para implantar o Nano Server, consulte a seção ["Implante o Nano Server."](#)

Implantação

Para provisionar LUNs NetApp em um servidor Nano, execute as seguintes etapas:

1. Conecte-se remotamente ao Nano Server usando as instruções na seção ["Conecte-se ao Nano Server."](#)
2. Para configurar iSCSI, execute os seguintes cmdlets do PowerShell no 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 retrieved 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. Adicione um iniciador ao grupo de iniciadores.

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

4. Configure o 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. Detectar storage de bloco.

```
# 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
```

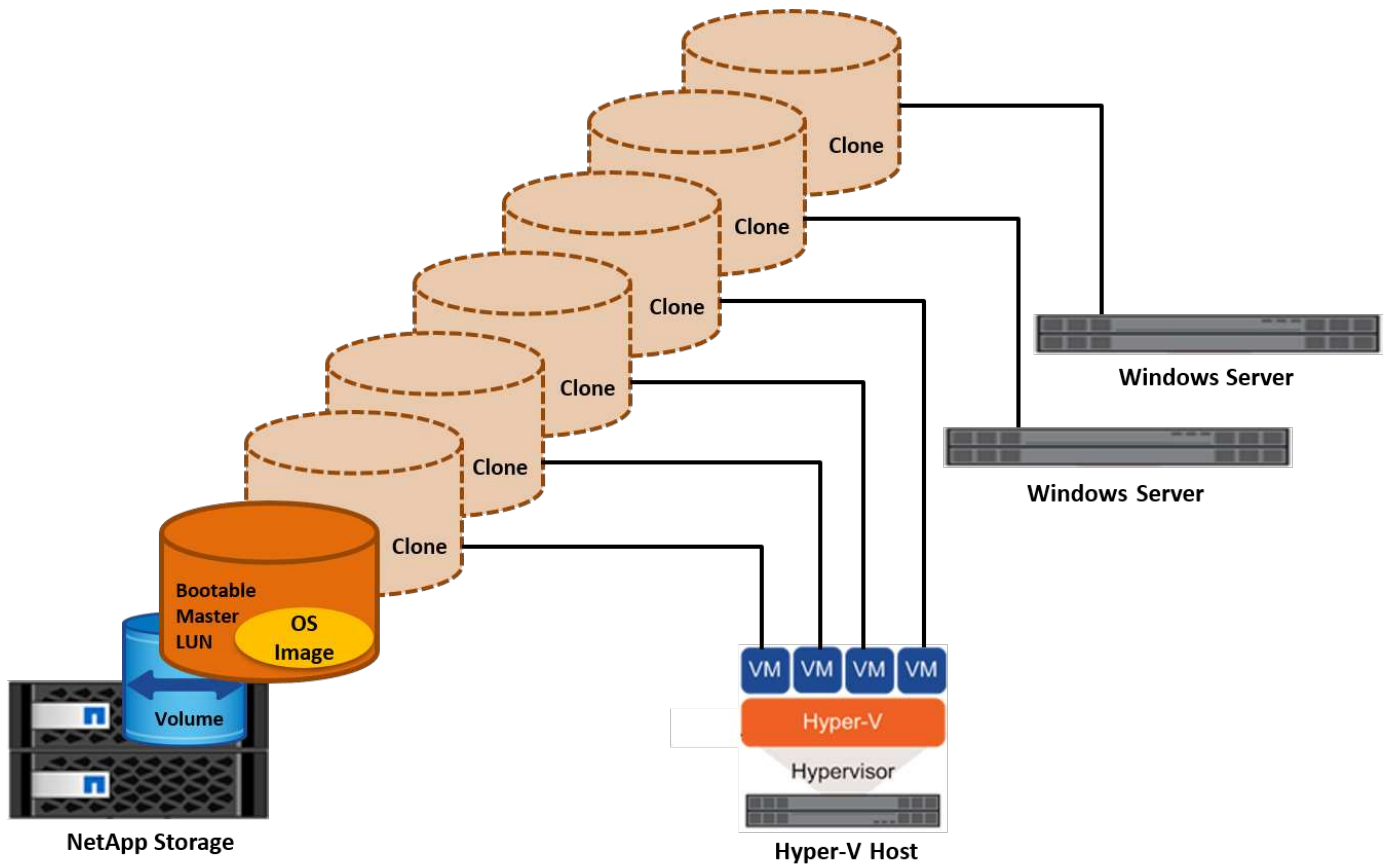
```
# To choose the size and drive letter use -Size and -DriveLetter
parameters
# Format the volume
Format-Volume -DriveLetter <DriveLetter> -FileSystem <FAT32 or NTFS or
REFS>
```

Arranque a partir de SAN

Um host físico (servidor) ou uma VM Hyper-V podem inicializar o sistema operacional do Windows diretamente de um LUN NetApp em vez de seu disco rígido interno. Na abordagem boot-from-SAN, a imagem do sistema operacional a ser inicializada reside em um LUN NetApp que é anexado a um host físico ou VM. Para um host físico, o HBA do host físico é configurado para usar o LUN NetApp para inicialização. Para uma VM, o LUN NetApp é conectado como um disco de passagem para inicialização.

Abordagem da NetApp FlexClone

Usando a tecnologia NetApp FlexClone, os LUNs de inicialização com uma imagem do sistema operacional podem ser clonados instantaneamente e conectados aos servidores e VMs para fornecer rapidamente imagens limpas do sistema operacional, como mostra a figura a seguir.



Arranque a partir de SAN para anfitrião físico

Pré-requisitos

- O host físico (servidor) tem um iSCSI ou FC HBA adequado.
- Você baixou um driver de dispositivo HBA adequado para o servidor que suporta o Windows Server.
- O servidor tem uma unidade de CD/DVD adequada ou um suporte virtual para inserir a imagem ISO do Windows Server e o controlador do dispositivo HBA foi transferido.
- Um iSCSI NetApp ou LUN FC é provisionado no controlador de storage NetApp.

Implantação

Para configurar a inicialização a partir de SAN para um host físico, execute as seguintes etapas:

1. Ative o BootBIOS no servidor HBA.
2. Para HBAs iSCSI, configure o IP do iniciador, o nome do nó iSCSI e o modo de inicialização do adaptador nas configurações do BIOS de inicialização.
3. Ao criar um grupo de iniciadores para iSCSI e/ou FC em um controlador de storage NetApp, adicione o iniciador HBA do servidor ao grupo. O iniciador HBA do servidor é o WWPN para o nome do nó FC HBA ou iSCSI para iSCSI HBA.
4. Crie um LUN no controlador de armazenamento NetApp com um ID LUN de 0 e associe-o ao grupo de iniciadores criado na etapa anterior. Este LUN serve como um LUN de arranque.
5. Restrinja o HBA a um único caminho para o LUN de inicialização. Caminhos adicionais podem ser adicionados depois que o Windows Server é instalado no LUN de inicialização para explorar o recurso multipathing.

6. Use o utilitário BootBIOS do HBA para configurar o LUN como um dispositivo de inicialização.
7. Reinicie o host e insira o utilitário BIOS do host.
8. Configure o BIOS do host para tornar o LUN de inicialização o primeiro dispositivo na ordem de inicialização.
9. A partir do ISO do Windows Server, inicie a instalação.
10. Quando a instalação perguntar "onde você deseja instalar o Windows?", clique em carregar driver na parte inferior da tela de instalação para iniciar a página Selecionar driver para instalar. Forneça o caminho do driver do dispositivo HBA baixado anteriormente e conclua a instalação do driver.
11. Agora, o LUN de inicialização criado anteriormente deve estar visível na página de instalação do Windows. Selecione o LUN de inicialização para instalação do Windows Server no LUN de inicialização e conclua a instalação.

Arranque a partir de SAN para máquina virtual

Para configurar a inicialização a partir de SAN para uma VM, execute as seguintes etapas:

Implantação

1. Ao criar um grupo de iniciadores para iSCSI ou FC em um controlador de armazenamento NetApp, adicione o IQN para iSCSI ou WWN para FC do servidor Hyper-V ao controlador.
2. Crie LUNs ou clones LUN no controlador de storage NetApp e associe-os ao grupo de iniciadores criado na etapa anterior. Essas LUNs servem como LUNs de inicialização das VMs.
3. Detecte os LUNs no servidor Hyper-V, coloque-os online e inicialize-os.
4. Coloque os LUNs offline.
5. Crie VMs com a opção Anexar um disco rígido virtual mais tarde na página conectar disco rígido virtual.
6. Adicione um LUN como um disco de passagem a uma VM.
 - a. Abra as definições da VM.
 - b. Clique em IDE Controller 0, selecione Hard Drive e clique em Add. Selecionar o controlador IDE 0 torna este disco o primeiro dispositivo de inicialização para a VM.
 - c. Selecione disco rígido físico nas opções disco rígido e selecione um disco da lista como um disco de passagem. Os discos são os LUNs configurados nas etapas anteriores.
7. Instale o Windows Server no disco de passagem.

Práticas recomendadas

- Certifique-se de que os LUNs estão offline. Caso contrário, o disco não pode ser adicionado como um disco de passagem a uma VM.
- Quando existirem vários LUNs, certifique-se de anotar o número de disco do LUN no gerenciamento de disco. Isso é necessário porque os discos listados para a VM são listados com o número do disco. Além disso, a seleção do disco como um disco de passagem para a VM é baseada neste número de disco.
- A NetApp recomenda evitar o agrupamento de NIC para NICs iSCSI.
- A NetApp recomenda o uso do ONTAP MPIO configurado no host para fins de armazenamento.

Provisionamento em ambientes SMB

O ONTAP fornece armazenamento nas resiliente e de alto desempenho para máquinas

virtuais Hyper-V usando o protocolo SMB3.

Quando um SVM é criado com o protocolo CIFS, um servidor CIFS é executado sobre o SVM que faz parte do domínio do Windows active Directory. Os compartilhamentos SMB podem ser usados para um diretório inicial e para hospedar cargas de trabalho Hyper-V e SQL Server. Os seguintes recursos do SMB 3,0 são suportados no ONTAP:

- Alças persistentes (compartilhamentos de arquivos disponíveis continuamente)
- Protocolo de testemunhas
- Failover de cliente em cluster
- Reconhecimento com escalabilidade horizontal
- ODX
- VSS remoto

Provisionamento de compartilhamentos SMB no Windows Server

Pré-requisitos

O uso do armazenamento NetApp em ambientes nas no Windows Server tem os seguintes requisitos:

- O cluster ONTAP tem uma licença CIFS válida.
- Pelo menos um agregado é criado.
- Uma interface lógica de dados (LIF) é criada e o LIF de dados deve ser configurado para CIFS.
- Um servidor de domínio do Windows active Directory configurado por DNS e credenciais de administrador de domínio estão presentes.
- Cada nó no cluster do NetApp é sincronizado com o controlador de domínio do Windows.

Controlador de domínio do active Directory

Um controlador de armazenamento NetApp pode ser ligado e operar dentro de um active Directory semelhante a um servidor Windows. Durante a criação do SVM, você pode configurar o DNS fornecendo os detalhes do nome de domínio e do servidor de nomes. O SVM tenta procurar um controlador de domínio do active Directory consultando o DNS para um servidor LDAP (Protocolo de Acesso ao active Directory/Lightweight Directory Access Protocol) de uma forma semelhante ao Windows Server.

Para que a configuração CIFS funcione corretamente, os controladores de armazenamento NetApp devem ser sincronizados com o controlador de domínio do Windows. A NetApp recomenda ter um desvio de tempo entre o controlador de domínio do Windows e o controlador de armazenamento NetApp de não mais de cinco minutos. É uma prática recomendada configurar o servidor NTP (Network Time Protocol) para que o cluster ONTAP sincronize com uma fonte de tempo externa. Para configurar o controlador de domínio do Windows como servidor NTP, execute o seguinte comando no cluster do ONTAP:

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

Implantação

1. Crie um novo SVM com o protocolo nas CIFS ativado. É possível criar um novo SVM com qualquer um dos seguintes métodos:

- Comandos CLI no ONTAP
 - System Manager
 - O kit de ferramentas do NetApp PowerShell
2. Configurar o protocolo CIFS
 - a. Forneça o nome do servidor CIFS.
 - b. Forneça o ativo Directory ao qual o servidor CIFS deve ser conectado. Você deve ter as credenciais de administrador de domínio para ingressar no servidor CIFS ao ativo Directory.
 3. Atribua o SVM com LIFs em cada nó de cluster.
 4. Inicie o serviço CIFS no SVM.
 5. Crie um volume com o estilo de segurança NTFS a partir do agregado.
 6. Crie uma qtree no volume (opcional).
 7. Crie compartilhamentos que correspondam ao volume ou diretório de qtree para que possam ser acessados a partir do Windows Server. Selecione Ativar disponibilidade contínua para Hyper-V durante a criação do compartilhamento se o compartilhamento for usado para armazenamento Hyper-V. Com isso, é possível obter alta disponibilidade para compartilhamentos de arquivos.
 8. Edite o compartilhamento criado e modifique as permissões conforme necessário para acessar o compartilhamento. As permissões para o compartilhamento SMB devem ser configuradas para conceder acesso às contas de computador de todos os servidores que acessam esse compartilhamento.

Integração de host

O protocolo nas CIFS é nativamente integrado ao ONTAP. Portanto, o Windows Server não requer nenhum software cliente adicional para acessar dados no ONTAP. Um controlador de armazenamento NetApp aparece na rede como um servidor de arquivos nativo e suporta autenticação do Microsoft ativo Directory.

Para detetar o compartilhamento CIFS criado anteriormente com o Windows Server, execute as seguintes etapas:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Vá para run.exe e digite o caminho completo do compartilhamento CIFS criado para acessar o compartilhamento.
3. Para mapear permanentemente a partilha no Windows Server, clique com o botão direito do rato neste PC, clique em Map Network Drive (Mapear unidade de rede) e forneça o caminho da partilha CIFS.
4. Algumas tarefas de gerenciamento CIFS podem ser executadas usando o MMC (Console de Gerenciamento da Microsoft). Antes de executar essas tarefas, você deve conectar o MMC ao armazenamento ONTAP usando os comandos de menu do MMC.
 - a. Para abrir o MMC no Windows Server, clique em Gerenciamento de computador na seção Ferramentas do Gerenciador de servidores.
 - b. Clique em mais ações e conecte-se a outro computador, que abre a caixa de diálogo Selecionar computador.
 - c. Digite o nome do servidor CIFS ou o endereço IP do SVM LIF para se conectar ao servidor CIFS.
 - d. Expanda Ferramentas do sistema e pastas compartilhadas para exibir e gerenciar arquivos abertos, sessões e compartilhamentos.

Práticas recomendadas

- Para confirmar que não há tempo de inatividade quando um volume é movido de um nó para outro ou em caso de falha de nó, a NetApp recomenda que você ative a opção de disponibilidade contínua no compartilhamento de arquivos.
- Ao provisionar VMs para um ambiente Hyper-V-over-SMB, a NetApp recomenda que você ative a descarga de cópia no sistema de storage. Isso reduz o tempo de provisionamento das VMs.
- Se o cluster de storage hospedar vários workloads SMB, como SQL Server, Hyper-V e servidores CIFS, a NetApp recomenda hospedar diferentes workloads SMB em SVMs separadas em agregados separados. Essa configuração é benéfica porque cada uma dessas cargas de trabalho garante uma rede de storage exclusiva e layouts de volume.
- A NetApp recomenda a conexão de hosts Hyper-V e o armazenamento ONTAP com uma rede 10GB, se houver um disponível. No caso da conectividade de rede 1GBG, a NetApp recomenda a criação de um grupo de interfaces composto por várias portas 1GBG.
- Ao migrar VMs de um compartilhamento SMB 3,0 para outro, a NetApp recomenda habilitar a funcionalidade de descarga de cópia CIFS no sistema de storage para que a migração seja mais rápida.

Coisas para se lembrar

- Quando você provisiona volumes para ambientes SMB, os volumes devem ser criados com o estilo de segurança NTFS.
- As configurações de tempo nos nós do cluster devem ser configuradas de acordo. Use o NTP se o servidor CIFS do NetApp tiver de participar no domínio do ativo Directory do Windows.
- As alças persistentes funcionam apenas entre nós em um par de HA.
- O protocolo testemunha funciona apenas entre nós em um par de HA.
- Compartilhamentos de arquivo continuamente disponíveis são compatíveis apenas para cargas de trabalho Hyper-V e SQL Server.
- O multicanal SMB é suportado a partir do ONTAP 9.4 em diante.
- RDMA não é suportado.
- Refs não é suportado.

Provisionamento de compartilhamentos SMB no Nano Server

Nano servidor não requer software cliente adicional para acessar dados no compartilhamento CIFS em um controlador de armazenamento NetApp.

Para copiar arquivos do Nano Server para um compartilhamento CIFS, execute os seguintes cmdlets no servidor 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
```

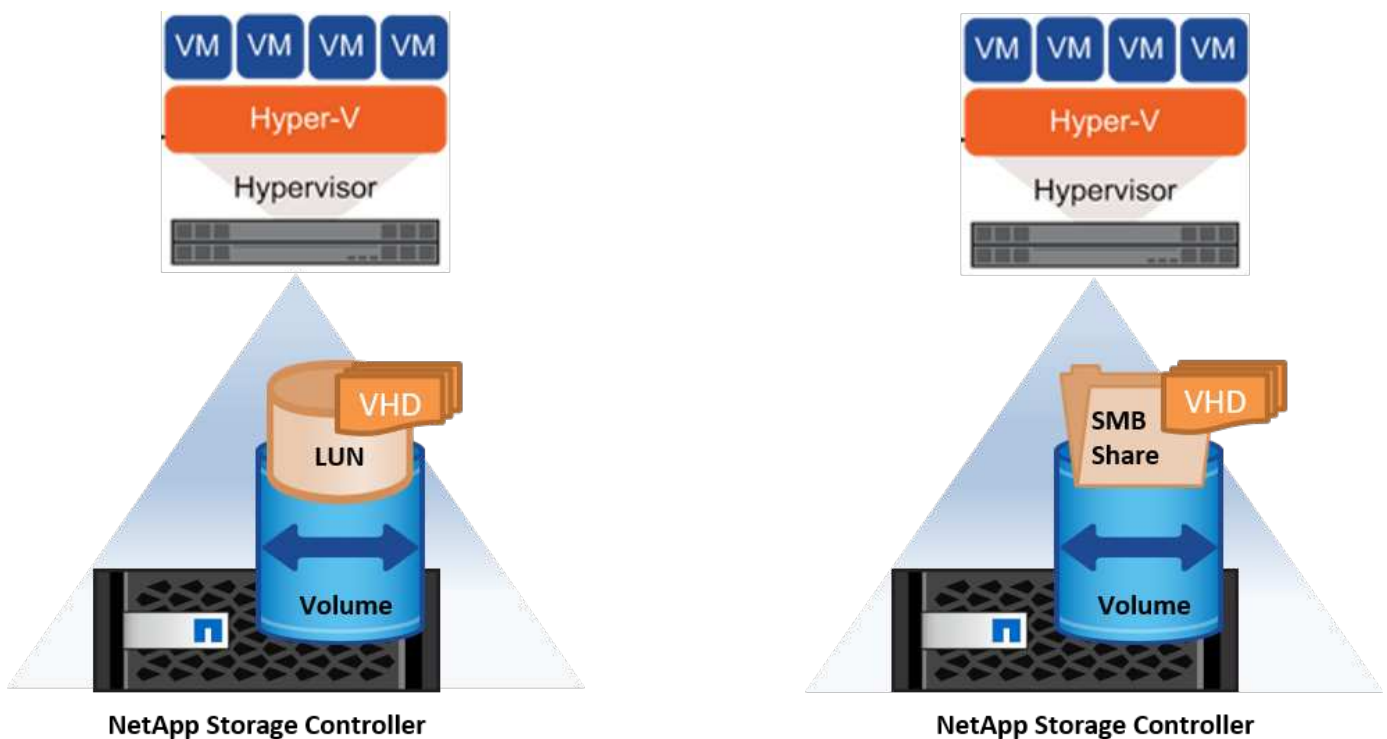
```
Copy-Item -FromSession $s -Path C:\Windows\Logs\DISM\dism.log
-Destination \\cifsshare
* `cifsshare` É o compartilhamento CIFS na controladora de storage do
NetApp.
* Para copiar arquivos para o Nano Server, execute o seguinte cmdlet:
```

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

Para copiar todo o conteúdo de uma pasta, especifique o nome da pasta e use o parâmetro -Recurse no final do cmdlet.

Infraestrutura de storage Hyper-V no NetApp

Uma infraestrutura de storage Hyper-V pode ser hospedada em sistemas de storage ONTAP. O armazenamento para o Hyper-V para armazenar os arquivos da VM e seus discos pode ser fornecido usando LUNs NetApp ou compartilhamentos NetApp CIFS, como mostrado na figura a seguir.



Armazenamento Hyper-V em LUNs NetApp

- Provisione um LUN NetApp na máquina do servidor Hyper-V. Para obter mais informações, consulte a seção "[Provisionamento em ambientes SAN](#)".
- Abra o Gerenciador do Hyper-V na seção Ferramentas do Gerenciador de servidores.
- Selecione o servidor Hyper-V e clique em Configurações Hyper-V.
- Especifique a pasta padrão para armazenar a VM e seu disco como LUN. Isso define o caminho padrão

como LUN para o armazenamento Hyper-V. Se você quiser especificar o caminho explicitamente para uma VM, então você pode fazê-lo durante a criação da VM.

Armazenamento Hyper-V no NetApp CIFS

Antes de iniciar as etapas listadas nesta seção, revise a seção ["Provisionamento em ambientes SMB"](#). Para configurar o armazenamento Hyper-V no compartilhamento CIFS do NetApp, execute as seguintes etapas:

1. Abra o Gerenciador do Hyper-V na seção Ferramentas do Gerenciador de servidores.
2. Selecione o servidor Hyper-V e clique em Configurações Hyper-V.
3. Especifique a pasta padrão para armazenar a VM e seu disco como o compartilhamento CIFS. Isso define o caminho padrão como o compartilhamento CIFS para o armazenamento Hyper-V. Se você quiser especificar o caminho explicitamente para uma VM, então você pode fazê-lo durante a criação da VM.

Cada VM no Hyper-V pode, por sua vez, ser fornecida com os LUNs e compartilhamentos CIFS do NetApp que foram fornecidos ao host físico. Este procedimento é o mesmo que para qualquer host físico. Os métodos a seguir podem ser usados para provisionar storage a uma VM:

- Adicionando um LUN de storage usando o iniciador FC na VM
- Adicionar um LUN de armazenamento utilizando o iniciador iSCSI na VM
- Adicionando um disco físico de passagem a uma VM
- Adicionando VHD/VHDX a uma VM do host

Práticas recomendadas

- Quando uma VM e seus dados são armazenados no storage NetApp, a NetApp recomenda a execução da deduplicação do NetApp no nível do volume em intervalos regulares. Essa prática resulta em economia significativa de espaço quando VMs idênticas são hospedadas em um compartilhamento CSV ou SMB. A deduplicação é executada no controlador de storage e não afeta o sistema host e o desempenho da VM.
- Ao usar iSCSI LUNs para Hyper-V, certifique-se de ativar `iSCSI Service (TCP-In) for Inbound` e `iSCSI Service (TCP-Out) for Outbound` nas configurações de firewall no host Hyper-V. Isso permite que o tráfego iSCSI passe de e para o host Hyper-V e para o controlador NetApp.
- A NetApp recomenda desmarcar a opção permitir que o sistema operacional de Gerenciamento compartilhe este adaptador de rede para o switch virtual Hyper-V. Isso cria uma rede dedicada para as VMs.

Coisas para se lembrar

- O provisionamento de uma VM usando Fibre Channel virtual requer um `N_Port ID Virtualization` e `FC HBA` habilitado. Suporte para um máximo de quatro portas FC.
- Se o sistema host for configurado com várias portas FC e apresentado à VM, o MPIO deverá ser instalado na VM para habilitar o multipathing.
- Os discos pass-through não podem ser provisionados para o host se o MPIO estiver sendo usado nesse host, porque os discos pass-through não suportam MPIO.
- O disco usado para arquivos VHD/VHDX deve usar a formatação 64K para alocação.

Leitura adicional

- Para obter informações sobre HBAs FC, consulte ["Matriz de interoperabilidade do NetApp"](#).
- Para obter mais informações sobre o Fibre Channel virtual, consulte a página da Microsoft ["Visão geral do](#)

Transferência de dados descarregada

O Microsoft ODX, também conhecido como descarga de cópia, permite transferências diretas de dados dentro de um dispositivo de armazenamento ou entre dispositivos de armazenamento compatíveis sem transferir os dados através do computador host. O ONTAP oferece suporte ao recurso ODX para protocolos CIFS e SAN. O ODX pode potencialmente melhorar o desempenho se as cópias estiverem dentro do mesmo volume, reduzir a utilização da CPU e da memória no cliente e reduzir a utilização da largura de banda de e/S.

Com o ODX, é mais rápido e eficiente copiar arquivos dentro dos compartilhamentos SMB, dentro dos LUNs e entre os compartilhamentos SMB e LUNs se estiver no mesmo volume. Essa abordagem é mais útil em um cenário para o qual várias cópias da imagem dourada de um SO (VHD/VHDX) são necessárias dentro do mesmo volume. Várias cópias da mesma imagem dourada podem ser feitas em menos tempo se as cópias estiverem dentro do mesmo volume. O ODX também é aplicado na migração ao vivo de armazenamento Hyper-V para a movimentação do armazenamento de VM.

Se a cópia estiver em volumes, talvez não haja ganhos significativos de desempenho em comparação com as cópias baseadas em host.

Para ativar o recurso ODX no CIFS, execute os seguintes comandos CLI no controlador de armazenamento NetApp:

1. Ative o ODX para CIFS. defina o nível de privilégio para o grupo de diagnóstico::> definir - diagnóstico de privilégios

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

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

2. Para ativar o recurso ODX na SAN, execute os seguintes comandos CLI no controlador de armazenamento NetApp: N defina o nível de privilégio para o cluster de diagnóstico::> defina o diagnóstico de privilégio

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

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

Coisas para se lembrar

- Para CIFS, o ODX só está disponível quando o cliente e o servidor de armazenamento suportam SMB 3,0

e o recurso ODX.

- Para ambientes SAN, o ODX só está disponível quando o cliente e o servidor de armazenamento suportam o recurso ODX.

Leitura adicional

Para obter informações sobre o ODX, "[Melhorando o desempenho da cópia remota da Microsoft](#)" consulte e "[Transferências de dados descarregadas da Microsoft](#)".

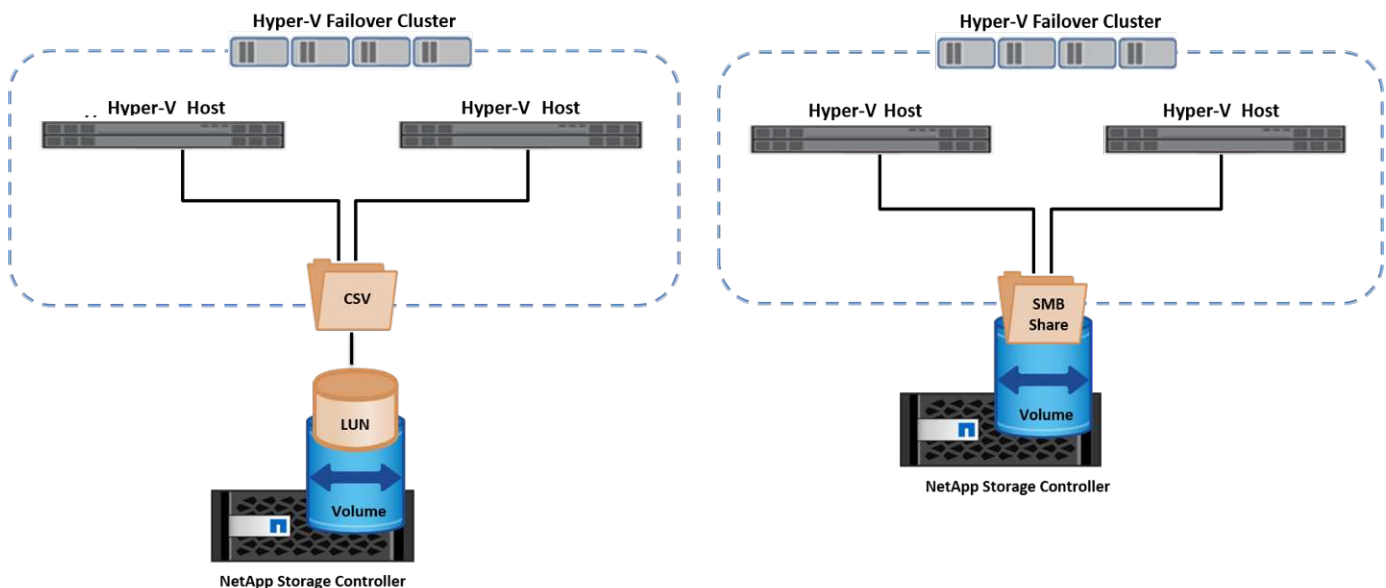
Cluster Hyper-V: Alta disponibilidade e escalabilidade para máquinas virtuais

Os clusters de failover fornecem alta disponibilidade e escalabilidade para servidores Hyper-V. Um cluster de failover é um grupo de servidores Hyper-V independentes que trabalham em conjunto para aumentar a disponibilidade e a escalabilidade das VMs.

Os servidores em cluster do Hyper-V (chamados nós) são conectados pela rede física e pelo software de cluster. Esses nós usam o storage compartilhado para armazenar os arquivos de VM, que incluem configuração, arquivos de disco rígido virtual (VHD) e cópias Snapshot. O storage compartilhado pode ser um compartilhamento NetApp SMB/CIFS ou um CSV em cima de um LUN NetApp, como mostrado abaixo. Esse storage compartilhado fornece um namespace consistente e distribuído que pode ser acessado simultaneamente por todos os nós do cluster. Portanto, se um nó falhar no cluster, o outro nó fornece serviço por um processo chamado failover. Os clusters de failover podem ser gerenciados usando o snap-in Gerenciador de Cluster de failover e os cmdlets do Windows PowerShell de clustering de failover.

Volumes partilhados de cluster

Os CSVs permitem que vários nós em um cluster de failover tenham simultaneamente acesso de leitura/gravação ao mesmo LUN NetApp que é provisionado como um volume NTFS ou refs. Com os CSVs, as funções em cluster podem fazer failover rapidamente de um nó para outro sem exigir uma alteração na propriedade da unidade ou desmontar e montar novamente um volume. Os CSVs também simplificam o gerenciamento de um número potencialmente grande de LUNs em um cluster de failover. Os CSVs fornecem um sistema de arquivos em cluster de uso geral que está em camadas acima de NTFS ou refs.



Práticas recomendadas

- A NetApp recomenda desativar a comunicação do cluster na rede iSCSI para impedir que a comunicação interna do cluster e o tráfego CSV fluam na mesma rede.
- A NetApp recomenda ter caminhos de rede redundantes (vários switches) para fornecer resiliência e QoS.

Coisas para se lembrar

- Os discos usados para CSV devem ser particionados com NTFS ou refs. Os discos formatados com FAT ou FAT32 não podem ser usados para um CSV.
- Os discos usados para CSVs devem usar a formatação 64K para alocação.

Leitura adicional

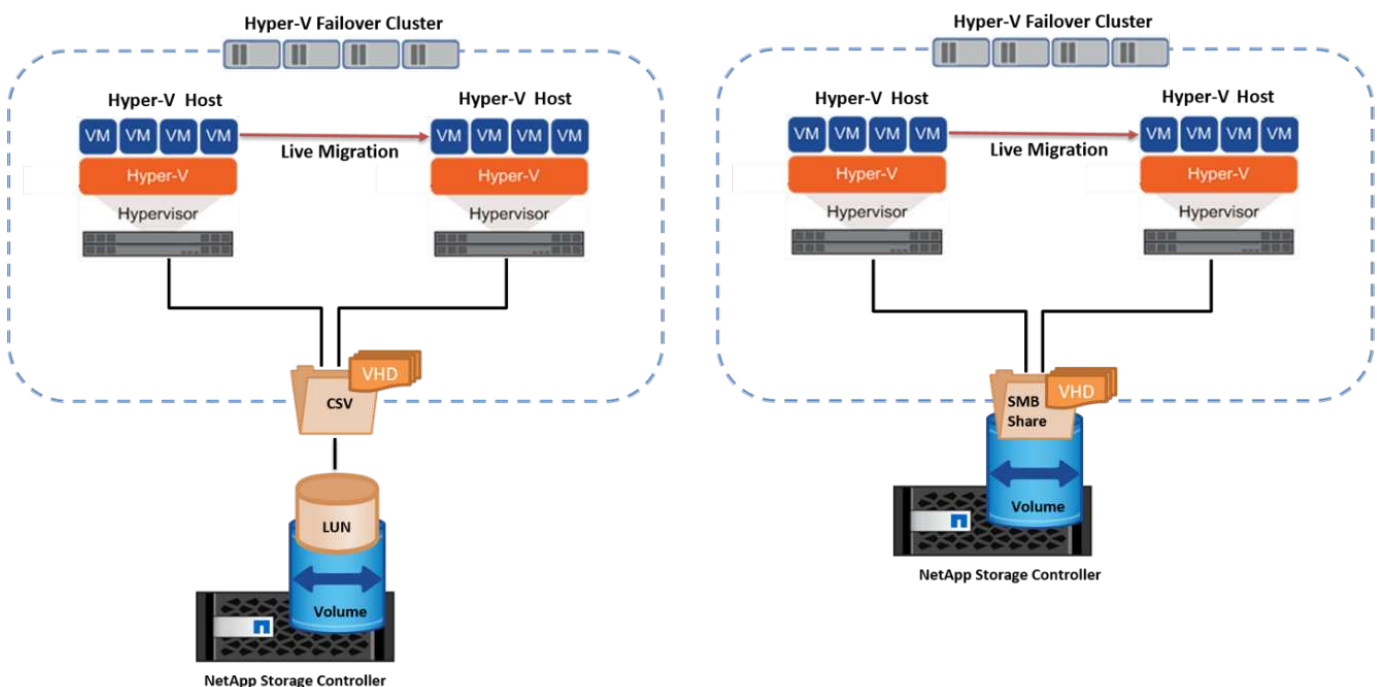
Para obter informações sobre como implantar um cluster Hyper-V, consulte o Apêndice B: "[Implantar o Hyper-V Cluster](#)".

Migração do Hyper-V Live: Migração de VMs

Às vezes, é necessário, durante a vida útil das VMs, movê-las para um host diferente no cluster do Windows. Isso pode ser necessário se o host estiver ficando sem recursos do sistema ou se o host for necessário para reinicializar por motivos de manutenção. Da mesma forma, pode ser necessário mover uma VM para um compartilhamento de LUN ou SMB diferente. Isso pode ser necessário se o LUN ou compartilhamento atual estiver ficando sem espaço ou produzindo desempenho inferior ao esperado. A migração em tempo real do Hyper-V move VMs em execução de um servidor Hyper-V físico para outro sem efeito na disponibilidade da VM para os usuários. Você pode migrar VMs em tempo real entre servidores Hyper-V que fazem parte de um cluster de failover ou entre servidores Hyper-V independentes que não fazem parte de nenhum cluster.

Migração ao vivo em um ambiente em cluster

As VMs podem ser movidas de forma otimizada entre os nós de um cluster. A migração de VM é instantânea porque todos os nós do cluster compartilham o mesmo storage e têm acesso à VM e ao disco. A figura a seguir mostra a migração ao vivo em um ambiente em cluster.



Prática recomendada

- Tenha uma porta dedicada para o tráfego de migração em tempo real.
- Tenha uma rede de migração ao vivo de host dedicada para evitar problemas relacionados à rede durante a migração.

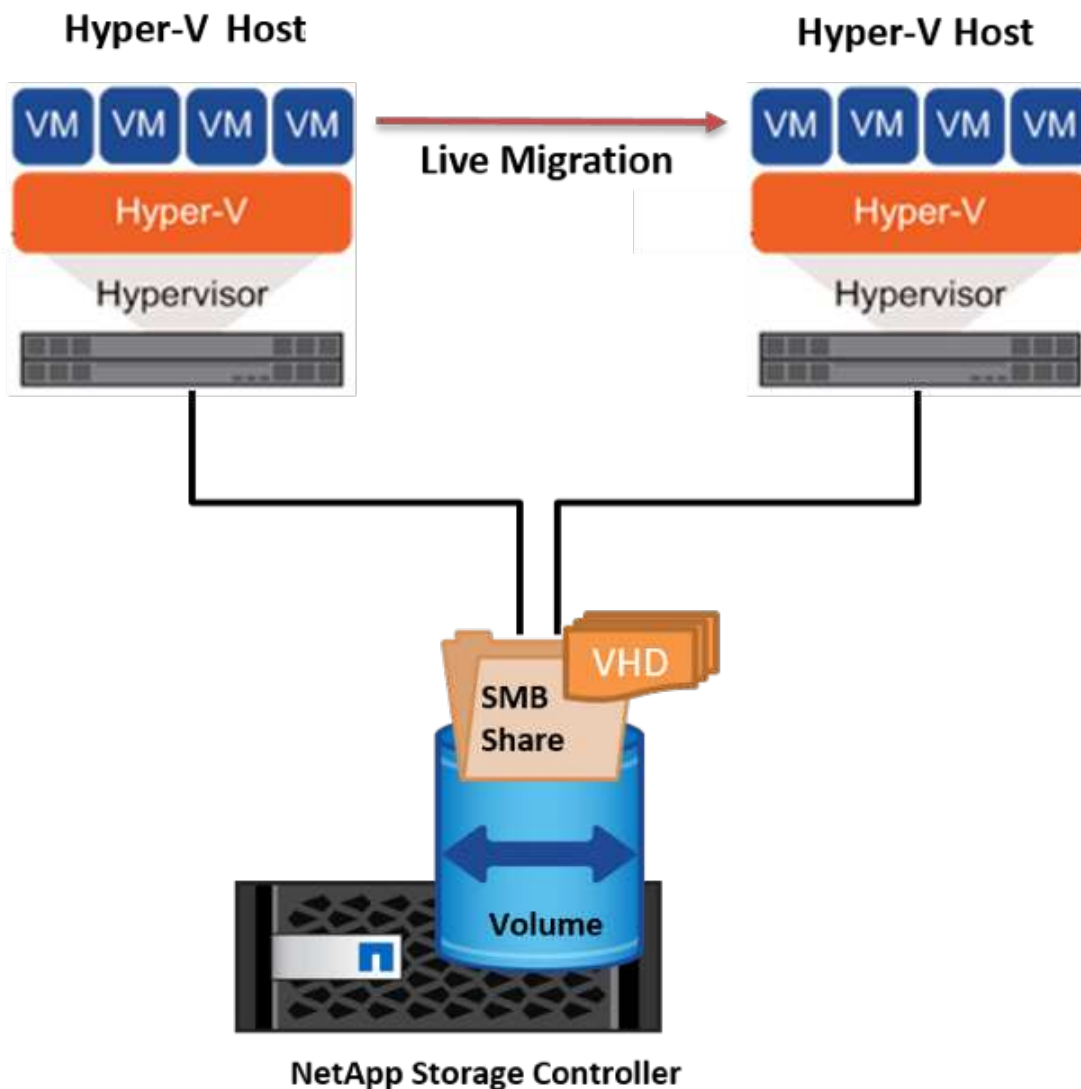
Leitura adicional

Para obter informações sobre como implantar a migração ao vivo em um ambiente em cluster, "[Apêndice C: Implantar a migração do Hyper-V Live em um ambiente em cluster](#)" consulte .

Migração ao vivo fora de um ambiente em cluster

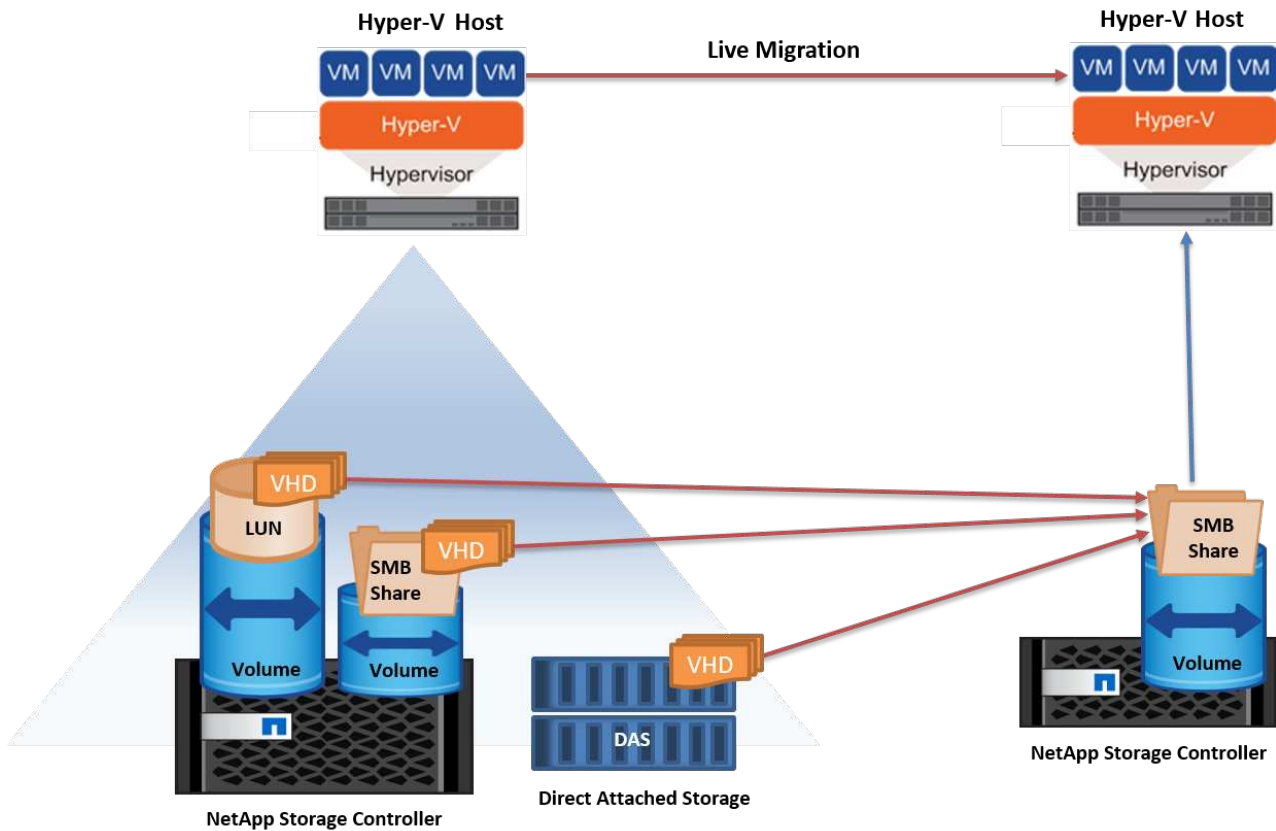
É possível migrar uma VM em tempo real entre dois servidores Hyper-V independentes e não agrupados. Esse processo pode usar migração ao vivo compartilhada ou compartilhada.

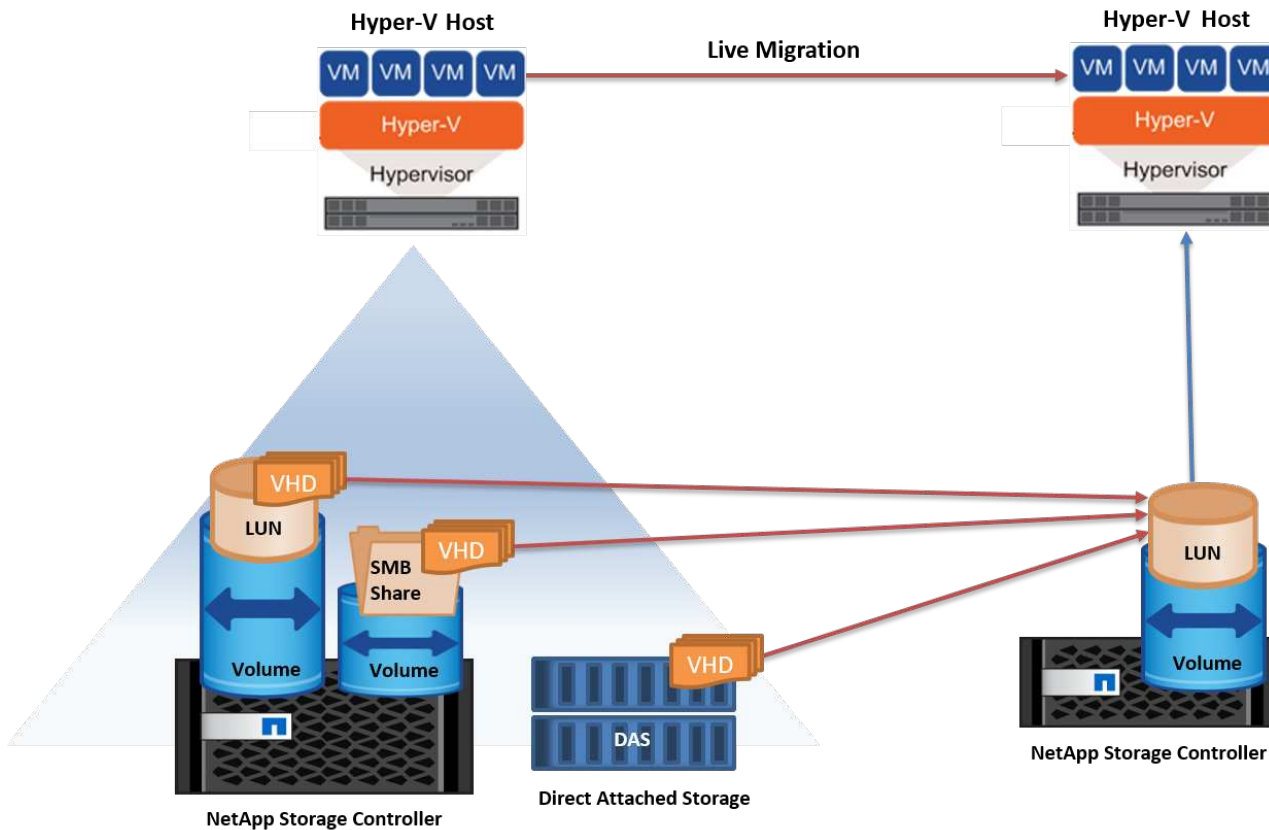
- Na migração ao vivo compartilhada, a VM é armazenada em um compartilhamento SMB. Portanto, quando você migra uma VM ativa, o armazenamento da VM permanece no compartilhamento central de SMB para acesso instantâneo pelo outro nó, como mostrado abaixo.



- Na migração compartilhada nada ao vivo, cada servidor Hyper-V tem seu próprio armazenamento local

(pode ser um compartilhamento SMB, um LUN ou DAS), e o armazenamento da VM é local para seu servidor Hyper-V. Quando uma VM é migrada em tempo real, o armazenamento da VM é espelhado para o servidor de destino através da rede cliente e, em seguida, a VM é migrada. A VM armazenada em DAS, um LUN ou um compartilhamento SMB/CIFS pode ser movida para um compartilhamento SMB/CIFS no outro servidor Hyper-V, como mostrado na figura a seguir. Ele também pode ser movido para um LUN, como mostrado na segunda figura.





Leitura adicional

Para obter informações sobre como implantar a migração ao vivo fora de um ambiente em cluster, ["Apêndice D: Implantar a migração do Hyper-V Live fora de um ambiente em cluster"](#) consulte .

Migração ao vivo do Hyper-V Storage

Durante a vida útil de uma VM, talvez seja necessário mover o armazenamento da VM (VHD/VHDX) para um compartilhamento de LUN ou SMB diferente. Isso pode ser necessário se o LUN ou compartilhamento atual estiver ficando sem espaço ou produzindo desempenho inferior ao esperado.

O LUN ou o compartilhamento que atualmente hospeda a VM podem ficar sem espaço, ser reutilizados ou fornecer desempenho reduzido. Nessas circunstâncias, a VM pode ser movida sem tempo de inatividade para outro LUN ou compartilhar em um volume, agregado ou cluster diferente. Esse processo será mais rápido se o sistema de storage tiver funcionalidades de descarga de cópia. Os sistemas de storage NetApp são habilitados por padrão para descarga de cópia em ambientes CIFS e SAN.

O recurso ODX executa cópias de arquivo completo ou subarquivo entre dois diretórios residentes em servidores remotos. Uma cópia é criada copiando dados entre os servidores (ou o mesmo servidor se ambos os arquivos de origem e de destino estiverem no mesmo servidor). A cópia é criada sem que o cliente leia os dados da origem ou escreva para o destino. Esse processo reduz o uso de processador e memória para o cliente ou servidor e minimiza a largura de banda de e/S de rede. A cópia é mais rápida se estiver dentro do mesmo volume. Se a cópia estiver em volumes, talvez não haja ganhos significativos de desempenho em comparação com as cópias baseadas em host. Antes de prosseguir com uma operação de cópia no host, confirme se as configurações de descarga de cópia estão configuradas no sistema de storage.

Quando a migração ativa do armazenamento de VM é iniciada a partir de um host, a origem e o destino são identificados e a atividade de cópia é descarregada para o sistema de storage. Como a atividade é realizada pelo sistema de armazenamento, há um uso insignificante da CPU, memória ou rede do host.

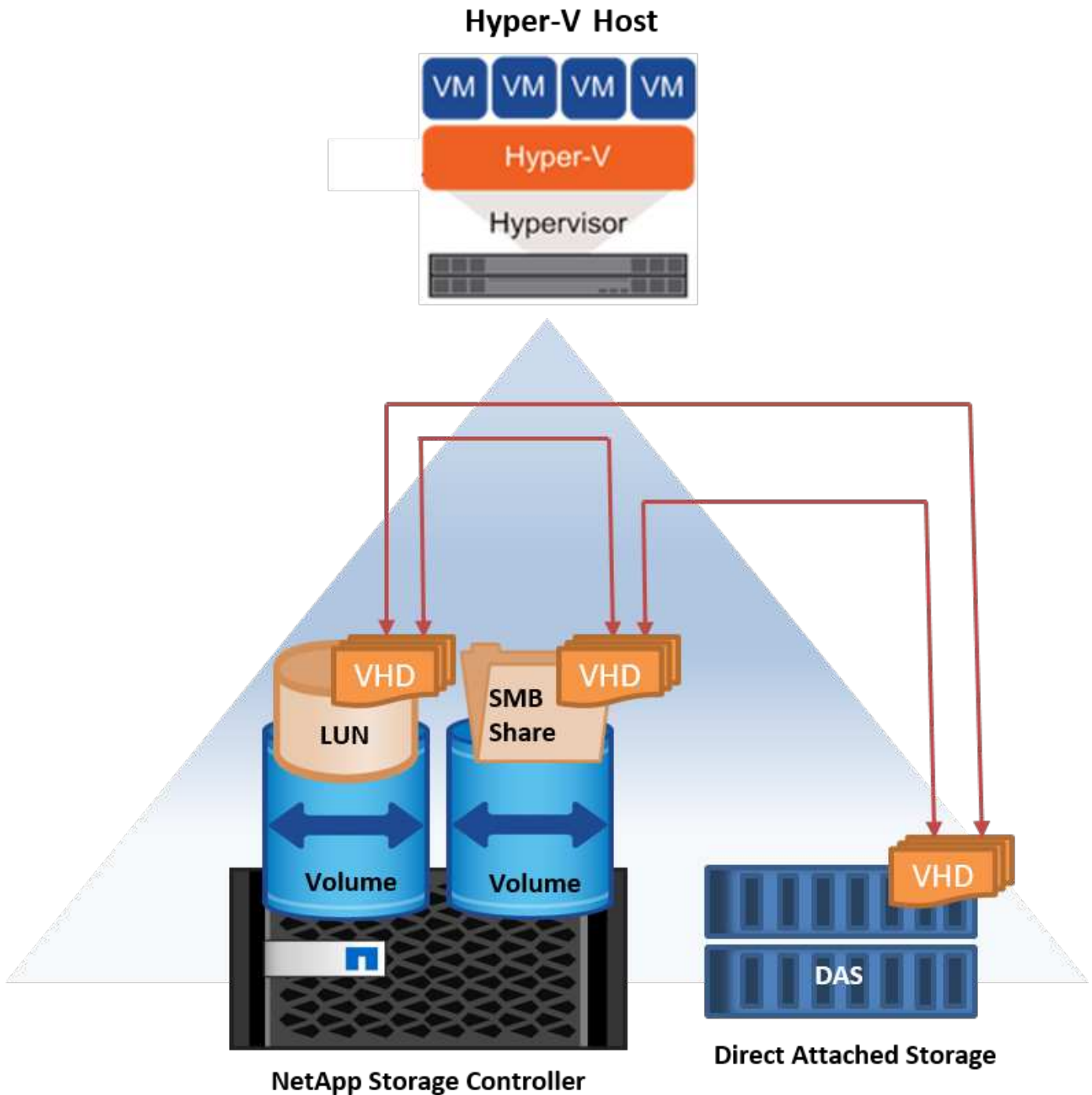
Os controladores de armazenamento NetApp suportam os seguintes cenários diferentes de ODX:

- **IntraSVM.** Os dados pertencem ao mesmo SVM:
- **Intradvolume, intranode.** Os arquivos de origem e destino ou LUNs residem no mesmo volume. A cópia é realizada com a tecnologia de arquivos FlexClone, que fornece benefícios adicionais de desempenho de cópia remota.
- **Intervolume, intranode.** Os arquivos de origem e destino ou LUNs estão em volumes diferentes que estão no mesmo nó.
- **Intervolume, internós.** Os arquivos de origem e destino ou LUNs estão em volumes diferentes localizados em nós diferentes.
- **InterSVM.** Os dados pertencem a diferentes SVMs.
- **Intervolume, intranode.** Os arquivos de origem e destino ou LUNs estão em volumes diferentes que estão no mesmo nó.
- **Intervolume, internós.** Os arquivos de origem e destino ou LUNs estão em volumes diferentes que estão em nós diferentes.
- **Intercluster.** A partir do ONTAP 9.0, o ODX também é compatível com transferências LUN entre clusters em ambientes SAN. O ODX é suportado apenas para protocolos SAN, não para SMB.

Após a conclusão da migração, as políticas de backup e replicação devem ser reconfiguradas para refletir o novo volume que contém as VMs. Quaisquer backups anteriores que foram feitos não podem ser usados.

O armazenamento de VM (VHD/VHDX) pode ser migrado entre os seguintes tipos de armazenamento:

- DAS e o compartilhamento SMB
- DAS e LUN
- Um compartilhamento SMB e um LUN
- Entre LUNs
- Entre compartilhamentos SMB



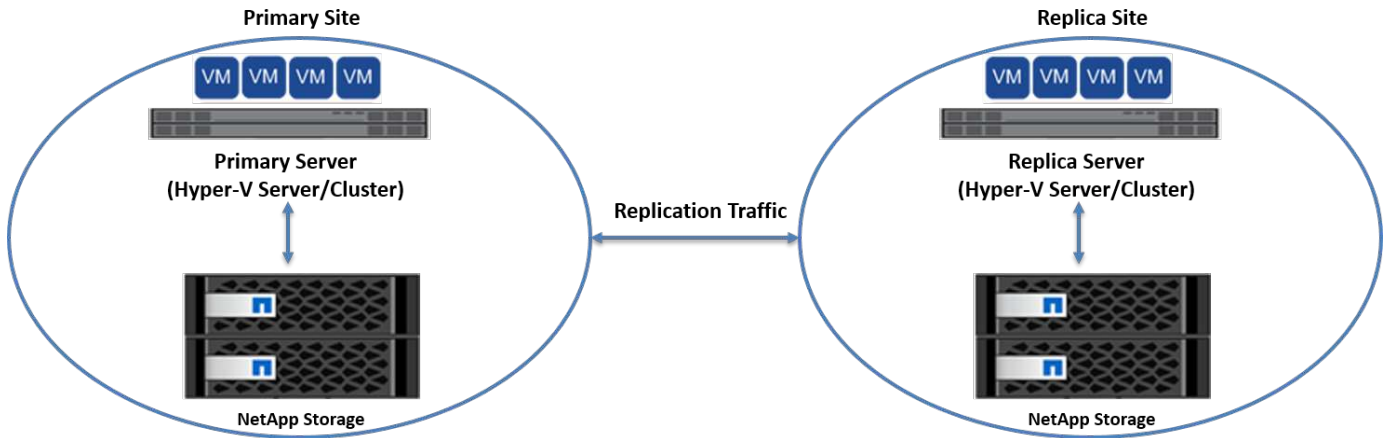
Leitura adicional

Para obter informações sobre como implantar a migração ao vivo de storage, ["Apêndice e: Implantar a migração ao vivo do Hyper-V Storage"](#) consulte .

Réplica do Hyper-V: Recuperação de desastres para máquinas virtuais

A réplica do Hyper-V replica as VMs do Hyper-V de um local primário para VMs de réplica em um local secundário, fornecendo, de forma assíncrona, recuperação de desastres para as VMs. O servidor Hyper-V no site principal que hospeda as VMs é conhecido como servidor primário; o servidor Hyper-V no site secundário que recebe VMs replicadas é conhecido como servidor de réplica. Um cenário de exemplo de réplica do Hyper-V é mostrado na figura a seguir. Você pode usar a réplica do Hyper-V para VMs entre servidores Hyper-

V que fazem parte de um cluster de failover ou entre servidores Hyper-V independentes que não fazem parte de nenhum cluster.



Replicação

Depois que a réplica do Hyper-V estiver habilitada para uma VM no servidor primário, a replicação inicial cria uma VM idêntica no servidor de réplica. Após a replicação inicial, a réplica do Hyper-V mantém um arquivo de log para os VHDs da VM. O ficheiro de registo é reproduzido na ordem inversa à réplica VHD de acordo com a frequência de replicação. Esse log e o uso da ordem inversa garantem que as alterações mais recentes sejam armazenadas e replicadas assincronamente. Se a replicação não ocorrer de acordo com a frequência esperada, é emitido um alerta.

Replicação estendida

A réplica do Hyper-V suporta replicação estendida na qual um servidor de réplica secundário pode ser configurado para recuperação de desastres. Um servidor de réplica secundário pode ser configurado para que o servidor de réplica receba as alterações nas VMs de réplica. Em um cenário de replicação estendida, as alterações nas VMs primárias no servidor primário são replicadas para o servidor de réplica. Em seguida, as alterações são replicadas para o servidor de réplica estendido. As VMs podem ser falhadas para o servidor de réplica estendido somente quando os servidores principal e de réplica estiverem inativos.

Failover

O failover não é automático; o processo deve ser acionado manualmente. Existem três tipos de failover:

- **Failover de teste.** Esse tipo é usado para verificar se uma VM de réplica pode ser iniciada com sucesso no servidor de réplica e é iniciada na VM de réplica. Esse processo cria uma VM de teste duplicada durante o failover e não afeta a replicação regular da produção.
- **Failover planejado.** Esse tipo é usado para fazer failover de VMs durante a inatividade planejada ou as interrupções esperadas. Esse processo é iniciado na VM principal, que deve ser desligado no servidor principal antes de um failover planejado ser executado. Após o failover da máquina, a réplica Hyper-V inicia a réplica VM no servidor de réplica.
- **Failover não planejado.** Este tipo é utilizado quando ocorrem interrupções inesperadas. Esse processo é iniciado na VM de réplica e deve ser usado somente se a máquina principal falhar.

Recuperação

Ao configurar a replicação para uma VM, você pode especificar o número de pontos de recuperação. Os pontos de recuperação representam pontos no tempo a partir dos quais os dados podem ser recuperados de uma máquina replicada.

Leitura adicional

- Para obter informações sobre como implantar a réplica do Hyper-V fora de um ambiente em cluster, consulte a seção ["Implantar a réplica do Hyper-V fora de um ambiente em cluster"](#).
- Para obter informações sobre como implantar a réplica do Hyper-V em um ambiente em cluster, consulte a seção ["Implantar a réplica do Hyper-V em um ambiente em cluster"](#).

Eficiência de storage

O ONTAP fornece eficiência de storage líder do setor para ambientes virtualizados, incluindo o Microsoft Hyper-V. o NetApp também oferece programas de garantia de eficiência de storage.

Deduplicação da NetApp

A deduplicação do NetApp funciona removendo blocos duplicados no nível do volume de storage, armazenando apenas uma cópia física, independentemente de quantas cópias lógicas estejam presentes. Portanto, a deduplicação cria a ilusão de que existem inúmeras cópias desse bloco. A deduplicação remove automaticamente blocos de dados duplicados em um nível de bloco de 4KB U em todo o volume. Esse processo recupera o storage para obter economia de espaço e potencial de desempenho, reduzindo o número de gravações físicas no disco. A deduplicação pode proporcionar mais de 70% de economia de espaço em ambientes Hyper-V.

Thin Provisioning

O thin Provisioning é uma maneira eficiente de provisionar storage porque o storage não é pré-alocado de antemão. Em outras palavras, quando um volume ou LUN é criado usando thin Provisioning, o espaço no sistema de storage não é usado. O espaço permanece não utilizado até que os dados sejam gravados no LUN ou volume e apenas o espaço necessário para armazenar os dados é usado. A NetApp recomenda ativar o provisionamento de thin Provisioning no volume e desativar a reserva de LUN.

Qualidade do serviço

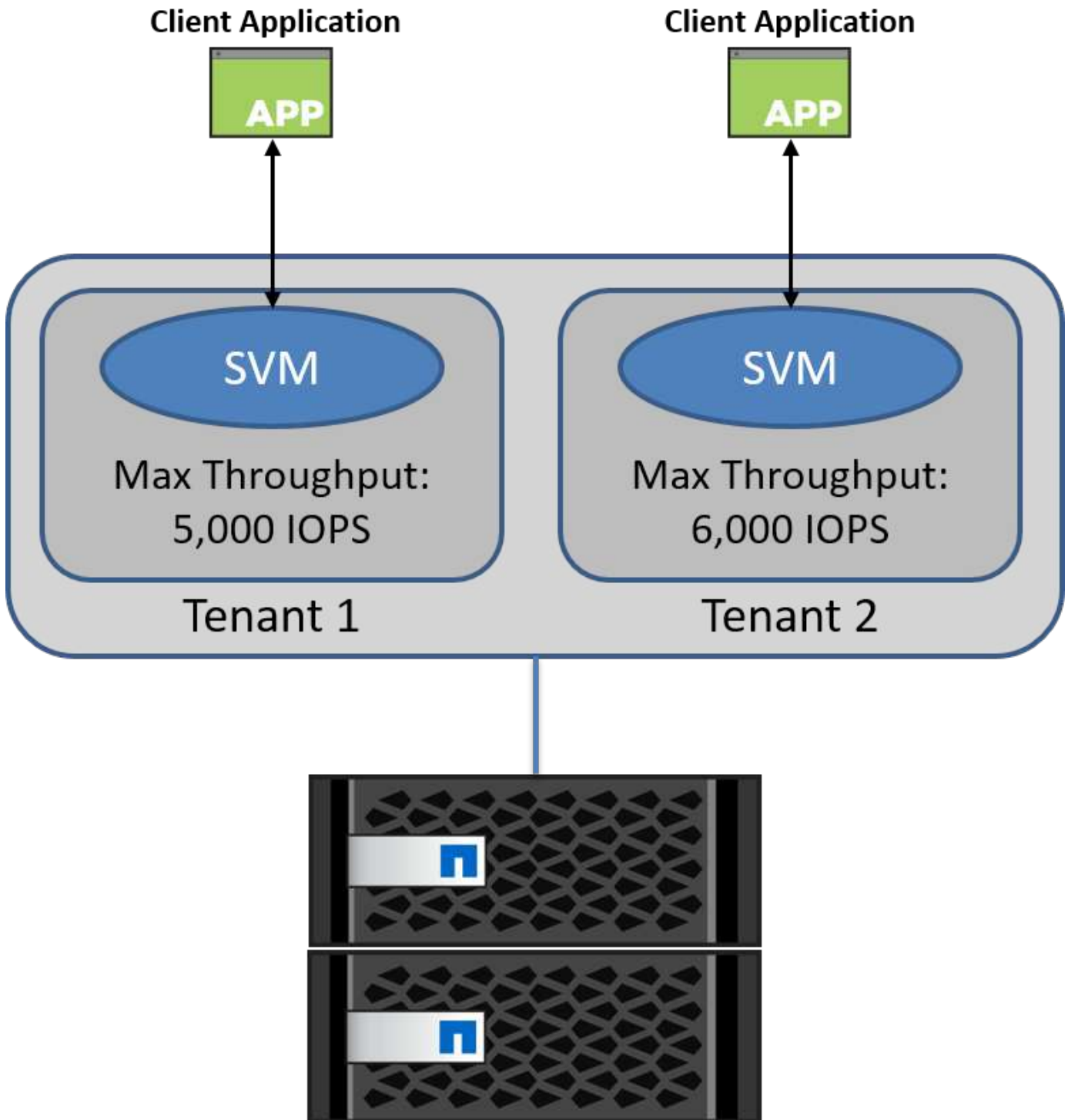
A QoS de storage no Clustered ONTAP permite agrupar objetos de storage e definir limites de taxa de transferência no grupo. A QoS de storage pode ser usada para limitar a taxa de transferência a workloads e monitorar a performance de workload. Com isso, um administrador de storage pode separar workloads por organização, aplicação, unidade de negócios ou ambientes de produção ou desenvolvimento.

Em ambientes empresariais, a QoS de storage ajuda a conseguir o seguinte:

- Impede que os workloads do usuário afetem uns aos outros.
- Protege aplicações críticas que têm tempos de resposta específicos que precisam ser atendidos em AMBIENTES DE TI como serviço (ITaaS).
- Impede que os inquilinos se afetem.
- Evita a degradação do desempenho com a adição de cada novo inquilino.

O QoS permite limitar a quantidade de e/S enviada para um SVM, um volume flexível, um LUN ou um arquivo. A e/S pode ser limitada pelo número de operações ou pela taxa de transferência bruta.

A figura a seguir ilustra o SVM com sua própria política de QoS que impõe um limite máximo de taxa de transferência.



Para configurar um SVM com sua própria política de QoS e monitorar o grupo de políticas, execute os seguintes comandos no cluster do 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
```

Segurança

O ONTAP fornece um sistema de armazenamento seguro para o sistema operacional Windows.

Windows Defender Antivirus

O Windows Defender é um software antimalware instalado e ativado no Windows Server por padrão. Este software protege ativamente o Windows Server contra malware conhecido e pode atualizar regularmente definições de antimalware através do Windows Update. Os LUNs e compartilhamentos SMB do NetApp podem ser digitalizados usando o Windows Defender.

Leitura adicional

Para obter mais informações, consulte ["Visão geral do Windows Defender"](#).

BitLocker

A criptografia de unidade BitLocker é um recurso de proteção de dados continuado do Windows Server 2012. Essa criptografia protege discos físicos, LUNs e CSVs.

Prática recomendada

Antes de ativar o BitLocker, o CSV deve ser colocado no modo de manutenção. Portanto, a NetApp recomenda que as decisões relativas à segurança baseada no BitLocker sejam tomadas antes de criar VMs no CSV para evitar o tempo de inatividade.

Implante o servidor Nano

Saiba mais sobre como implantar o Microsoft Windows Nano Server.

Implantação

Para implantar um Nano Server como um host Hyper-V, execute as seguintes etapas:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Copie a pasta NanoServerImageGenerator da pasta NanoServer na ISO do Windows Server para a unidade de disco rígido local.

3. Para criar um Nano Server VHD/VHDX, execute as seguintes etapas:

- a. Inicie o Windows PowerShell como administrador, navegue até a pasta NanoServerImageGenerator copiada no disco rígido local e execute o seguinte cmdlet:

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

- b. Crie um VHD para o Nano Server como um host Hyper-V executando o cmdlet do PowerShell a seguir. Este comando solicita uma senha de administrador para o novo 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
.. No exemplo a seguir, criamos um Nano Server VHD com o recurso
Hyper-V host com cluster de failover ativado. Este exemplo cria um
Nano Server VHD a partir de um ISO montado em f: '. O VHD recém-
criado é colocado em uma pasta chamada NanoServer na pasta de onde o
cmdlet é executado. O nome do computador é NanoServer e o VHD
resultante contém a edição padrão do Windows Server.
```

```
New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath f:\ -TargetPath .\NanoServer.vhd -ComputerName NanoServer
-Compute -Clustering
.. Com o cmdlet New-NanoServerImage, configure parâmetros que definem
o endereço IP, a máscara de sub-rede, o gateway padrão, o servidor
DNS, o nome do domínio, etc.
```

4. Use o VHD em uma VM ou host físico para implantar o Nano Server como um host Hyper-V:

- a. Para implantação em uma VM, crie uma nova VM no Hyper-V Manager e use o VHD criado na Etapa 3.
- b. Para implantação em um host físico, copie o VHD para o computador físico e configure-o para inicializar a partir deste novo VHD. Primeiro, monte o VHD, execute o bcdboot e: Janelas (onde o VHD está montado sob e:), desmonte o VHD, reinicie o computador físico e inicialize no Nano Server.

5. Junte o Nano Server a um domínio (opcional):

- a. Faça login em qualquer computador no domínio e crie um blob de dados executando o seguinte cmdlet do 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
.. Copie o arquivo odjblob para o Nano Server executando os seguintes
cmdlets do PowerShell em uma máquina remota:
```

```
$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
$nanosecuredcred = new-object management.automation.pscredential
$nanouname, $securenanopwd
```

```
Invoke-Command -VMName $nanoserver -Credential $nanosecuredcred
-ArgumentList @($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. Reinicie o Nano Server.

Conecte-se ao Nano Server

Para se conectar ao Nano Server remotamente usando o PowerShell, execute as seguintes etapas:

1. Adicione o Nano Server como um host confiável no computador remoto executando o seguinte cmdlet no servidor remoto:

```
Set-Item WSMAN:\LocalHost\Client\TrustedHosts "<input IP Address of the Nano Server>"
```

. Se o ambiente estiver seguro e se você quiser definir todos os hosts para serem adicionados como hosts confiáveis em um servidor, execute o seguinte comando:

```
Set-Item WSMAN:\LocalHost\Client\TrustedHosts *
```

. Inicie a sessão remota executando o cmdlet a seguir no servidor remoto. Forneça a senha para o Nano Server quando solicitado.

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

Para se conectar ao Nano Server remotamente usando ferramentas de gerenciamento de GUI de um Windows Server remoto, execute os seguintes comandos:

1. Inicie sessão no Windows Server como membro do grupo de administradores.
2. Inicie o Server Manager.
3. Para gerenciar um Nano Server remotamente a partir do Gerenciador de servidores, clique com o botão direito em todos os servidores, clique em Adicionar servidores, forneça as informações do Nano Server e adicione-o. Agora você pode ver o Nano Server listado na lista de servidores. Selecione o Nano Server, clique com o botão direito do Mouse nele e comece a gerenciá-lo com as várias opções fornecidas.
4. Para gerenciar serviços em um Nano Server remotamente, execute as seguintes etapas:
 - a. Abra Serviços na seção Ferramentas do Gerenciador de servidores.
 - b. Clique com o botão direito em Serviços (local).
 - c. Clique em conectar ao servidor.
 - d. Forneça os detalhes do Nano Server para visualizar e gerenciar os serviços no Nano Server.
5. Se a função Hyper-V estiver ativada no Nano Server, execute as seguintes etapas para gerenciá-la remotamente a partir do Gerenciador Hyper-V:
 - a. Abra o Gerenciador do Hyper-V na seção Ferramentas do Gerenciador de servidores.
 - b. Clique com o botão direito do rato em Gestor Hyper-V.
 - c. Clique em conectar ao servidor e forneça os detalhes do Nano Server. Agora, o Nano Server pode ser gerenciado como um servidor Hyper-V para criar e gerenciar VMs em cima dele.
6. Se a função de cluster de failover estiver ativada no Nano Server, execute as seguintes etapas para gerenciá-lo remotamente a partir do gerenciador de cluster de failover:
 - a. Abra o Gerenciador de Cluster de failover na seção Ferramentas do Gerenciador de servidores.
 - b. Execute operações relacionadas a clustering com o Nano Server.

Implantar o cluster Hyper-V.

Este apêndice descreve a implantação de um cluster Hyper-V.

Pré-requisitos

- Existem pelo menos dois servidores Hyper-V conectados entre si.
- Pelo menos um switch virtual é configurado em cada servidor Hyper-V.
- O recurso de cluster de failover é ativado em cada servidor Hyper-V.
- Compartilhamentos SMB ou CSVs são usados como armazenamento compartilhado para armazenar VMs e seus discos para clustering Hyper-V.
- O storage não deve ser compartilhado entre clusters diferentes. Você deve ter apenas um compartilhamento CSV/CIFS por cluster.
- Se o compartilhamento SMB for usado como armazenamento compartilhado, as permissões no compartilhamento SMB devem ser configuradas para conceder acesso às contas de computador de todos os servidores Hyper-V no cluster.

Implantação

1. Faça login em um dos servidores do Windows Hyper-V como membro do grupo de administradores.
2. Inicie o Server Manager.
3. Na seção Ferramentas, clique em Gerenciador de Cluster de failover.
4. Clique no menu criar cluster a partir de ações.
5. Forneça detalhes para o servidor Hyper-V que faz parte deste cluster.
6. Valide a configuração do cluster. Selecione Sim quando solicitado para validar a configuração do cluster e selecione os testes necessários para validar se os servidores Hyper-V passam os pré-requisitos para fazer parte do cluster.
7. Após a validação ser bem-sucedida, o assistente criar cluster é iniciado. No assistente, forneça o nome do cluster e o endereço IP do cluster para o novo cluster. Um novo cluster de failover é então criado para o servidor Hyper-V.
8. Clique no cluster recém-criado no Gerenciador de Cluster de failover e gerencie-o.
9. Defina o armazenamento partilhado para o cluster a utilizar. Pode ser um compartilhamento SMB ou um CSV.
10. Usar um compartilhamento SMB como armazenamento compartilhado não requer etapas especiais.
 - Configurar um compartilhamento CIFS em um controlador de storage NetApp. Para o fazer, consulte a seção ["Provisionamento em ambientes SMB"](#).
11. Para usar um CSV como armazenamento compartilhado, execute as seguintes etapas:
 - a. Configurar LUNs em um controlador de storage NetApp. Para fazer isso, consulte a seção "provisionamento em ambientes SAN".
 - b. Certifique-se de que todos os servidores Hyper-V no cluster de failover possam ver os LUNs NetApp. Para fazer isso para todos os servidores Hyper-V que fazem parte do cluster de failover, verifique se seus iniciadores são adicionados ao grupo de iniciadores no armazenamento NetApp. Certifique-se também de que seus LUNs sejam descobertos e verifique se o MPIO está habilitado.
 - c. Em qualquer um dos servidores Hyper-V no cluster, execute as seguintes etapas:
 - i. Leve o LUN online, inicialize o disco, crie um novo volume simples e formate-o usando NTFS ou

refs.

- ii. No Gerenciador de Cluster de failover, expanda o cluster, expanda armazenamento, clique com o botão direito do Mouse em discos e clique em Adicionar discos. Isso abre o assistente Adicionar discos a um cluster mostrando o LUN como um disco. Clique em OK para adicionar o LUN como um disco.
 - iii. Agora, o LUN é nomeado disco agrupado e é mostrado como armazenamento disponível em discos.
 - d. Clique com o botão direito do rato no LUN (Clustered Disk) e clique em Adicionar ao Cluster Shared volumes. Agora, o LUN é mostrado como um CSV.
 - e. O CSV é simultaneamente visível e acessível a partir de todos os servidores Hyper-V do cluster de failover em sua localização local C: /ClusterStorage.
12. Crie uma VM altamente disponível:
- a. No Gerenciador de Cluster de failover, selecione e expanda o cluster criado anteriormente.
 - b. Clique em funções e, em seguida, clique em máquinas virtuais em ações. Clique em Nova máquina virtual.
 - c. Selecione o nó no cluster onde a VM deve residir.
 - d. No assistente de criação de máquinas virtuais, forneça o armazenamento compartilhado (compartilhamento SMB ou CSV) como o caminho para armazenar a VM e seus discos.
 - e. Use o Hyper-V Manager para definir o armazenamento compartilhado (compartilhamento SMB ou CSV) como o caminho padrão para armazenar a VM e seus discos para um servidor Hyper-V.
13. Testar failover planejado. Mova as VMs para outro nó usando a migração ao vivo, a migração rápida ou a migração de storage (mover). Reveja "[Migração ao vivo em um ambiente em cluster](#)" para obter mais detalhes.
14. Teste failover não planejado. Pare o serviço de cluster no servidor que possui a VM.

Implante a migração ao vivo do Hyper-V em um ambiente em cluster

Este apêndice descreve a implantação da migração ao vivo em um ambiente em cluster.

Pré-requisitos

Para implantar a migração em tempo real, você precisa ter servidores Hyper-V configurados em um cluster de failover com armazenamento compartilhado. Reveja "[Implantar o Hyper-V Cluster](#)" para obter mais detalhes.

Implantação

Para usar a migração em tempo real em um ambiente em cluster, execute as seguintes etapas:

1. No Gerenciador de Cluster de failover, selecione e expanda o cluster. Se o cluster não estiver visível, clique em Gerenciador de Cluster de failover, clique em conectar ao cluster e forneça o nome do cluster.
2. Clique em funções, que lista todas as VMs disponíveis em um cluster.
3. Clique com o botão direito do rato na VM e clique em mover. Isso fornece três opções:
 - **Migração ao vivo.** Você pode selecionar um nó manualmente ou permitir que o cluster selecione o melhor nó. Na migração em tempo real, o cluster copia a memória usada pela VM do nó atual para outro nó. Portanto, quando a VM é migrada para outro nó, as informações de memória e estado necessárias pela VM já estão em vigor para a VM. Esse método de migração é quase instantâneo, mas apenas uma VM pode ser migrada ao vivo de cada vez.

- **Migração rápida.** Você pode selecionar um nó manualmente ou permitir que o cluster selecione o melhor nó. Na migração rápida, o cluster copia a memória usada por uma VM para um disco no armazenamento. Portanto, quando a VM é migrada para outro nó, as informações de memória e estado necessárias pela VM podem ser lidas rapidamente do disco pelo outro nó. Com a migração rápida, várias VMs podem ser migradas simultaneamente.
- * Migração de armazenamento de máquina virtual.* Este método usa o assistente mover armazenamento de máquina virtual. Com este assistente, você pode selecionar o disco da VM juntamente com outros arquivos a serem movidos para outro local, que pode ser um CSV ou um compartilhamento SMB.

Implante a migração do Hyper-V Live fora de um ambiente em cluster

Esta seção descreve a implantação da migração em tempo real do Hyper-V fora de um ambiente em cluster.

Pré-requisitos

- Servidores Hyper-V autônomos com storage independente ou armazenamento SMB compartilhado.
- A função Hyper-V instalada nos servidores de origem e de destino.
- Ambos os servidores Hyper-V pertencem ao mesmo domínio ou a domínios que confiam uns nos outros.

Implantação

Para executar a migração em tempo real em um ambiente não agrupado, configure os servidores Hyper-V de origem e destino para que eles possam enviar e receber operações de migração em tempo real. Em ambos os servidores Hyper-V, execute as seguintes etapas:

1. Abra o Gerenciador do Hyper-V na seção Ferramentas do Gerenciador de servidores.
2. Em ações, clique em Configurações do Hyper-V.
3. Clique em Live Migrations (migrações em tempo real) e selecione Enable Incoming (Ativar migrações em tempo real de entrada)
4. Escolha se deseja permitir tráfego de migração em tempo real em qualquer rede disponível ou apenas em redes específicas.
5. Opcionalmente, você pode configurar o protocolo de autenticação e as opções de desempenho na seção Avançado de Live Migrations.
6. Se CredSSP for usado como protocolo de autenticação, certifique-se de fazer login no servidor Hyper-V de origem do servidor Hyper-V de destino antes de mover a VM.
7. Se o Kerberos for usado como protocolo de autenticação, configure a delegação restrita. Isso requer acesso ao controlador de domínio do ativo Directory. Para configurar a delegação, execute as seguintes etapas:
 - a. Faça login no controlador de domínio do ativo Directory como administrador.
 - b. Inicie o Server Manager.
 - c. Na seção Ferramentas, clique em usuários e computadores do ativo Directory.
 - d. Expanda o domínio e clique em computadores.
 - e. Selecione o servidor Hyper-V de origem na lista, clique nele com o botão direito e clique em Propriedades.

- f. Na guia Delegação, selecione confiar neste computador somente para Delegação aos Serviços especificados.
 - g. Selecione usar somente Kerberos.
 - h. Clique em Adicionar, que abre o assistente Adicionar serviços.
 - i. Em Adicionar Serviços, clique em usuários e computadores, que abre Select Users or Computers**
 - j. Forneça o nome do servidor Hyper-V de destino e clique em OK.
 - Para mover o armazenamento de VM, selecione CIFS.
 - Para mover VMs, selecione o serviço Microsoft Virtual System Migration.
 - k. Na guia Delegação, clique em OK.
 - l. Na pasta computadores, selecione o servidor Hyper-V de destino na lista e repita o processo. Em Selecionar usuários ou computadores, forneça o nome do servidor Hyper-V de origem.
8. Mova a VM.
- a. Abra o Gerenciador do Hyper-V.
 - b. Clique com o botão direito em uma VM e clique em mover.
 - c. Escolha mover a Máquina Virtual.
 - d. Especifique o servidor Hyper-V de destino para a VM.
 - e. Escolha as opções mover. Para migração ao vivo compartilhada, escolha mover somente a Máquina Virtual. Para Shared Nothing Live Migration, escolha qualquer uma das outras duas opções com base nas suas preferências.
 - f. Forneça a localização da VM no servidor Hyper-V de destino com base nas suas preferências.
 - g. Revise o resumo e clique em OK para mover a VM.

Implante a migração ao vivo do Hyper-V Storage

Saiba como configurar a migração ao vivo de armazenamento Hyper-V

Pré-requisitos

- Você deve ter um servidor Hyper-V autônomo com armazenamento independente (DAS ou LUN) ou armazenamento SMB (local ou compartilhado entre outros servidores Hyper-V).
- O servidor Hyper-V deve ser configurado para migração em tempo real. Revise a seção sobre implantação no "[Migração ao vivo fora de um ambiente em cluster](#)".

Implantação

1. Abra o Gerenciador do Hyper-V.
2. Clique com o botão direito em uma VM e clique em mover.
3. Selecione mover o armazenamento da máquina virtual.
4. Selecione opções para mover o armazenamento de acordo com as suas preferências.
5. Forneça o novo local para os itens da VM.
6. Revise o resumo e clique em OK para mover o armazenamento da VM.

Implantar a réplica do Hyper-V fora de um ambiente em cluster

Este apêndice descreve a implantação da réplica do Hyper-V fora de um ambiente em cluster.

Pré-requisitos

- Você precisa de servidores Hyper-V autônomos localizados nas mesmas localizações geográficas ou separadas, servindo como servidores primários e de réplica.
- Se forem usados sites separados, o firewall em cada site deve ser configurado para permitir a comunicação entre os servidores principal e de réplica.
- O servidor de réplica deve ter espaço suficiente para armazenar as cargas de trabalho replicadas.

Implantação

1. Configure o servidor de réplica.

- a. Para que as regras de firewall de entrada permitam tráfego de replicação de entrada, execute o seguinte cmdlet do PowerShell:

```
Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP Listener (TCP-In) "  
.. Abra o Gerenciador do Hyper-V na seção Ferramentas do Gerenciador de servidores.  
.. Clique em Configurações do Hyper-V em ações.  
.. Clique em Configuração de replicação e selecione Ativar este computador como um servidor de réplica.  
.. Na seção Autenticação e portas, selecione o método e a porta de autenticação.  
.. Na seção autorização e armazenamento, especifique o local para armazenar as VMs e arquivos replicados.
```

2. Habilite a replicação de VM para VMs no servidor primário. A replicação da VM é ativada por VM e não para todo o servidor Hyper-V.

- a. No Hyper-V Manager, clique com o botão direito do rato numa VM e clique em Ativar replicação para abrir o assistente Ativar replicação.
- b. Forneça o nome do servidor de réplica em que a VM deve ser replicada.
- c. Forneça o tipo de autenticação e a porta do servidor de réplica que foi configurada para receber tráfego de replicação no servidor de réplica.
- d. Selecione os VHDs a serem replicados.
- e. Escolha a frequência (duração) em que as alterações são enviadas para o servidor de réplica.
- f. Configure os pontos de recuperação para especificar o número de pontos de recuperação a serem mantidos no servidor de réplica.
- g. Escolha método de replicação inicial para especificar o método para transferir a cópia inicial dos dados da VM para o servidor de réplica.
- h. Reveja o resumo e clique em concluir.

- i. Esse processo cria uma réplica de VM no servidor de réplica.

Replicação

1. Execute um failover de teste para garantir que a réplica VM funcione corretamente no servidor de réplica. O teste cria uma VM temporária no servidor de réplica.
 - a. Faça login no servidor de réplica.
 - b. No Gerenciador do Hyper-V, clique com o botão direito em uma réplica VM, clique em replicação e clique em failover de teste.
 - c. Escolha o ponto de recuperação a utilizar.
 - d. Este processo cria uma VM com o mesmo nome anexado com -Test.
 - e. Verifique a VM para se certificar de que tudo funciona bem.
 - f. Após o failover, a VM de teste de réplica é excluída se você selecionar Parar failover de teste para ela.
2. Execute um failover planejado para replicar as alterações mais recentes na VM principal para a VM réplica.
 - a. Inicie sessão no servidor principal.
 - b. Desligue a VM para falha.
 - c. No Gerenciador do Hyper-V, clique com o botão direito do Mouse na VM desativada, clique em replicação e clique em failover planejado.
 - d. Clique em failover para transferir as alterações mais recentes da VM para o servidor de réplica.
3. Executar um failover não planejado no caso de falha de VM principal.
 - a. Faça login no servidor de réplica.
 - b. No Hyper-V Manager, clique com o botão direito do Mouse em uma réplica VM, clique em replicação e clique em failover.
 - c. Escolha o ponto de recuperação a utilizar.
 - d. Clique em failover para fazer failover na VM.

Implantar a réplica do Hyper-V em um ambiente em cluster

Saiba como implantar e configurar a réplica do Hyper-V com o Cluster de failover do Windows Server.

Pré-requisitos

- Você precisa ter clusters do Hyper-V localizados no mesmo local ou em locais geográficos separados, servindo como clusters primários e de réplica. Reveja "[Implantar o Hyper-V Cluster](#)" para obter mais detalhes.
- Se forem usados sites separados, o firewall em cada local deve ser configurado para permitir a comunicação entre os clusters primário e de réplica.
- O cluster de réplica deve ter espaço suficiente para armazenar as cargas de trabalho replicadas.

Implantação

1. Ative regras de firewall em todos os nós de um cluster. Execute o cmdlet do PowerShell a seguir com o admin Privileges em todos os nós nos clusters primário e de réplica.

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

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

2. Configure o cluster de réplica.

- a. Configure o corretor de réplica Hyper-V com um nome NetBIOS e endereço IP para usar como ponto de conexão para o cluster que é usado como o cluster de réplica.
 - i. Abra o Gerenciador de Cluster de failover.
 - ii. Expanda o cluster, clique em funções e clique no painel Configurar função a partir de ações.
 - iii. Selecione Hyper-V Replica Broker na página Selecionar função.
 - iv. Forneça o nome NetBIOS e o endereço IP a serem usados como ponto de conexão para o cluster (ponto de acesso do cliente).
 - v. Este processo cria uma função de corretor de réplica Hyper-V. Verifique se ele está online com sucesso.
- b. Configure as definições de replicação.
 - i. Clique com o botão direito do Mouse no corretor de réplica criado nas etapas anteriores e clique em Configurações de replicação.
 - ii. Selecione Ativar este cluster como um servidor de réplica.
 - iii. Na seção Autenticação e portas, selecione o método e a porta de autenticação.
 - iv. Na seção autorização e armazenamento, selecione os servidores autorizados a replicar VMs para este cluster. Além disso, especifique o local padrão onde as VMs replicadas são armazenadas.

Replicação

A replicação é semelhante ao processo descrito na "[Réplica fora de um ambiente em cluster](#)" seção .

Onde encontrar informações adicionais

Recursos adicionais para Microsoft Windows e Hyper-V.

- Conceitos de ONTAP <https://docs.netapp.com/us-en/ontap/concepts/introducing-data-management-software-concept.html>
- Práticas recomendadas para SAN moderna <https://www.netapp.com/media/10680-tr4080.pdf>
- Disponibilidade e integridade de dados de array all-SAN NetApp com o NetApp ASA <https://www.netapp.com/pdf.html?item=/media/85671-tr-4968.pdf>
- Documentação SMB <https://docs.netapp.com/us-en/ontap/smb-admin/index.html>

- Introdução ao Nano Server <https://technet.microsoft.com/library/mt126167.aspx>
- Novidades no Hyper-V no Windows Server <https://technet.microsoft.com/windows-server-docs/compute/hyper-v/what-s-new-in-hyper-v-on-windows>

Microsoft SQL Server

Visão geral

O ONTAP oferece uma solução de desempenho e segurança de nível empresarial para seus bancos de dados Microsoft SQL Server, além de fornecer ferramentas de classe mundial para gerenciar seu ambiente.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-4590: Guia de práticas recomendadas para Microsoft SQL Server com ONTAP*

NetApp assume que o leitor tem conhecimento prático do seguinte:

- ONTAP
- Arquitetura e administração do Microsoft SQL Server
- Software de backup SnapCenter, incluindo os seguintes componentes:
- NetApp SnapCenter como software de backup, que inclui:
 - Plug-in do SnapCenter para Microsoft Windows
 - Plug-in do SnapCenter para SQL Server

O escopo desta seção de práticas recomendadas está limitado ao design técnico com base nos princípios e padrões preferenciais recomendados pela NetApp para a infraestrutura de storage. A implementação de ponta a ponta está fora do escopo.

Para obter informações sobre compatibilidade entre os produtos NetApp, consulte "[Ferramenta de Matriz de interoperabilidade NetApp \(IMT\)](#)".

Cargas de trabalho do Microsoft SQL Server

Antes de implantar o SQL Server, você deve entender os requisitos de carga de trabalho dos aplicativos suportados pelas instâncias de banco de dados do SQL Server. Cada aplicativo tem requisitos diferentes de capacidade, desempenho e disponibilidade, e, portanto, cada banco de dados deve ser projetado para oferecer suporte ideal a esses requisitos. Muitas organizações classificam os bancos de dados em várias camadas de gerenciamento, usando os requisitos da aplicação para definir SLAs. As cargas de trabalho do SQL Server são frequentemente categorizadas como descrito abaixo:

- OLTP, que geralmente são os bancos de dados mais críticos de uma organização. Esses bancos de dados geralmente apoiam aplicativos voltados para o cliente e são considerados essenciais para as operações principais da empresa. Os bancos de dados OLTP de missão crítica e as aplicações com suporte geralmente têm SLAs que exigem alta performance, são sensíveis à degradação da performance e exigem disponibilidade máxima. Eles também podem ser candidatos para clusters de failover sempre em ou grupos de disponibilidade sempre em disponibilidade. A mistura de e/S nesses tipos de bancos de dados é geralmente caracterizada por 75% a 90% de leitura aleatória e 25% a 10% de gravação.
- Bancos de dados do sistema de suporte à decisão (DSS), às vezes chamados de data warehouses. Esses bancos de dados são essenciais para muitas organizações que dependem de análises para seus negócios. Esses bancos de dados são sensíveis à utilização da CPU e operações de leitura do disco quando as consultas estão sendo executadas. Em muitas organizações, os bancos de dados DSS são os mais críticos durante o mês, trimestre e fim do ano. Em geral, esse workload tem uma combinação de I/O de leitura de quase 100% e a taxa de transferência de e/S geralmente é mais importante do que o IOPS.

Configuração do banco de dados

Configuração CPU

O desempenho do SQL Server tem várias dependências na CPU e na configuração do núcleo.

Hyperthreading

O hyperthreading refere-se à implementação simultânea de multithreading (SMT), o que melhora a paralelização dos cálculos realizados em processadores x86. SMT está disponível em ambos os processadores Intel e AMD.

O hyperthreading resulta em CPUs lógicas que aparecem como CPUs físicas para o sistema operacional. O SQL Server então vê essas CPUs adicionais e as usa como se houvesse mais núcleos do que fisicamente presentes. Isso pode melhorar substancialmente o desempenho aumentando a paralelização.

A ressalva aqui é que cada versão do SQL Server tem suas próprias limitações sobre o poder de computação que pode usar. Para obter mais informações, consulte limites de capacidade de computação por edição do SQL Server.

Núcleos e licenciamento

Existem duas opções para o licenciamento do SQL Server. O primeiro é conhecido como um modelo de licença de acesso de servidor e cliente (CAL); o segundo é o modelo de núcleo por processador. Embora você possa acessar todos os recursos do produto disponíveis no SQL Server com a estratégia de servidor e CAL, há um limite de hardware de 20 núcleos de CPU por soquete. Mesmo que você tenha o SQL Server Enterprise Edition para um servidor com mais de 20 núcleos de CPU por soquete, o aplicativo não pode usar todos esses núcleos de uma vez nessa instância.

A imagem abaixo mostra a mensagem de log do SQL Server após a inicialização indicando a imposição do limite do núcleo.

```

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:
-a "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

```

Portanto, para usar todas as CPUs, você deve usar a licença de núcleo por processador. Para obter informações detalhadas sobre o licenciamento do SQL Server, ["SQL Server 2022: Sua plataforma de dados moderna"](#) consulte .

Afinidade de CPU

É improvável que você precise alterar os padrões de afinidade do processador a menos que você encontre problemas de desempenho, mas ainda vale a pena entender o que eles são e como eles funcionam.

O SQL Server suporta afinidade de processador por duas opções:

- Máscara de afinidade da CPU
- Máscara de e/S de afinidade

O SQL Server usa todas as CPUs disponíveis no sistema operacional (se a licença de núcleo por processador for escolhida). Ele também cria agendadores f0r cada CPU para fazer o melhor uso dos recursos para qualquer carga de trabalho. Ao multitarefa, o sistema operacional ou outros aplicativos no servidor podem alternar processos de um processador para outro. O SQL Server é um aplicativo com uso intenso de recursos e o desempenho pode ser afetado quando isso ocorre. Para minimizar o impactos, você pode configurar os processadores para que toda a carga do SQL Server seja direcionada para um grupo pré-selecionado de processadores. Isso é conseguido usando a máscara de afinidade da CPU.

A opção de máscara de e/S de afinidade vincula e/S de disco do SQL Server a um subconjunto de CPUs. Em

ambientes OLTP do SQL Server, essa extensão pode melhorar significativamente o desempenho de threads do SQL Server que emitem operações de e/S.

Grau máximo de paralelismo (MAXDOP)

Por padrão, o SQL Server usa todas as CPUs disponíveis durante a execução da consulta se a licença de núcleo por processador escolhida.

Embora isso seja útil para consultas grandes, pode causar problemas de desempenho e limitar a simultaneidade. Uma melhor abordagem é limitar o paralelismo ao número de núcleos físicos em um único soquete de CPU. Por exemplo, em um servidor com dois soquetes físicos de CPU com 12 núcleos por soquete, independentemente do hyperthreading, `MAXDOP` deve ser definido como 12. `MAXDOP` Não é possível restringir ou ditar qual CPU deve ser usada. Em vez disso, ele restringe o número de CPUs que podem ser usadas por uma única consulta em lote.



A NetApp recomenda para DSS, como data warehouses, comece com `MAXDOP` a 50 e explore o ajuste para cima ou para baixo, se necessário. Certifique-se de medir as consultas críticas em seu aplicativo ao fazer alterações.

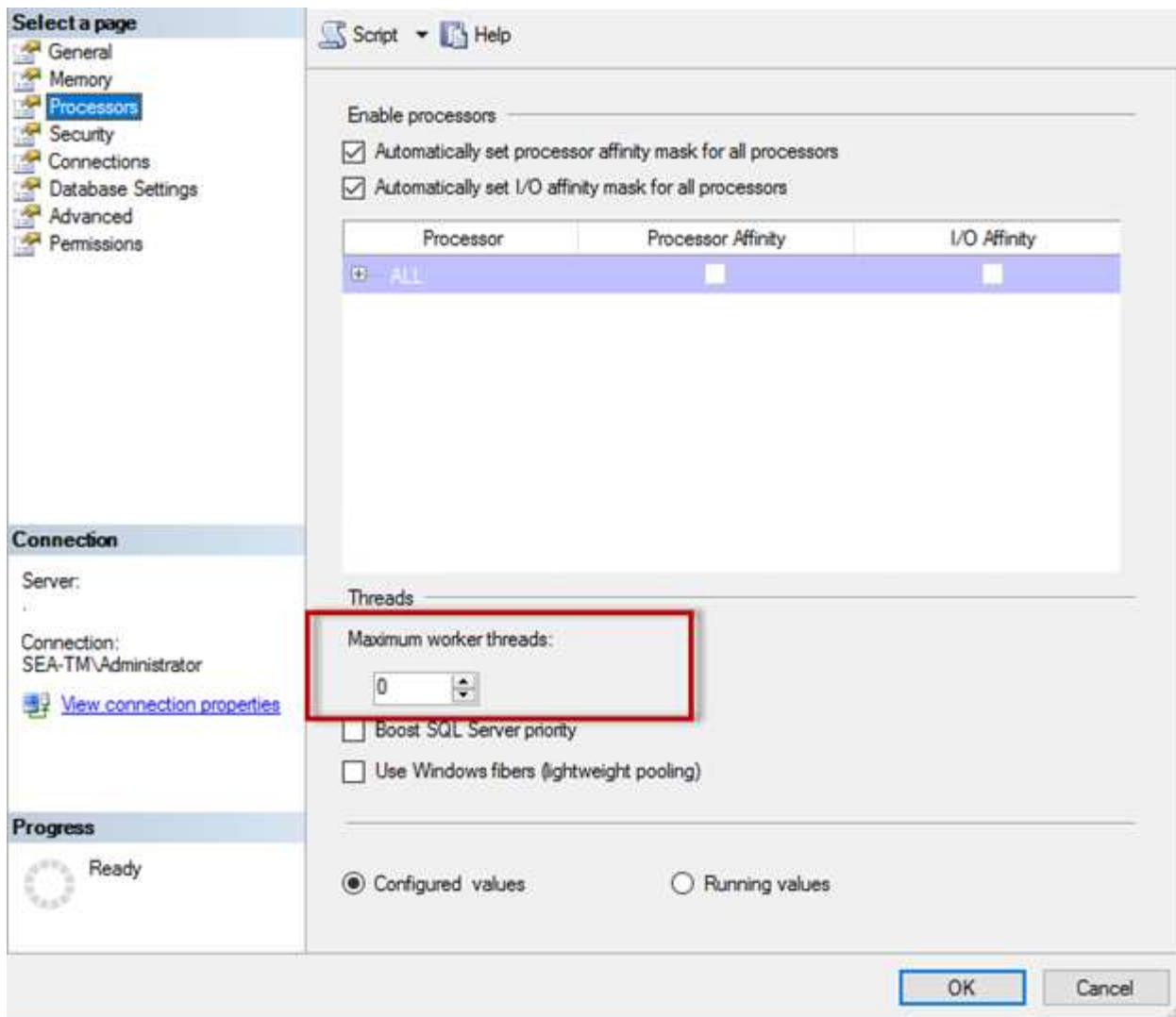
Max worker threads

A opção `Max worker threads` ajuda a otimizar o desempenho quando um grande número de clientes está conectado ao SQL Server.

Normalmente, um thread de sistema operacional separado é criado para cada consulta. Se centenas de conexões simultâneas forem feitas ao SQL Server, a configuração de um thread por consulta pode consumir recursos excessivos do sistema. A `max worker threads` opção ajuda a melhorar o desempenho ao permitir que o SQL Server crie um conjunto de threads de trabalho que podem atender coletivamente um número maior de solicitações de consulta.

O valor padrão é 0, o que permite que o SQL Server configure automaticamente o número de threads de trabalho na inicialização. Isso funciona para a maioria dos sistemas. `Max worker threads` é uma opção avançada e não deve ser alterada sem a ajuda de um administrador de banco de dados experiente (DBA).

Quando você deve configurar o SQL Server para usar mais threads de trabalho? Se o comprimento médio da fila de trabalho para cada agendador for superior a 1, você pode se beneficiar da adição de mais threads ao sistema, mas somente se a carga não estiver ligada à CPU ou tiver outras esperas pesadas. Se qualquer uma dessas situações estiver acontecendo, adicionar mais threads não ajuda porque acabam esperando outros gargalos do sistema. Para obter mais informações sobre threads de trabalho máximo, "[Configure a opção Max Worker threads Server Configuration](#)" consulte .



Configurando o máximo de threads de trabalho usando o SQL Server Management Studio.

O exemplo a seguir mostra como configurar a opção Max Work threads usando T-SQL.

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

Configuração da memória

A seção a seguir explica as configurações de memória do SQL Server necessárias para otimizar o desempenho do banco de dados.

Memória máx. Do servidor

A opção Max Server Memory (memória máxima do servidor) define a quantidade máxima de memória que a instância do SQL Server pode usar. Geralmente é usado se vários aplicativos estiverem sendo executados no mesmo servidor em que o SQL Server está sendo executado e você deseja garantir que esses aplicativos tenham memória suficiente para funcionar corretamente.

Algumas aplicações utilizam apenas qualquer memória disponível quando iniciam e não solicitam memória adicional, mesmo que estejam sob pressão de memória. É aí que entra em jogo a configuração de memória máxima do servidor.

Em um cluster do SQL Server com várias instâncias do SQL Server, cada instância pode estar competindo por recursos. Definir um limite de memória para cada instância do SQL Server pode ajudar a garantir o melhor desempenho para cada instância.



A NetApp recomenda deixando pelo menos 4GB GB a 6GB GB de RAM para o sistema operacional para evitar problemas de desempenho.

The screenshot shows the 'Server Memory' configuration window in SQL Server Enterprise Manager. The 'Memory' tab is active. The 'Server memory options' section is highlighted with a red box, showing the following settings:

- Minimum server memory (in MB): 0
- Maximum server memory (in MB): 120832

Below this, the 'Other memory options' section shows the following settings:

- Index creation memory (in KB, 0 = dynamic memory): 0
- Minimum memory per query (in KB): 1024

At the bottom, the 'Configured values' radio button is selected, and the 'Running values' radio button is unselected. The 'OK' and 'Cancel' buttons are visible at the bottom right.

Ajustando a memória mínima e máxima do servidor usando o SQL Server Management Studio.

Usar o SQL Server Management Studio para ajustar a memória mínima ou máxima do servidor requer uma reinicialização do serviço SQL Server. Você também pode ajustar a memória do servidor usando Transact

SQL (T-SQL) usando este código:

```
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
```

Acesso à memória não uniforme

O Acesso à memória não uniforme (NUMA) é uma tecnologia de otimização de acesso à memória que ajuda a evitar a carga extra no barramento do processador.

Se O NUMA estiver configurado num servidor onde o SQL Server está instalado, não é necessária nenhuma configuração adicional porque o SQL Server tem UMA consciência NUMA e tem um bom desempenho no hardware NUMA.

Index criar memória

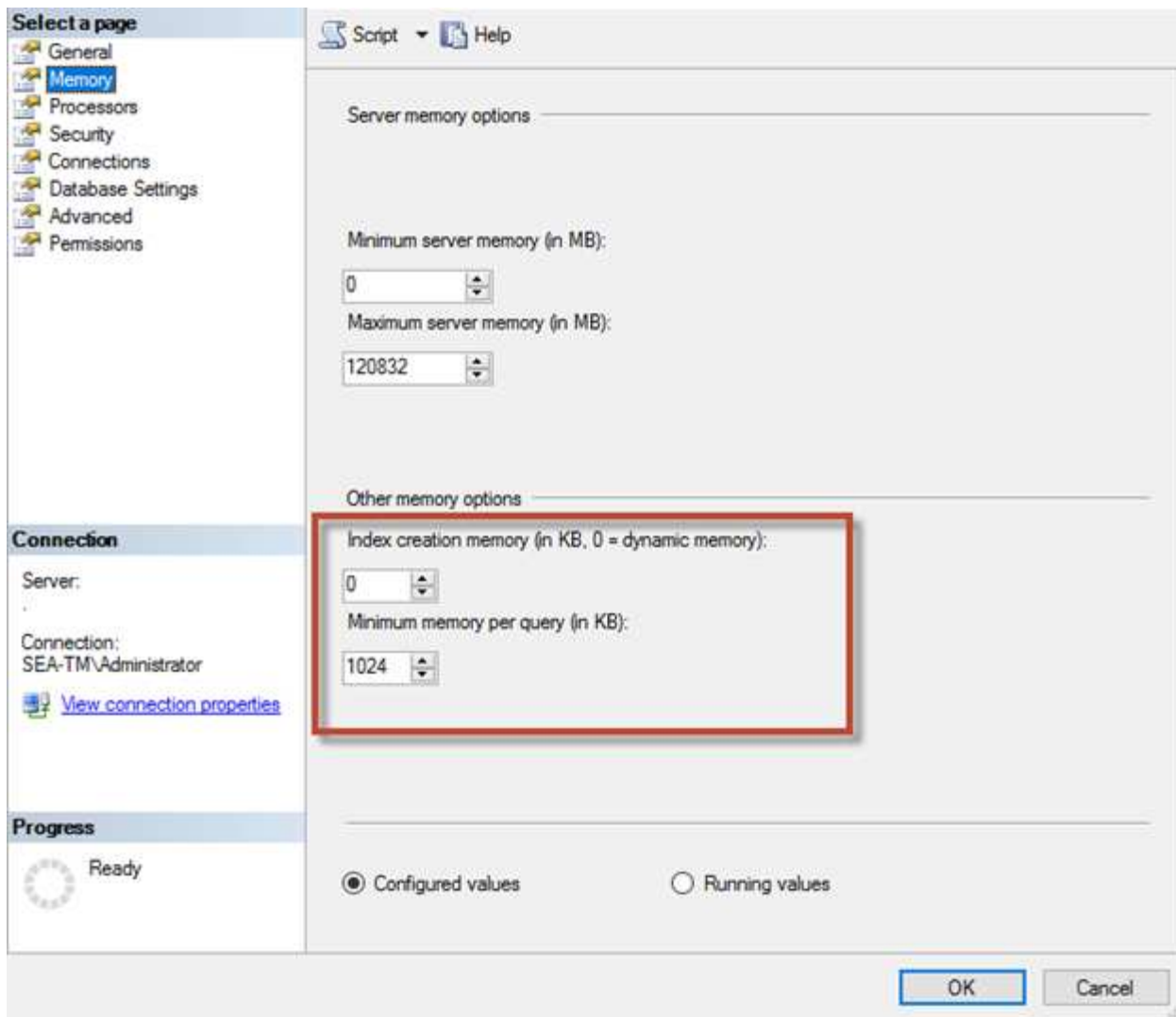
A opção index create memory (criar memória) é outra opção avançada que normalmente não deve ser alterada a partir dos padrões.

Ele controla a quantidade máxima de RAM inicialmente alocada para a criação de índices. O valor padrão para essa opção é 0, o que significa que ela é gerenciada automaticamente pelo SQL Server. No entanto, se você encontrar dificuldades para criar índices, considere aumentar o valor dessa opção.

Memória mínima por consulta

Quando uma consulta é executada, o SQL Server tenta alocar a quantidade ideal de memória para ser executado com eficiência.

Por padrão, a configuração de memória mínima por consulta aloca > 1024KB para cada consulta ser executada. É uma prática recomendada é deixar essa configuração no valor padrão para permitir que o SQL Server gerencie dinamicamente a quantidade de memória alocada para operações de criação de índice. No entanto, se o SQL Server tiver mais RAM do que precisa para ser executado de forma eficiente, o desempenho de algumas consultas pode ser aumentado se você aumentar essa configuração. Portanto, desde que a memória esteja disponível no servidor que não está sendo usado pelo SQL Server, por quaisquer outros aplicativos ou pelo sistema operacional, então aumentar essa configuração pode ajudar o desempenho geral do SQL Server. Se não houver memória livre disponível, aumentar essa configuração pode prejudicar o desempenho geral.



Instância compartilhada versus instância dedicada

O SQL Server pode ser configurado como uma única instância por servidor ou como várias instâncias. A decisão certa geralmente depende de fatores como se o servidor deve ser usado para produção ou desenvolvimento, se a instância é considerada crítica para as operações de negócios e metas de desempenho.

Configurações de instâncias compartilhadas podem ser inicialmente mais fáceis de configurar, mas podem levar a problemas nos quais os recursos ficam divididos ou bloqueados, o que, por sua vez, causa problemas de desempenho para outros aplicativos que têm bancos de dados hospedados na instância compartilhada do SQL Server.

A solução de problemas de desempenho pode ser complicada porque você precisa descobrir qual instância é a causa principal. Esta questão é ponderada em relação aos custos das licenças do sistema operacional e das licenças do SQL Server. Se o desempenho do aplicativo for primordial, uma instância dedicada é altamente recomendada.

A Microsoft licencia o SQL Server por núcleo no nível do servidor e não por instância. Por esse motivo, os administradores de banco de dados são tentados a instalar quantas instâncias do SQL Server o servidor pode lidar para economizar nos custos de licenciamento, o que pode levar a grandes problemas de desempenho posteriormente.



A NetApp recomenda escolher instâncias dedicadas do SQL Server sempre que possível para obter o desempenho ideal.

Configuração de armazenamento

Visão geral

A combinação de soluções de storage ONTAP e Microsoft SQL Server permite designs de storage de banco de dados de nível empresarial que podem atender aos requisitos de aplicativos mais exigentes dos dias de hoje.

A otimização de uma solução SQL Server no ONTAP requer a compreensão do padrão e/S do SQL Server. Um layout de armazenamento bem projetado para um banco de dados do SQL Server deve suportar os requisitos de desempenho do SQL Server, além de fornecer o máximo de gerenciamento da infraestrutura como um todo. Um bom layout de storage também permite que a implantação inicial seja bem-sucedida e o ambiente cresça suavemente ao longo do tempo à medida que a empresa cresce.

Design de storage de dados

Para bancos de dados do SQL Server que não usam o SnapCenter para executar backups, a Microsoft recomenda colocar os dados e arquivos de log em unidades separadas. Para aplicativos que simultaneamente atualizam e solicitam dados, o arquivo de log é intenso de gravação e o arquivo de dados (dependendo do aplicativo) é intenso de leitura/gravação. Para a recuperação de dados, o ficheiro de registo não é necessário. Portanto, as solicitações de dados podem ser satisfeitas a partir do arquivo de dados colocado em sua própria unidade.

Ao criar um novo banco de dados, a Microsoft recomenda especificar unidades separadas para os dados e logs. Para mover arquivos após a criação do banco de dados, o banco de dados deve ser offline. Para obter mais recomendações da Microsoft, "[Coloque dados e arquivos de log em unidades separadas](#)" consulte .

Agregados

Agregados são os contêineres de storage de nível mais baixo para configurações de storage NetApp. Alguma documentação legada existe na internet que recomenda separar o IO em diferentes conjuntos de unidades subjacentes. Isso não é recomendado com o ONTAP. A NetApp realizou vários testes de caracterização de carga de trabalho de e/S usando agregados compartilhados e dedicados com arquivos de dados e arquivos de log de transações separados. Os testes mostram que um agregado grande com mais grupos RAID e unidades otimiza e melhora o desempenho do armazenamento, além de ser mais fácil para os administradores gerenciarem por dois motivos:

- Um agregado grande torna os recursos de e/S de todas as unidades disponíveis para todos os arquivos.
- Um agregado grande permite o uso mais eficiente do espaço em disco.

Para alta disponibilidade (HA), coloque a réplica síncrona secundária do SQL Server Always On Availability Group em uma máquina virtual de storage (SVM) separada no agregado. Para fins de recuperação de desastre, coloque a réplica assíncrona em um agregado que faça parte de um cluster de storage separado no local de DR, com conteúdo replicado pelo uso da tecnologia NetApp SnapMirror. A NetApp recomenda ter pelo menos 10% de espaço livre disponível em um agregado para obter um desempenho ideal de storage.

Volumes

os volumes são criados e residem dentro de agregados. Este termo às vezes causa confusão porque um volume ONTAP não é um LUN. Um volume de ONTAP é um contêiner de gerenciamento para os dados. Um volume pode conter arquivos, LUNs ou até mesmo objetos S3D. Um volume não ocupa espaço, é usado apenas para o gerenciamento dos dados contidos.

Considerações sobre design de volume

Antes de criar um design de volume de banco de dados, é importante entender como o padrão e as características de e/S do SQL Server variam dependendo da carga de trabalho e dos requisitos de backup e recuperação. Consulte as seguintes recomendações da NetApp para volumes flexíveis:

- Evite compartilhar volumes entre hosts. Por exemplo, embora seja possível criar 2 LUNs em um único volume e compartilhar cada LUN com um host diferente, isso deve ser evitado porque pode complicar o gerenciamento.
- Use pontos de montagem NTFS em vez de letras de unidade para superar a limitação de 26 unidades no Windows. Ao usar pontos de montagem de volume, é uma recomendação geral dar ao rótulo de volume o mesmo nome que o ponto de montagem.
- Quando apropriado, configure uma política de dimensionamento automático de volume para ajudar a evitar condições fora do espaço. 17 Guia de melhores práticas para o Microsoft SQL Server com ONTAP 2022 NetApp, Inc. Todos os direitos reservados.
- Se você instalar o SQL Server em um compartilhamento SMB, verifique se o Unicode está habilitado nos volumes SMB para criar pastas.
- Defina o valor da reserva do snapshot no volume para zero para facilitar o monitoramento do ponto de vista operacional.
- Desative as programações de snapshot e as políticas de retenção. Em vez disso, use o SnapCenter para coordenar cópias Snapshot dos volumes de dados do SQL Server.
- Coloque os bancos de dados do sistema do SQL Server em um volume dedicado.
- Tempdb é um banco de dados de sistema usado pelo SQL Server como um espaço de trabalho temporário, especialmente para operações de e/S intensivas DBCC CHECKDB. Portanto, coloque esse banco de dados em um volume dedicado com um conjunto separado de fusos. Em ambientes grandes em que a contagem de volume é um desafio, você pode consolidar tempdb em menos volumes e armazená-lo no mesmo volume que outros bancos de dados do sistema após um Planejamento cuidadoso. A proteção de dados para tempdb não é uma prioridade alta porque este banco de dados é recriado sempre que o SQL Server é reiniciado.
- Coloque os arquivos de dados do usuário (.mdf) em volumes separados porque eles são cargas de trabalho de leitura/gravação aleatórias. É comum criar backups de log de transações com mais frequência do que backups de banco de dados. Por esse motivo, coloque arquivos de log de transações (.ldf) em um volume separado ou VMDK dos arquivos de dados para que programações de backup independentes possam ser criadas para cada um. Essa separação também isola a e/S de gravação sequencial dos arquivos de log da e/S de leitura/gravação aleatória de arquivos de dados e melhora significativamente o desempenho do SQL Server.

LUNs

- Certifique-se de que os arquivos do banco de dados do usuário e o diretório de log para armazenar o backup de log estejam em volumes separados para impedir que a política de retenção substitua os snapshots quando eles forem usados com a tecnologia SnapVault.
- Não misture arquivos de banco de dados e não de banco de dados, como arquivos relacionados à

pesquisa de texto completo, no mesmo LUN.

- Colocar arquivos secundários de banco de dados (como parte de um grupo de arquivos) em volumes separados melhora o desempenho do banco de dados SQL Server. Essa separação só é válida se o arquivo do banco de dados `.mdf`` não compartilhar seu LUN com `.mdf` outros arquivos.
- Se você criar LUNs com o DiskManager ou outras ferramentas, certifique-se de que o tamanho da unidade de alocação está definido como 64K para partições ao formatar os LUNs.
- Consulte "[Microsoft Windows e MPIO nativo sob as práticas recomendadas do ONTAP para SAN moderna](#)" para aplicar suporte multipathing no Windows a dispositivos iSCSI nas propriedades MPIO.

Arquivos de banco de dados e grupos de arquivos

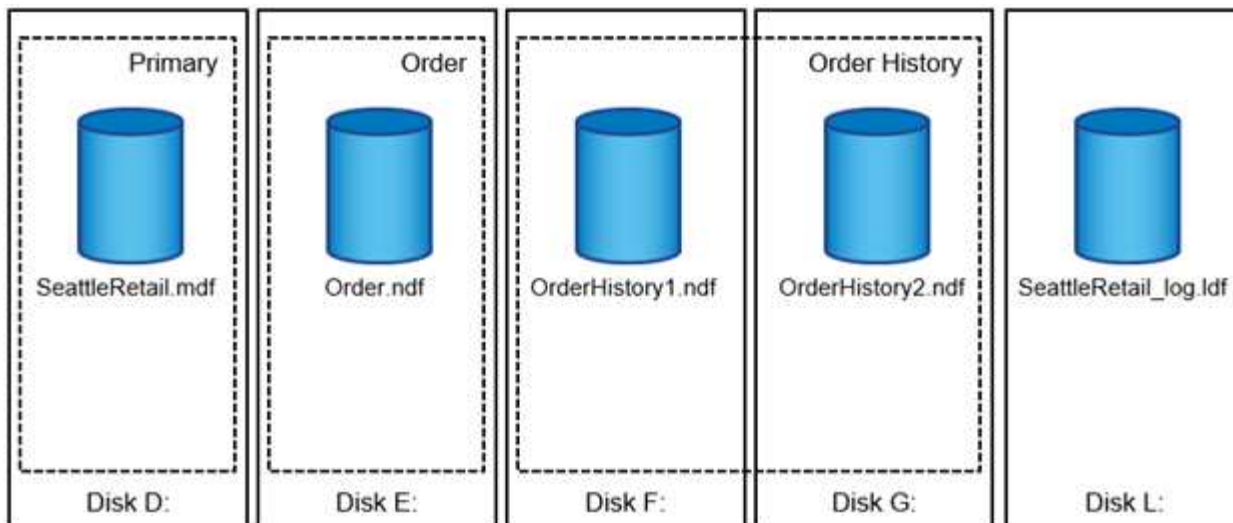
O posicionamento adequado do arquivo de banco de dados do SQL Server no ONTAP é fundamental durante a fase inicial de implantação. Isso garante um desempenho ideal, gerenciamento de espaço, backup e tempos de restauração que podem ser configurados para atender aos requisitos da sua empresa.

Em teoria, o SQL Server (64 bits) suporta 32.767 bancos de dados por instância e 524.272TB GB de tamanho de banco de dados, embora a instalação típica geralmente tenha vários bancos de dados. No entanto, o número de bancos de dados que o SQL Server pode manipular depende da carga e do hardware. Não é incomum ver instâncias do SQL Server hospedando dezenas, centenas ou até milhares de bancos de dados pequenos.

Cada banco de dados consiste em um ou mais arquivos de dados e um ou mais arquivos de log de transações. O log de transações armazena as informações sobre transações de banco de dados e todas as modificações de dados feitas por cada sessão. Sempre que os dados são modificados, o SQL Server armazena informações suficientes no log de transações para desfazer (reverter) ou refazer (reproduzir) a ação. Um log de transações do SQL Server é parte integrante da reputação do SQL Server em relação à integridade e robustez dos dados. O log de transações é vital para os recursos de atomicidade, consistência, isolamento e durabilidade (ACID) do SQL Server. O SQL Server grava no log de transações assim que qualquer alteração na página de dados ocorrer. Cada declaração DML (Data Manipulation Language) (por exemplo, selecionar, inserir, atualizar ou excluir) é uma transação completa, e o log de transações garante que toda a operação baseada em conjunto ocorra, certificando-se da atomicidade da transação.

Cada banco de dados tem um arquivo de dados primário, que, por padrão, tem a extensão `.mdf`. Além disso, cada banco de dados pode ter arquivos de banco de dados secundários. Esses arquivos, por padrão, têm extensões `.ndf`.

Todos os arquivos de banco de dados são agrupados em grupos de arquivos. Um grupo de arquivos é a unidade lógica, o que simplifica a administração do banco de dados. Eles permitem a separação entre o posicionamento lógico de objetos e arquivos de banco de dados físicos. Quando você cria as tabelas de objetos de banco de dados, você especifica em que grupo de arquivos eles devem ser colocados sem se preocupar com a configuração do arquivo de dados subjacente.



A capacidade de colocar vários arquivos de dados dentro do grupo de arquivos permite que você espalhe a carga entre diferentes dispositivos de armazenamento, o que ajuda a melhorar o desempenho de e/S do sistema. O log de transação em contraste não se beneficia dos vários arquivos porque o SQL Server grava no log de transação de forma sequencial.

A separação entre o posicionamento lógico de objetos nos grupos de arquivos e arquivos físicos de banco de dados permite ajustar o layout do arquivo do banco de dados, obtendo o máximo do subsistema de armazenamento. O número de arquivos de dados que suportam uma determinada carga de trabalho pode ser variado conforme necessário para dar suporte aos requisitos de e/S e à capacidade esperada, sem afetar a aplicação. Essas variações no layout do banco de dados são transparentes para os desenvolvedores de aplicativos, que estão colocando os objetos do banco de dados nos grupos de arquivos em vez de arquivos do banco de dados.



NetApp recomenda evitar o uso do grupo de arquivos primário para qualquer coisa além de objetos do sistema. Criar um grupo de arquivos separado ou um conjunto de grupos de arquivos para os objetos do usuário simplifica a administração do banco de dados e a recuperação de desastres, especialmente no caso de bancos de dados grandes.

Você pode especificar o tamanho inicial do arquivo e os parâmetros de crescimento automático no momento em que você cria o banco de dados ou adiciona novos arquivos a um banco de dados existente. O SQL Server usa um algoritmo de preenchimento proporcional ao escolher em qual arquivo de dados ele deve gravar dados. Ele grava uma quantidade de dados proporcionalmente ao espaço livre disponível nos arquivos. Quanto mais espaço livre no arquivo, mais escreve ele lida.



A NetApp recomenda que todos os arquivos no único grupo de arquivos tenham o mesmo tamanho inicial e parâmetros de crescimento automático, com o tamanho de crescimento definido em megabytes em vez de porcentagens. Isso ajuda o algoritmo de preenchimento proporcional equilibrar uniformemente as atividades de gravação em todos os arquivos de dados.

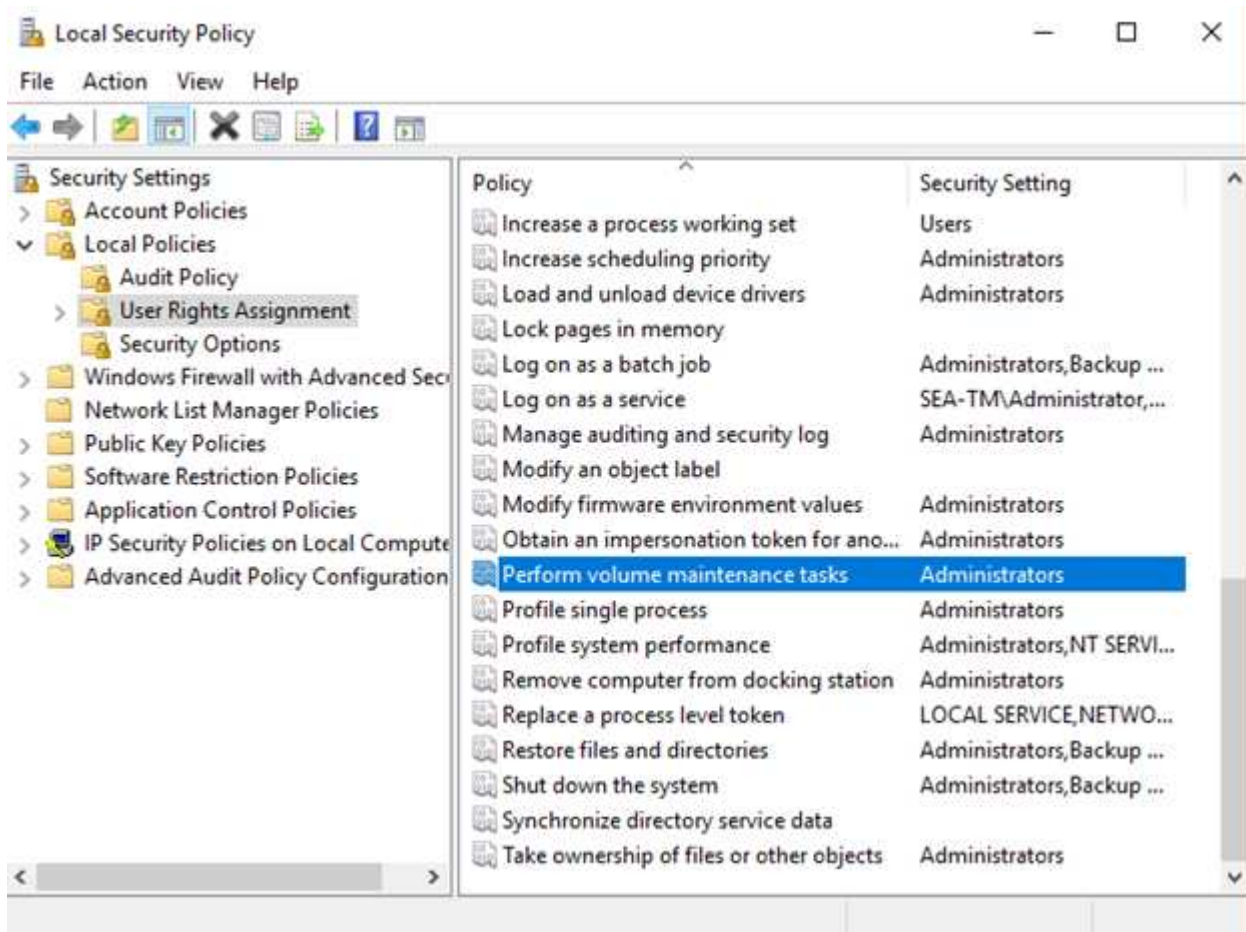
Toda vez que o SQL Server cresce arquivos, ele preenche o espaço recém-alocado com zeros. Esse processo bloqueia todas as sessões que precisam gravar no arquivo correspondente ou, em caso de crescimento de log de transações, gerar Registros de log de transações.

O SQL Server sempre apaga o log de transações e esse comportamento não pode ser alterado. No entanto, você pode controlar se os arquivos de dados estão zerando para fora habilitando ou desativando a inicialização instantânea de arquivos. Ativar a inicialização instantânea de arquivos ajuda a acelerar o

crescimento de arquivos de dados e reduz o tempo necessário para criar ou restaurar o banco de dados.

Um pequeno risco de segurança está associado à inicialização instantânea de arquivos. Quando esta opção está ativada, partes não alocadas do arquivo de dados podem conter informações de arquivos do sistema operacional excluídos anteriormente. Os administradores de banco de dados podem examinar esses dados.

Você pode habilitar a inicialização instantânea de arquivos adicionando a permissão SA_MANAGE_VOLUME_NAME, também conhecida como "executar tarefa de manutenção de volume", à conta de inicialização do SQL Server. Você pode fazer isso sob o aplicativo de gerenciamento de políticas de segurança local (secpol.msc), como mostrado na figura a seguir. Abra as propriedades da permissão "Executar tarefa de manutenção de volume" e adicione a conta de inicialização do SQL Server à lista de usuários lá.



Para verificar se a permissão está ativada, você pode usar o código do exemplo a seguir. Esse código define dois sinalizadores de rastreamento que forçam o SQL Server a gravar informações adicionais no log de erros, criar um pequeno banco de dados e ler o conteúdo do log.

```

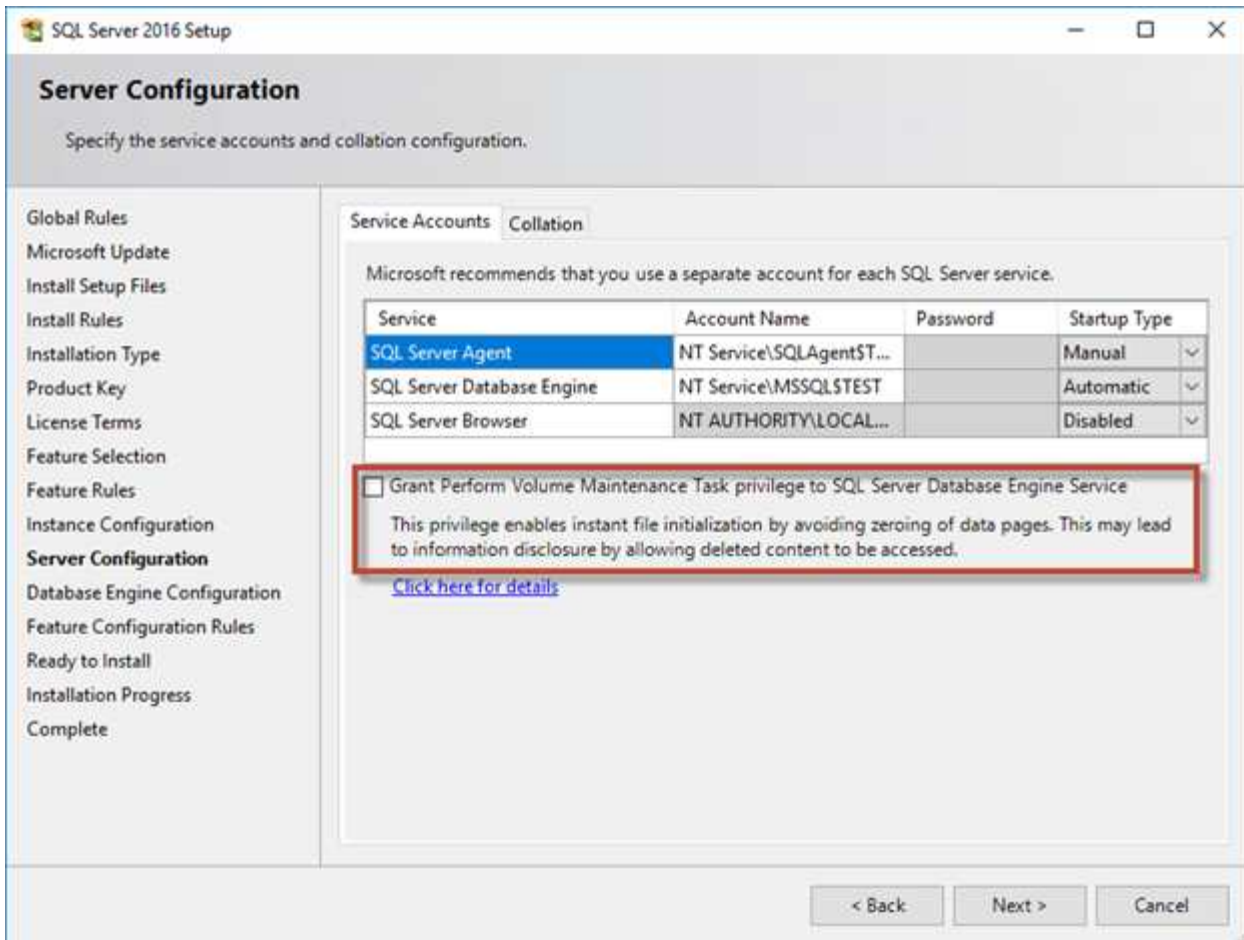
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO

```

Quando a inicialização instantânea do arquivo não está ativada, o log de erro do SQL Server mostra que o SQL Server está zerando o arquivo de dados do mdf, além de zerar o arquivo de log ldf, como mostrado no exemplo a seguir. Quando a inicialização instantânea do arquivo está ativada, ele exibe apenas a restauração do arquivo de log.

	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\Microsoft SQL Server\MSSQL
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQL
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.

A tarefa realizar manutenção de volume é simplificada no SQL Server 2016 e é fornecida posteriormente como uma opção durante o processo de instalação. Esta figura exibe a opção de conceder ao serviço do mecanismo de banco de dados SQL Server o privilégio de executar a tarefa de manutenção de volume.



Outra opção importante de banco de dados que controla os tamanhos de arquivo de banco de dados é o auto-retrátil. Quando essa opção está ativada, o SQL Server diminui regularmente os arquivos do banco de dados, reduz seu tamanho e libera espaço para o sistema operacional. Esta operação é intensiva em recursos e raramente é útil porque os arquivos de banco de dados crescem novamente após algum tempo quando novos dados entram no sistema. O Autoshink não deve ser ativado no banco de dados.

Diretório de log

O diretório de log é especificado no SQL Server para armazenar dados de backup de log de transação no nível do host. Se você estiver usando o SnapCenter para fazer backup de arquivos de log, cada host do SQL Server usado pelo SnapCenter deve ter um diretório de log do host configurado para executar backups de log. O SnapCenter tem um repositório de banco de dados, portanto, os metadados relacionados a operações de backup, restauração ou clonagem são armazenados em um repositório de banco de dados central.

Os tamanhos do diretório de log do host são calculados da seguinte forma: Tamanho do diretório de log do host ((tamanho máximo de DB LDF x taxa de alteração diária de log %) x (retenção de snapshot) ÷ (1 - LUN overhead space %) a fórmula de dimensionamento do diretório de log do host assume um espaço de sobrecarga de LUN de 10%

Coloque o diretório de log em um volume dedicado ou LUN. A quantidade de dados no diretório de log do host depende do tamanho dos backups e do número de dias em que os backups são mantidos. O SnapCenter permite apenas um diretório de log de host por host do SQL Server. Você pode configurar os diretórios de log

do host em SnapCenter --> Host --> Configurar Plug-in.

A NetApp recomenda o seguinte para um diretório de log do host:

- Certifique-se de que o diretório de log do host não é compartilhado por nenhum outro tipo de dados que possa potencialmente corromper os dados instantâneos do backup.
- Não coloque bancos de dados do usuário ou bancos de dados do sistema em um LUN que hospeda pontos de montagem.
- Crie o diretório de log do host em um volume dedicado ao qual o SnapCenter copia logs de transações.
- Use os assistentes do SnapCenter para migrar bancos de dados para o storage do NetApp, de modo que os bancos de dados sejam armazenados em locais válidos, permitindo operações de backup e restauração bem-sucedidas do SnapCenter. Tenha em mente que o processo de migração é disruptivo e pode fazer com que os bancos de dados fiquem offline enquanto a migração está em andamento.
- As seguintes condições devem estar em vigor para instâncias de cluster de failover (FCIs) do SQL Server:
 - Se você estiver usando uma instância de cluster de failover, o LUN do diretório de log do host deve ser um recurso de disco de cluster no mesmo grupo de cluster que a instância do SQL Server que está sendo feita backup do SnapCenter.
 - Se você estiver usando uma instância de cluster de failover, os bancos de dados de usuário devem ser colocados em LUNs compartilhados que são recursos de cluster de disco físico atribuídos ao grupo de cluster associado à instância do SQL Server.



Arquivos tempdb

O banco de dados tempdb pode ser muito utilizado. Além do posicionamento ideal de arquivos de banco de dados de usuários no ONTAP, o posicionamento de datafiles tempdb também é fundamental para reduzir a contenção de alocação

A contenção de páginas pode ocorrer nas páginas mapa de alocação global (GAM), mapa de alocação global compartilhado (SGAM) ou espaço livre de páginas (PFS) quando o SQL Server deve gravar em páginas especiais do sistema para alocar novos objetos. Os trincos bloqueiam estas páginas na memória. Em uma instância ocupada do SQL Server, pode levar muito tempo para obter uma trava em uma página do sistema no tempdb. Isso resulta em tempos de execução de consulta mais lentos e é conhecido como contenção de trava. Veja as práticas recomendadas a seguir para criar arquivos de dados tempdb:

- Para 8 núcleos: Arquivos de dados tempdb: Número de núcleos
- Para > 8 núcleos: 8 arquivos de dados tempdb

O script de exemplo a seguir modifica tempdb criando oito arquivos tempdb e movendo tempdb para o ponto de montagem C:\MSSQL\tempdb do SQL Server 2012 e posterior.

```
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%);
```

A partir do SQL Server 2016, o número de núcleos de CPU visíveis para o sistema operacional é detetado automaticamente durante a instalação e, com base nesse número, o SQL Server calcula e configura o número de arquivos tempdb necessários para um desempenho ideal.

Eficiência de storage

A eficiência de storage do ONTAP é otimizada para armazenar e gerenciar dados do SQL Server de uma forma que consome a menor quantidade de espaço de storage sem afetar a performance.

Os recursos de eficiência de espaço, como compressão, compactação e deduplicação, foram projetados para aumentar a quantidade de dados lógicos compatíveis com uma determinada quantidade de storage físico. O resultado são custos mais baixos e sobrecarga de gerenciamento.

Em um alto nível, a compressão é um processo matemático pelo qual padrões nos dados são detetados e codificados de uma forma que reduz os requisitos de espaço. Em contraste, a deduplicação deteta blocos repetidos reais de dados e remove as cópias externas. A compactação permite que vários blocos lógicos de dados compartilhem o mesmo bloco físico na Mídia.



Veja as seções abaixo sobre provisionamento de thin para uma explicação da interação entre eficiência de armazenamento e reserva fracionária.

Compactação

Antes da disponibilidade de sistemas de storage all-flash, a compactação baseada em array era de valor limitado porque a maioria dos workloads com uso intenso de e/S exigia um número muito grande de fusos para fornecer desempenho aceitável. Os sistemas de armazenamento invariavelmente continham muito mais capacidade do que o necessário como um efeito colateral do grande número de unidades. A situação mudou com o aumento do armazenamento de estado sólido. Não há mais a necessidade de superprovisionamento vastamente de unidades puramente para obter um bom desempenho. O espaço da unidade em um sistema de armazenamento pode ser correspondente às necessidades reais de capacidade.

A maior funcionalidade de IOPS das solid-State drives (SSDs) quase sempre produz economias de custo em comparação com as unidades giratórias, mas a compactação pode gerar mais economias ao aumentar a capacidade efetiva da Mídia de estado sólido.

Existem várias maneiras de compactar dados. Muitos bancos de dados incluem suas próprias capacidades de compressão, mas isso raramente é observado em ambientes de clientes. A razão é geralmente a penalidade de desempenho para uma **mudança** para dados compactados, além disso, com alguns aplicativos, há altos custos de licenciamento para compactação no nível do banco de dados. Finalmente, há as consequências gerais de desempenho para as operações do banco de dados. Faz pouco sentido pagar um alto custo de licença por CPU para uma CPU que executa compactação de dados e descompressão em vez de trabalho real de banco de dados. Uma opção melhor é descarregar o trabalho de compressão para o sistema de storage.

Compressão adaptável

A compactação adaptável foi totalmente testada com workloads empresariais sem efeito observado no desempenho, mesmo em um ambiente all-flash em que a latência é medida em microssegundos. Alguns

clientes até relataram um aumento de desempenho com o uso de compactação porque os dados permanecem compactados no cache, aumentando efetivamente a quantidade de cache disponível em um controlador.

O ONTAP gerencia blocos físicos em 4KB unidades. A compactação adaptável usa um tamanho de bloco de compressão padrão de 8KB, o que significa que os dados são compactados em unidades 8KB. Isso corresponde ao tamanho do bloco 8KB mais frequentemente usado por bancos de dados relacionais. Os algoritmos de compressão tornam-se mais eficientes à medida que mais dados são compactados como uma única unidade. Um tamanho de bloco de compressão 32KB seria mais eficiente em termos de espaço do que uma unidade de bloco de compressão 8KB. Isso significa que a compactação adaptável usando o tamanho padrão de bloco 8KB leva a taxas de eficiência ligeiramente menores, mas também há um benefício significativo para o uso de um tamanho menor de bloco de compressão. As cargas de trabalho de banco de dados incluem uma grande quantidade de atividade de substituição. Substituir um 8KB de um bloco de dados comprimido 32KB requer a leitura de todo o 32KB de dados lógicos, descomprimindo-o, atualizando a região 8KB necessária, recomprimindo e gravando o 32KB inteiro de volta às unidades. Esta é uma operação muito cara para um sistema de storage e é o motivo pelo qual alguns storage arrays concorrentes baseados em tamanhos de bloco de compressão maiores também incorrer em uma penalidade de desempenho significativa nos workloads de banco de dados.



O tamanho do bloco usado pela compressão adaptável pode ser aumentado até 32KBMB. Isso pode melhorar a eficiência do storage e deve ser considerado para arquivos inativos, como logs de transações e arquivos de backup, quando uma quantidade substancial de tais dados é armazenada no array. Em algumas situações, os bancos de dados ativos que usam um tamanho de bloco 16KB ou 32KB também podem se beneficiar do aumento do tamanho de bloco da compactação adaptável a corresponder. Consulte um representante da NetApp ou do parceiro para obter orientação sobre se isso é apropriado para sua carga de trabalho.



Tamanhos de blocos de compactação maiores que 8KBMB não devem ser usados juntamente com a deduplicação em destinos de backup de streaming. A razão é que pequenas alterações nos dados de backup afetam a janela de compressão 32KB. Se a janela mudar, os dados compactados resultantes diferem em todo o arquivo. A deduplicação ocorre após a compactação, o que significa que o mecanismo de deduplicação vê cada backup compactado de forma diferente. Se a deduplicação de backups de streaming for necessária, somente a compactação adaptável de bloco 8KB deve ser usada. A compactação adaptável é preferível, porque funciona em um tamanho de bloco menor e não prejudica a eficiência da deduplicação. Por razões semelhantes, a compactação no lado do host também interfere na eficiência da deduplicação.

Alinhamento da compressão

A compactação adaptável em um ambiente de banco de dados requer alguma consideração do alinhamento do bloco de compressão. Fazer isso é apenas uma preocupação para os dados que estão sujeitos a substituições aleatórias de blocos muito específicos. Essa abordagem é semelhante em conceito ao alinhamento geral do sistema de arquivos, onde o início de um sistema de arquivos deve ser alinhado a um limite de dispositivo 4K e o tamanho de bloco de um sistema de arquivos deve ser um múltiplo de 4K.

Por exemplo, uma gravação 8KBD em um arquivo é compactada somente se ele se alinhar com um limite 8KBD dentro do próprio sistema de arquivos. Este ponto significa que ele deve cair no primeiro 8KB do arquivo, o segundo 8KB do arquivo, e assim por diante. A maneira mais simples de garantir o alinhamento correto é usar o tipo de LUN correto, qualquer partição criada deve ter um deslocamento desde o início do dispositivo que é um múltiplo de 8K, e usar um tamanho de bloco de sistema de arquivos que é um múltiplo do tamanho do bloco de banco de dados.

Dados como backups ou logs de transações são operações sequencialmente escritas que abrangem vários

blocos, todos eles compactados. Portanto, não há necessidade de considerar o alinhamento. O único padrão de e/S de preocupação é as substituições aleatórias de arquivos.

Compactação de dados

A compactação de dados é uma tecnologia que melhora a eficiência da compressão. Como dito anteriormente, a compactação adaptável por si só pode proporcionar, na melhor das hipóteses, economias de 2:1x porque está limitada a armazenar uma e/S de 8KBx em um bloco de 4KB WAFL. Os métodos de compressão com tamanhos de bloco maiores proporcionam melhor eficiência. No entanto, eles não são adequados para dados sujeitos a pequenas substituições de blocos. A descompressão de 32KB unidades de dados, a atualização de uma porção 8KB, a recompressão e a gravação de volta nas unidades cria sobrecarga.

A compactação de dados funciona permitindo que vários blocos lógicos sejam armazenados em blocos físicos. Por exemplo, um banco de dados com dados altamente compressíveis, como texto ou blocos parcialmente completos, pode compactar de 8KB a 1KB. Sem compactação, 1KB PB de dados ainda ocupariam um bloco inteiro de 4KB TB. A compactação de dados in-line permite que 1KB TB de dados compactados sejam armazenados em apenas 1KB TB de espaço físico, juntamente com outros dados compactados. Não é uma tecnologia de compressão; é simplesmente uma maneira mais eficiente de alocar espaço nas unidades e, portanto, não deve criar qualquer efeito de desempenho detetável.

O grau de poupança obtido varia. Os dados que já estão compactados ou criptografados geralmente não podem ser mais compactados e, portanto, esses conjuntos de dados não se beneficiam com a compactação. Em contraste, datafiles recém-inicializados que contêm pouco mais do que metadados de bloco e zeros compactam até 80:1.

Eficiência de armazenamento sensível à temperatura

A eficiência de armazenamento sensível à temperatura (TSSE) está disponível no ONTAP 9.8 e posterior. Ele depende de mapas de calor de acesso a blocos para identificar blocos acessados com pouca frequência e compactá-los com maior eficiência.

Deduplicação

A deduplicação é a remoção de tamanhos de bloco duplicados de um conjunto de dados. Por exemplo, se o mesmo bloco 4KB existisse em 10 arquivos diferentes, a deduplicação redirecionaria esse bloco 4KB dentro de todos os arquivos 10 para o mesmo bloco físico 4KB. O resultado seria uma melhoria de 10:1 na eficiência para esses dados.

Dados como LUNs de inicialização convidado da VMware geralmente deduplicam muito bem porque consistem em várias cópias dos mesmos arquivos do sistema operacional. A eficiência de 100:1 e maior foi observada.

Alguns dados não contêm dados duplicados. Por exemplo, um bloco Oracle contém um cabeçalho que é globalmente exclusivo para o banco de dados e um trailer que é quase único. Como resultado, a deduplicação de um banco de dados Oracle raramente oferece mais de 1% de economia. A deduplicação com bancos de dados MS SQL é um pouco melhor, mas metadados exclusivos no nível de bloco ainda são uma limitação.

Economia de espaço de até 15% em bancos de dados com 16KB e grandes blocos foram observadas em alguns casos. O 4KB inicial de cada bloco contém o cabeçalho globalmente exclusivo, e o último bloco de 4KB contém o trailer quase único. Os blocos internos são candidatos à deduplicação, embora na prática isso seja quase inteiramente atribuído à deduplicação de dados zerados.

Muitos arrays concorrentes afirmam a capacidade de deduplicar bancos de dados com base na presunção de que um banco de dados é copiado várias vezes. A esse respeito, a deduplicação NetApp também pode ser

usada, mas o ONTAP oferece uma opção melhor: A tecnologia NetApp FlexClone. O resultado final é o mesmo; várias cópias de um banco de dados que compartilham a maioria dos blocos físicos subjacentes são criadas. Usar o FlexClone é muito mais eficiente do que ter tempo para copiar arquivos de banco de dados e, em seguida, deduplicá-los. É, na verdade, não duplicação em vez de deduplicação, porque uma duplicata nunca é criada em primeiro lugar.

Eficiência e thin Provisioning

Os recursos de eficiência são formas de thin Provisioning. Por exemplo, um LUN de 100GB GB ocupando um volume de 100GB TB pode ser compactado para 50GB TB. Ainda não há economias reais realizadas porque o volume ainda é 100GB. O volume deve primeiro ser reduzido em tamanho para que o espaço guardado possa ser utilizado noutro local do sistema. Se as alterações posteriores ao LUN 100GBD resultarem em menos compressíveis os dados, o LUN crescerá em tamanho e o volume poderá ser preenchido.

O thin Provisioning é altamente recomendado porque pode simplificar o gerenciamento e fornecer melhorias substanciais na capacidade utilizável com economias de custo associadas. O motivo é simples: Os ambientes de banco de dados geralmente incluem muito espaço vazio, um grande número de volumes e LUNs e dados compressíveis. O provisionamento thick resulta na reserva de espaço no storage para volumes e LUNs, caso eles se tornem 100% cheios e contenham dados 100% não compactáveis. É pouco provável que isso ocorra. O thin Provisioning permite que esse espaço seja recuperado e usado em outros lugares e permite que o gerenciamento de capacidade seja baseado no próprio sistema de storage, em vez de muitos volumes e LUNs menores.

Alguns clientes preferem usar o provisionamento thick, seja para cargas de trabalho específicas ou, geralmente, com base em práticas operacionais e de aquisição estabelecidas.

Atenção: se um volume for provisionado de forma grossa, deve-se ter cuidado para desativar completamente todos os recursos de eficiência para esse volume, incluindo descompressão e remoção de deduplicação usando o `sis undo` comando. O volume não deve aparecer `volume efficiency show` na saída. Se isso acontecer, o volume ainda será parcialmente configurado para recursos de eficiência. Como resultado, as garantias de substituição funcionam de forma diferente, o que aumenta a chance de que a configuração seja ultrapassada fazendo com que o volume fique inesperadamente sem espaço, resultando em erros de e/S do banco de dados.

Práticas recomendadas de eficiência

A NetApp recomenda o seguinte:

Padrões do AFF

Os volumes criados no ONTAP executados em um sistema all-flash AFF são thin Provisioning com todos os recursos de eficiência in-line habilitados. Embora os bancos de dados geralmente não se beneficiem da deduplicação e possam incluir dados não compressíveis, as configurações padrão são, no entanto, apropriadas para quase todas as cargas de trabalho. O ONTAP foi projetado para processar com eficiência todos os tipos de dados e padrões de e/S, resultando ou não em economia. Os padrões só devem ser alterados se os motivos forem totalmente compreendidos e houver um benefício para se desviar.

Recomendações gerais

- Se os volumes e/ou LUNs não forem provisionados de forma fina, você deve desativar todas as configurações de eficiência porque o uso desses recursos não oferece economia e a combinação de provisionamento espesso com eficiência de espaço habilitada pode causar comportamento inesperado, incluindo erros fora do espaço.
- Se os dados não estiverem sujeitos a sobrescritas, como backups ou logs de transação de banco de

dados, você poderá obter maior eficiência ativando o TSSE com um período de resfriamento baixo.

- Alguns arquivos podem conter uma quantidade significativa de dados não compressíveis, por exemplo, quando a compactação já está ativada no nível de aplicativo de arquivos são criptografados. Se qualquer um desses cenários for verdadeiro, considere desativar a compactação para permitir uma operação mais eficiente em outros volumes que contenham dados compressíveis.
- Não use a compactação e a deduplicação do 32KB com backups de bancos de dados. Consulte a seção [Compressão adaptável](#) para obter detalhes.

Compactação de banco de dados

O próprio SQL Server também tem recursos para compactar e gerenciar dados de forma eficiente. O SQL Server atualmente suporta dois tipos de compactação de dados: Compactação de linha e compactação de página.

A compressão de linha altera o formato de armazenamento de dados. Por exemplo, ele muda inteiros e decimais para o formato de comprimento variável em vez de seu formato de comprimento fixo nativo. Ele também muda strings de caracteres de comprimento fixo para o formato de comprimento variável, eliminando espaços em branco. A compressão de página implementa a compressão de linha e duas outras estratégias de compressão (compressão de prefixo e compressão de dicionário). Pode encontrar mais detalhes sobre a compressão de páginas "[Implementação da compressão de páginas](#)" no .

Atualmente, a compactação de dados é suportada nas edições Enterprise, Developer e Evaluation do SQL Server 2008 e posterior. Embora a compactação possa ser executada pelo próprio banco de dados, isso raramente é observado em um ambiente SQL Server.

Aqui estão a recomendação para gerenciar espaço para arquivos de dados do SQL Server

- Use thin Provisioning em ambientes SQL Server para melhorar a utilização de espaço e reduzir os requisitos gerais de armazenamento quando a funcionalidade de garantia de espaço é usada.
 - Use o crescimento automático para as configurações de implantação mais comuns porque o administrador de storage precisa monitorar o uso de espaço no agregado.
- Não ative a deduplicação em quaisquer volumes que contenham arquivos de dados do SQL Server, a menos que o volume seja conhecido por conter várias cópias dos mesmos dados, como restaurar banco de dados de backups para um único volume.

Exigência de espaço

A recuperação de espaço pode ser iniciada periodicamente para recuperar espaço não utilizado em um LUN. Com o SnapCenter, você pode usar o seguinte comando do PowerShell para iniciar a recuperação de espaço.

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

Se você precisar executar a recuperação de espaço, esse processo deve ser executado durante períodos de baixa atividade, pois inicialmente consome ciclos no host.

Proteção de dados

As estratégias de backup de banco de dados devem ser baseadas em requisitos de negócios identificados, não em recursos teóricos. Ao combinar a tecnologia Snapshot da ONTAP e aproveitar as API do Microsoft SQL Server, você pode rapidamente fazer

backup consistente de aplicativos, independentemente do tamanho dos bancos de dados do usuário. Para requisitos de gerenciamento de dados mais avançados ou com escalabilidade horizontal, a NetApp oferece SnapCenter.

SnapCenter

O SnapCenter é o software de proteção de dados da NetApp para aplicações empresariais. Os bancos de dados do SQL Server podem ser protegidos rápida e facilmente com o plug-in do SnapCenter para SQL Server e com operações do SO gerenciadas pelo plug-in do SnapCenter para Microsoft Windows.

A instância do SQL Server pode ser uma configuração autônoma, uma instância de cluster de failover ou pode estar sempre no grupo de disponibilidade. O resultado é que, a partir de um painel único, os bancos de dados podem ser protegidos, clonados e restaurados da cópia primária ou secundária. O SnapCenter pode gerenciar bancos de dados do SQL Server tanto no local, na nuvem e configurações híbridas. Cópias de banco de dados também podem ser criadas em poucos minutos no host original ou alternativo para desenvolvimento ou para fins de geração de relatórios.



A NetApp recomenda usar o SnapCenter para criar cópias Snapshot. O método T-SQL descrito abaixo também funciona, mas o SnapCenter oferece automação completa sobre o processo de backup, restauração e clonagem. Ele também executa a descoberta para garantir que os snapshots corretos estejam sendo criados. Não é necessária pré-configuração.

O SQL Server também requer coordenação entre o sistema operacional e o armazenamento para garantir que os dados corretos estejam presentes em snapshots no momento da criação. Na maioria dos casos, o único método seguro para fazer isso é com SnapCenter ou T-SQL. Os instantâneos criados sem essa coordenação adicional podem não ser recuperáveis de forma confiável.

Para obter mais detalhes sobre o plug-in do SQL Server para SnapCenter, ["TR-4714: Guia de práticas recomendadas para SQL Server usando NetApp SnapCenter"](#) consulte .

Proteção de banco de dados usando snapshots T-SQL

No SQL Server 2022, a Microsoft introduziu snapshots T-SQL que oferece um caminho para a execução de scripts e automação de operações de backups. Em vez de executar cópias de tamanho completo, você pode preparar o banco de dados para snapshots. Depois que o banco de dados estiver pronto para backup, você poderá aproveitar as APIs REST do ONTAP para criar snapshots.

O seguinte é um exemplo de fluxo de trabalho de backup:

1. Congelar um banco de dados com comando ALTER. Ele prepara o banco de dados para um snapshot consistente no storage subjacente. Após o congelamento, você pode descongelar o banco de dados e gravar o instantâneo com o comando BACKUP.
2. Execute snapshots de vários bancos de dados nos volumes de storage simultaneamente com os novos comandos DO GRUPO DE backup e DO SERVIDOR DE BACKUP.
3. Execute backups COMPLETOS ou backups COMPLETOS Copy_ONLY. Esses backups também são gravados em msdb.
4. Execute recuperação pontual usando backups de log feitos com a abordagem de streaming normal após o backup COMPLETO do snapshot. Backups diferenciais de streaming também são suportados, se desejado.

Para saber mais, ["Documentação da Microsoft para saber sobre os snapshots T-SQL"](#) consulte .

Recuperação de desastres

Recuperação de desastres

Bancos de dados empresariais e infraestruturas de aplicações geralmente exigem replicação para proteger contra desastres naturais ou interrupções inesperadas dos negócios com o mínimo de tempo de inatividade.

O recurso de replicação de grupo de disponibilidade contínua do SQL Server pode ser uma excelente opção, e o NetApp oferece opções para integrar a proteção de dados sempre ativa. Em alguns casos, no entanto, você pode querer considerar a tecnologia de replicação do ONTAP. Existem três opções básicas.

SnapMirror

A tecnologia SnapMirror oferece uma solução empresarial rápida e flexível para replicação de dados em LANs e WANs. A tecnologia SnapMirror transfere apenas blocos de dados alterados para o destino após a criação do espelhamento inicial, reduzindo significativamente os requisitos de largura de banda da rede. Ele pode ser configurado em modo síncrono ou assíncrono.

Sincronização ativa do NetApp MetroCluster e do SnapMirror

Para muitos clientes, a recuperação de desastres requer mais do que apenas a posse de uma cópia remota de dados. Para isso, ela exige a capacidade de usar esses dados rapidamente. A NetApp oferece duas tecnologias que atendem a essa necessidade: MetroCluster e SnapMirror active Sync

O MetroCluster se refere ao ONTAP em uma configuração de hardware que inclui armazenamento de baixo nível espelhado e vários recursos adicionais. Soluções integradas, como o MetroCluster, simplificam as infraestruturas de virtualização, aplicações e banco de dados complicadas e com escalabilidade horizontal atuais. Ele substitui vários produtos e estratégias de proteção de dados externos por um storage array simples e central. Ele também fornece backup integrado, recuperação, recuperação de desastres e alta disponibilidade (HA) em um único sistema de storage em cluster.

A sincronização ativa do SnapMirror é baseada no SnapMirror síncrono. Com o MetroCluster, cada controlador ONTAP é responsável por replicar os dados da unidade para um local remoto. Com o SnapMirror active Sync, você tem essencialmente dois sistemas ONTAP diferentes que mantêm cópias independentes dos seus dados LUN, mas cooperam para apresentar uma única instância desse LUN. Do ponto de vista do host, é uma única entidade LUN.

Comparação de SM-as e MCC

O SM-as e o MetroCluster são semelhantes na funcionalidade geral, mas há diferenças importantes na maneira como a replicação RPO-0 foi implementada e como ela é gerenciada. O SnapMirror assíncrono e síncrono também pode ser usado como parte de um plano de recuperação de desastres, mas não foram desenvolvidos como tecnologias de repliação de HA.

- Uma configuração do MetroCluster é mais como um cluster integrado com nós distribuídos entre locais. O SM-as se comporta como dois clusters independentes que estão cooperando no fornecimento de LUNs replicados de forma síncrona RPO igual a 0.
- Os dados em uma configuração do MetroCluster só podem ser acessados de um determinado site a qualquer momento. Uma segunda cópia dos dados está presente no site oposto, mas os dados são passivos. Ele não pode ser acessado sem um failover do sistema de storage.
- O espelhamento de desempenho MetroCluster e SM-as ocorre em níveis diferentes. O espelhamento

MetroCluster é executado na camada RAID. Os dados de baixo nível são armazenados em um formato espelhado usando SyncMirror. O uso do espelhamento é praticamente invisível nas camadas de LUN, volume e protocolo.

- Em contraste, o espelhamento SM-as ocorre na camada de protocolo. Os dois clusters são, no geral, clusters independentes. Depois que as duas cópias dos dados estiverem sincronizadas, os dois clusters só precisam espelhar gravações. Quando ocorre uma gravação em um cluster, ela é replicada para o outro cluster. A gravação só é reconhecida para o host quando a gravação for concluída em ambos os sites. Além desse comportamento de divisão de protocolo, os dois clusters são, de outra forma, clusters ONTAP normais.
- A função principal do MetroCluster é a replicação em grande escala. É possível replicar um array inteiro com RPO igual a 0 e rto quase zero. Isso simplifica o processo de failover porque há apenas uma "coisa" a fazer failover e é dimensionado extremamente bem em termos de capacidade e IOPS.
- Um dos principais casos de uso para SM-as é a replicação granular. Às vezes, você não quer replicar todos os dados como uma única unidade ou precisa ser capaz de falhar seletivamente em determinados workloads.
- Outro importante caso de uso para SM-as é para operações ativas-ativas, em que você deseja que cópias totalmente utilizáveis de dados estejam disponíveis em dois clusters diferentes localizados em dois locais diferentes com características de desempenho idênticas e, se desejado, não é necessário estender a SAN entre locais. Você pode ter suas aplicações já em execução em ambos os locais, o que reduz o rto geral durante operações de failover.

SnapMirror

Veja a seguir as recomendações do SnapMirror para SQL Server:

- Se o SMB for usado, o SVM de destino deve ser membro do mesmo domínio do ativo Directory do qual o SVM de origem é membro, de modo que as listas de controle de acesso (ACLs) armazenadas em arquivos nas não sejam quebradas durante a recuperação de um desastre.
- Não é necessário usar nomes de volume de destino iguais aos nomes de volume de origem, mas pode tornar o processo de montagem de volumes de destino no destino mais simples de gerenciar. Se o SMB for usado, você deve tornar o namespace nas de destino idêntico em caminhos e estrutura de diretório para o namespace de origem.
- Para fins de consistência, não programe atualizações do SnapMirror a partir dos controladores. Em vez disso, ative as atualizações do SnapMirror do SnapCenter para atualizar o SnapMirror após a conclusão da cópia de segurança completa ou de registro.
- Distribua volumes que contêm dados do SQL Server entre diferentes nós do cluster para permitir que todos os nós do cluster compartilhem a atividade de replicação do SnapMirror. Essa distribuição otimiza o uso de recursos de nós.
- Usar a replicação síncrona, onde a demanda por recuperação rápida de dados é mais alta e soluções assíncronas para flexibilidade no RPO.

Para obter mais informações sobre o SnapMirror, ["TR-4015: Guia de práticas recomendadas e configuração do SnapMirror para ONTAP 9"](#) consulte .

MetroCluster

Arquitetura

A implantação do Microsoft SQL Server com ambiente MetroCluster requer alguma explicação do design físico de um sistema MetroCluster.

O MetroCluster espelha de forma síncrona os dados e a configuração entre dois clusters ONTAP em locais separados ou domínios de falha. O MetroCluster fornece storage disponível continuamente para aplicações, gerenciando automaticamente dois objetivos:

- Objetivo do ponto de restauração zero (RPO) espelhando de forma síncrona os dados gravados no cluster.
- Objetivo de tempo de recuperação quase zero (rto) espelhando a configuração e automatizando o acesso aos dados no segundo local.

O MetroCluster oferece simplicidade com o espelhamento automático de dados e a configuração entre os dois clusters independentes localizados nos dois locais. À medida que o storage é provisionado em um cluster, ele é espelhado automaticamente para o segundo cluster no segundo local. O NetApp SyncMirror fornece uma cópia completa de todos os dados com RPO zero. Isso significa que workloads de um local podem alternar a qualquer momento para o local oposto e continuar fornecendo dados sem perda de dados. O MetroCluster gerencia o processo de switchover fornecendo acesso a dados provisionados por SAN e nas no segundo local. O design do MetroCluster como uma solução validada contém dimensionamento e configuração que permitem que um switchover seja executado dentro dos períodos de tempo limite do protocolo ou antes (geralmente menos de 120 segundos). Isso resulta em um RPO quase zero e as aplicações podem continuar acessando os dados sem incorrer em falhas. O MetroCluster está disponível em várias variações definidas pela malha de storage de back-end.

O MetroCluster pode ser usado em 3 configurações diferentes

- Pares HA com conectividade IP
- Pares DE HA com conectividade FC
- Controladora única com conectividade FC



O termo 'conectividade' refere-se à conexão de cluster usada para replicação entre sites. Não se refere aos protocolos de host. Todos os protocolos do lado do host são suportados como de costume em uma configuração MetroCluster, independentemente do tipo de conexão usada para comunicação entre clusters.

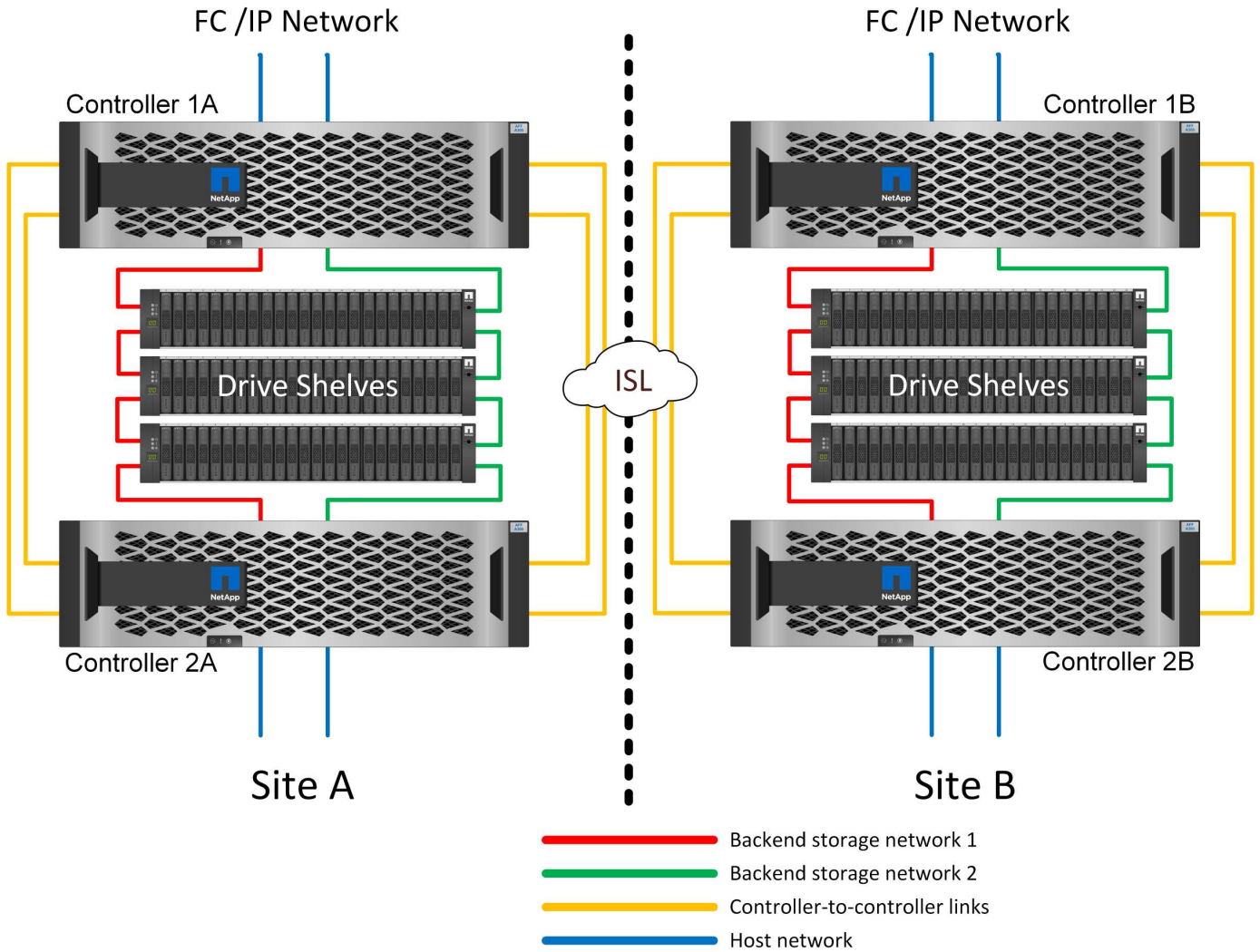
IP MetroCluster

A configuração MetroCluster IP de par de HA usa dois ou quatro nós por local. Essa opção de configuração aumenta a complexidade e os custos em relação à opção de dois nós, mas oferece um benefício importante: A redundância intrasite. Uma simples falha do controlador não requer acesso aos dados na WAN. O acesso aos dados permanece local por meio do controlador local alternativo.

A maioria dos clientes está escolhendo a conectividade IP porque os requisitos de infraestrutura são mais simples. No passado, a conectividade entre locais de alta velocidade era geralmente mais fácil de provisionar usando switches FC e fibra escura, mas hoje em dia os circuitos IP de baixa latência e alta velocidade estão mais prontamente disponíveis.

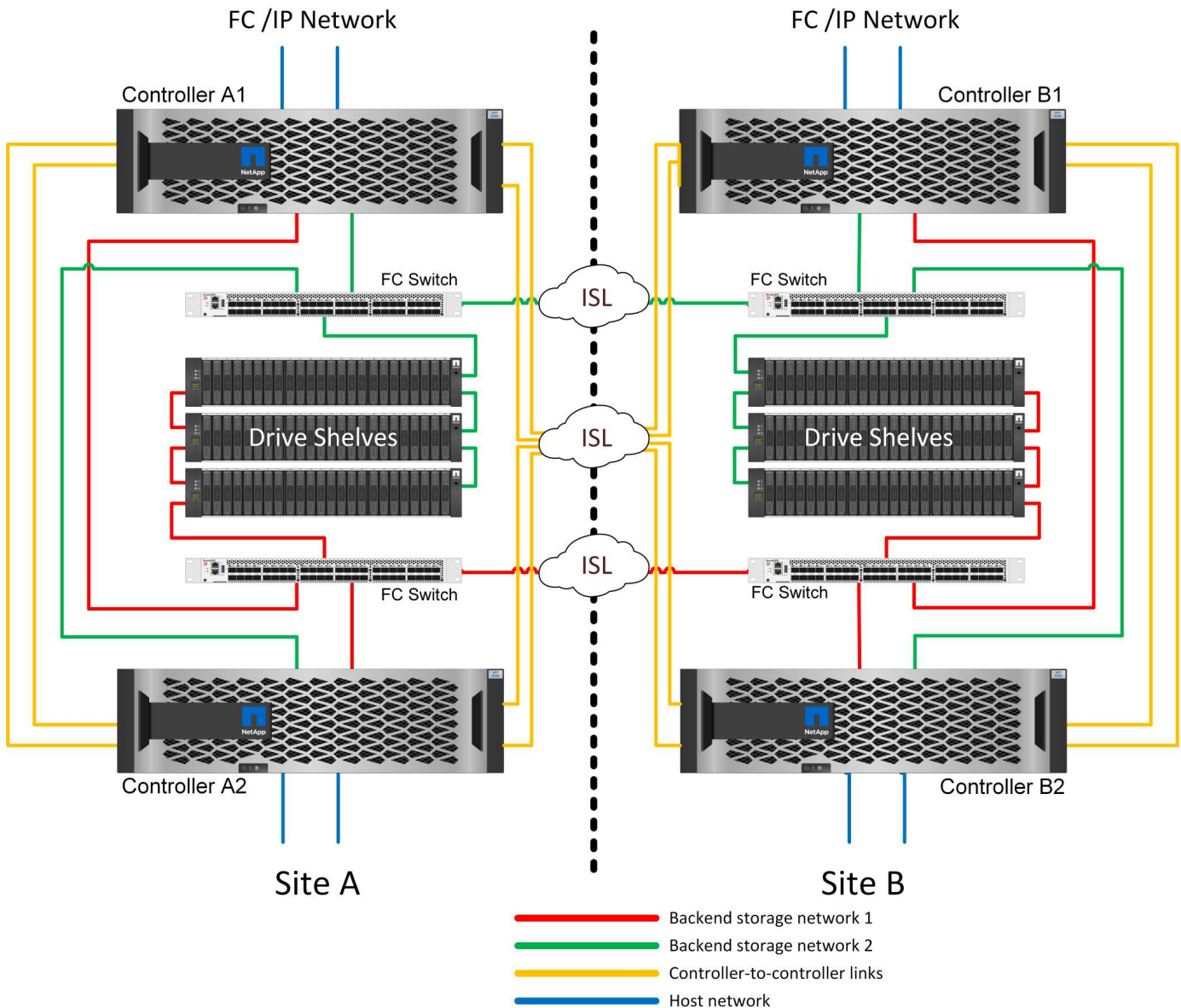
A arquitetura também é mais simples porque as únicas conexões entre locais são para os controladores. No Metroclusters FC SAN conectados, um controlador grava diretamente nas unidades no local oposto e, portanto, requer conexões, switches e bridges SAN adicionais. Em contraste, um controlador em uma configuração IP grava nas unidades opostas através do controlador.

Para obter informações adicionais, consulte a documentação oficial do ONTAP e "[Arquitetura e design da solução IP da MetroCluster](#)".



MetroCluster conectados a FC de par HA SAN

A configuração de MetroCluster FC de par de HA usa dois ou quatro nós por local. Essa opção de configuração aumenta a complexidade e os custos em relação à opção de dois nós, mas oferece um benefício importante: A redundância intrasite. Uma simples falha do controlador não requer acesso aos dados na WAN. O acesso aos dados permanece local por meio do controlador local alternativo.

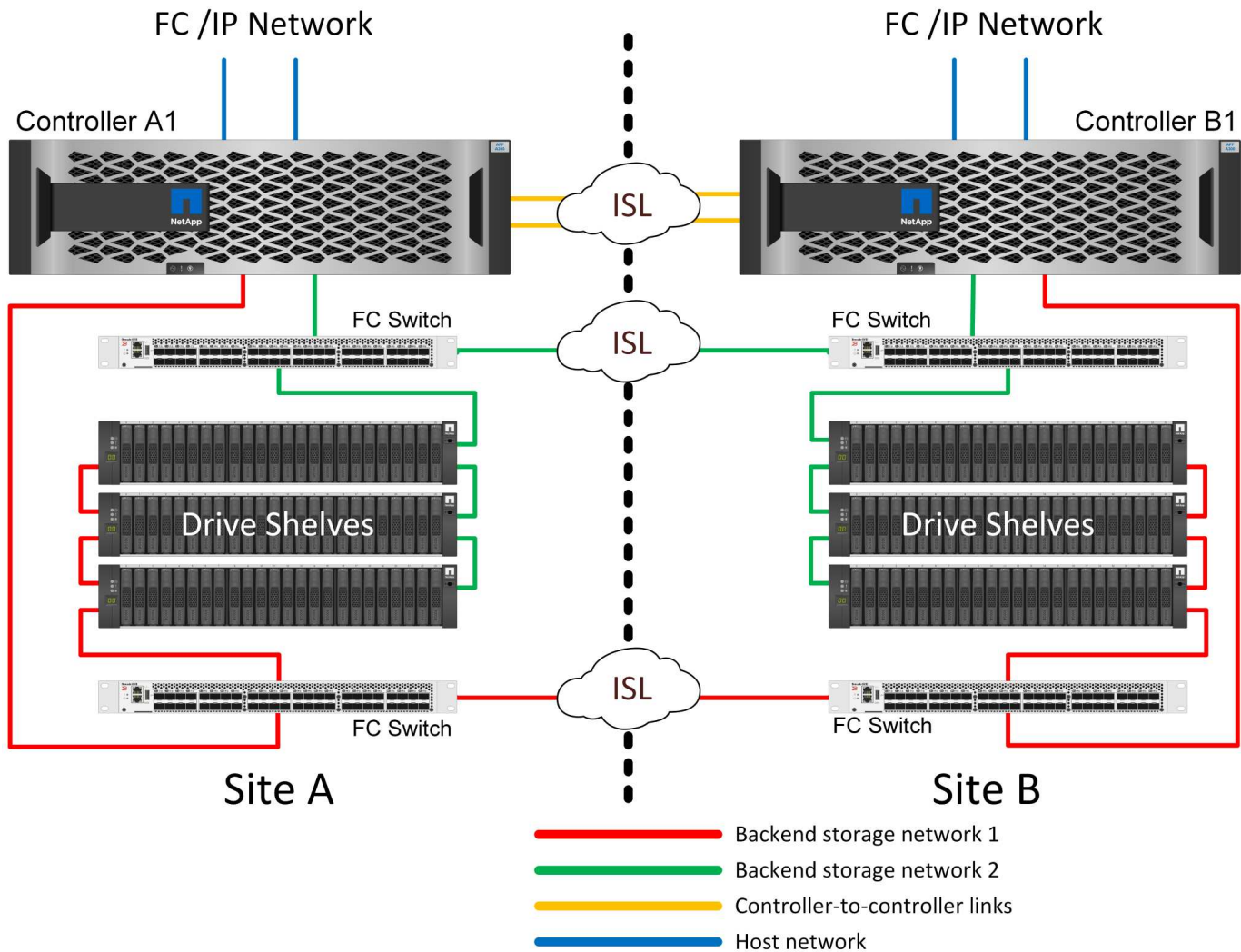


Algumas infraestruturas multisite não foram desenvolvidas para operações ativas-ativas, mas são usadas mais como local principal e local de recuperação de desastres. Nesta situação, uma opção MetroCluster de par de HA é geralmente preferível pelas seguintes razões:

- Embora um cluster de dois nós MetroCluster seja um sistema de HA, a falha inesperada de uma controladora ou a manutenção planejada exige que os serviços de dados fiquem online no local oposto. Se a conectividade de rede entre locais não puder suportar a largura de banda necessária, o desempenho é afetado. A única opção seria também falhar sobre os vários sistemas operacionais host e serviços associados ao site alternativo. O cluster de MetroCluster de par de HA elimina esse problema porque a perda de uma controladora resulta em failover simples no mesmo local.
- Algumas topologias de rede não são projetadas para acesso entre sites, mas usam sub-redes diferentes ou SANs FC isoladas. Nesses casos, o cluster MetroCluster de dois nós não funciona mais como um sistema de HA porque o controlador alternativo não pode fornecer dados aos servidores no local oposto. A opção MetroCluster de par de HA é necessária para fornecer redundância completa.
- Se uma infraestrutura de dois locais for vista como uma única infraestrutura altamente disponível, a configuração de dois nós do MetroCluster será adequada. No entanto, se o sistema precisar funcionar por um longo período de tempo após a falha do local, é preferível usar um par de HA porque ele continua fornecendo HA em um único local.

MetroCluster com conexão SAN FC de dois nós

A configuração do MetroCluster de dois nós usa apenas um nó por local. Esse design é mais simples do que a opção de par de HA, pois há menos componentes para configurar e manter. Ele também reduziu as demandas de infraestrutura em termos de cabeamento e switch FC. Finalmente, reduz custos.



O impacto óbvio desse projeto é que a falha do controlador em um único local significa que os dados estão disponíveis no local oposto. Esta restrição não é necessariamente um problema. Muitas empresas têm operações de data center multisite com redes estendidas, de alta velocidade e baixa latência, que funcionam essencialmente como uma única infraestrutura. Nesses casos, a versão de dois nós do MetroCluster é a configuração preferida. Sistemas de dois nós são usados atualmente em escala de petabyte por vários provedores de serviços.

Recursos de resiliência do MetroCluster

Não há pontos únicos de falha em uma solução MetroCluster:

- Cada controladora tem dois caminhos independentes para os compartimentos de unidades no local.
- Cada controladora tem dois caminhos independentes para o shelves de unidades no local remoto.
- Cada controlador tem dois caminhos independentes para os controladores no local oposto.
- Na configuração de par de HA, cada controladora tem dois caminhos para seu parceiro local.

Em resumo, qualquer componente na configuração pode ser removido sem comprometer a capacidade do MetroCluster de fornecer dados. A única diferença em termos de resiliência entre as duas opções é que a versão do par de HA ainda é um sistema de storage de HA geral após uma falha do local.

SyncMirror

A proteção do SQL Server com MetroCluster é baseada no SyncMirror, que oferece uma tecnologia de espelhamento síncrono com escalabilidade horizontal e alta performance.

Proteção de dados com o SyncMirror

No nível mais simples, a replicação síncrona significa que qualquer alteração deve ser feita em ambos os lados do storage espelhado antes que seja reconhecida. Por exemplo, se um banco de dados estiver escrevendo um log ou se um convidado VMware estiver sendo corrigido, uma gravação nunca deve ser perdida. Como um nível de protocolo, o sistema de storage não deve reconhecer a gravação até que ela tenha sido comprometida com a Mídia não volátil em ambos os locais. Só então é seguro prosseguir sem o risco de perda de dados.

O uso de uma tecnologia de replicação síncrona é a primeira etapa no projeto e gerenciamento de uma solução de replicação síncrona. A consideração mais importante é entender o que poderia acontecer durante vários cenários de falha planejados e não planejados. Nem todas as soluções de replicação síncrona oferecem os mesmos recursos. Se você precisa de uma solução que forneça um objetivo de ponto de restauração (RPO) zero, o que significa perda de dados zero, é necessário considerar todos os cenários de falha. Em particular, qual é o resultado esperado quando a replicação é impossível devido à perda de conectividade entre sites?

Disponibilidade de dados do SyncMirror

A replicação do MetroCluster é baseada na tecnologia NetApp SyncMirror, projetada para entrar e sair do modo síncrono com eficiência. Essa funcionalidade atende aos requisitos dos clientes que exigem replicação síncrona, mas que também precisam de alta disponibilidade para seus serviços de dados. Por exemplo, se a conectividade a um local remoto for cortada, geralmente é preferível que o sistema de armazenamento continue operando em um estado não replicado.

Muitas soluções de replicação síncrona só são capazes de operar no modo síncrono. Esse tipo de replicação tudo ou nada é às vezes chamado de modo domino. Esses sistemas de storage param de fornecer dados em vez de permitir que cópias locais e remotas dos dados fiquem não sincronizadas. Se a replicação for violada à força, a resincronização pode ser extremamente demorada e pode deixar um cliente exposto à perda completa de dados durante o tempo em que o espelhamento é restabelecido.

O SyncMirror não só pode alternar facilmente do modo síncrono se o local remoto não estiver acessível, como também pode sincronizar rapidamente para um estado RPO de 0 quando a conectividade é restaurada. A cópia obsoleta dos dados no local remoto também pode ser preservada em um estado utilizável durante a resincronização, o que garante que cópias locais e remotas dos dados existam em todos os momentos.

Quando o modo domino é necessário, o NetApp oferece SnapMirror Synchronous (SM-S). Opções de nível de aplicativo também existem, como o Oracle DataGuard ou o SQL Server Always On Availability Groups. O espelhamento de disco no nível DO SO pode ser uma opção. Consulte sua equipe de conta do NetApp ou do parceiro para obter informações e opções adicionais.

SQL Server com MetroCluster

Uma opção para proteger bancos de dados SQL Server com RPO zero é o MetroCluster. O MetroCluster é uma tecnologia de replicação simples e de alta performance RPO igual

a 0 que permite replicar facilmente toda uma infraestrutura nos locais.

O SQL Server pode escalar até milhares de bancos de dados em um único sistema MetroCluster. Pode haver instâncias autônomas do SQL Server ou instâncias de cluster de failover, o sistema MetroCluster não necessariamente adiciona ou altera quaisquer práticas recomendadas para gerenciar um banco de dados.

Uma explicação completa do MetroCluster está além do escopo deste documento, mas os princípios são simples. O MetroCluster pode fornecer uma solução de replicação RPO igual a 0 com failover rápido. O que você constrói sobre essa base depende de suas necessidades.

Por exemplo, um procedimento básico de DR rápido após perda súbita do local pode usar as seguintes etapas básicas:

- Forçar um switchover do MetroCluster
- Realizar descoberta de LUNs FC/iSCSI (somente SAN)
- Montar sistemas de arquivos
- Inicie o SQL Services

O principal requisito dessa abordagem é um sistema operacional em execução no local remoto. Ele deve ser pré-configurado com a configuração do SQL Server e deve ser atualizado com versão de compilação equivalente. Os bancos de dados do sistema do SQL Server também podem ser espelhados para o local remoto e montados se um desastre for declarado.

Se os volumes, sistemas de arquivos e bancos de dados virtualizados que hospedam bancos de dados não estiverem em uso no local de recuperação de desastres antes do switchover, não há nenhum requisito para definir `dr-force- nvfail` os volumes associados.

Sincronização ativa do SnapMirror

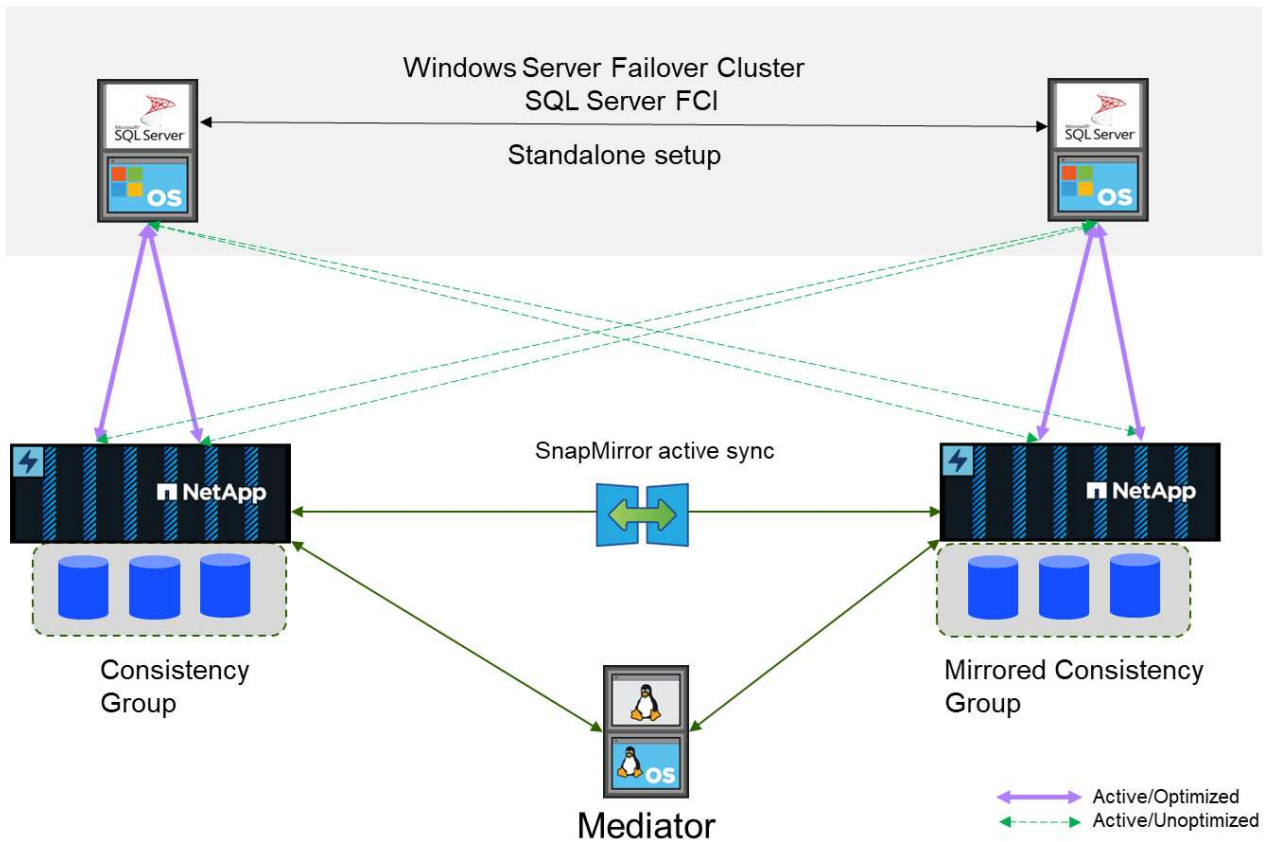
Visão geral

O SnapMirror active Sync permite que bancos de dados e aplicações individuais do SQL Server continuem as operações durante interrupções de storage e rede, com failover transparente de storage sem qualquer intervenção manual.

Iniciando o ONTAP 9.15.1, o SnapMirror ativo Sync suporta arquitetura ativa-ativa simétrica, além da configuração assimétrica existente. A funcionalidade ativo-ativo simétrica fornece replicação bidirecional síncrona para continuidade dos negócios e recuperação de desastres. Ele ajuda você a proteger o acesso a dados para workloads SAN críticos com acesso de leitura e gravação simultâneos a dados em vários domínios de falha, garantindo operações ininterruptas e minimizando o tempo de inatividade durante desastres ou falhas do sistema.

Os hosts do SQL Server acessam o storage usando Fibre Channel (FC) ou iSCSI LUNs. Replicação entre cada cluster que hospeda uma cópia dos dados replicados. Como esse recurso é replicação no nível de armazenamento, as instâncias do SQL Server executadas em instâncias de cluster de host ou failover independentes podem executar operações de leitura/gravação em cluster. Para obter informações sobre as etapas de Planejamento e configuração, "[Documentação do ONTAP na sincronização ativa do SnapMirror](#)" consulte .

Arquitetura do SnapMirror ativo com ativo-ativo simétrico



Replicação síncrona

Em operação normal, cada cópia é uma réplica síncrona RPO/0 em todos os momentos, com uma exceção. Se os dados não puderem ser replicados, o ONTAP cumprirá o requisito de replicar dados e retomar a distribuição de I/O em um local, enquanto os LUNs no outro local ficam offline.

Hardware de armazenamento

Ao contrário de outras soluções de recuperação de desastres de storage, o SnapMirror active Sync oferece flexibilidade assimétrica de plataforma. O hardware em cada local não precisa ser idêntico. Esse recurso permite dimensionar corretamente o hardware usado para suportar a sincronização ativa do SnapMirror. O sistema de storage remoto pode ser idêntico ao local principal se precisar dar suporte a uma carga de trabalho de produção completa, mas se um desastre resultar em e/S reduzida, do que um sistema menor no local remoto pode ser mais econômico.

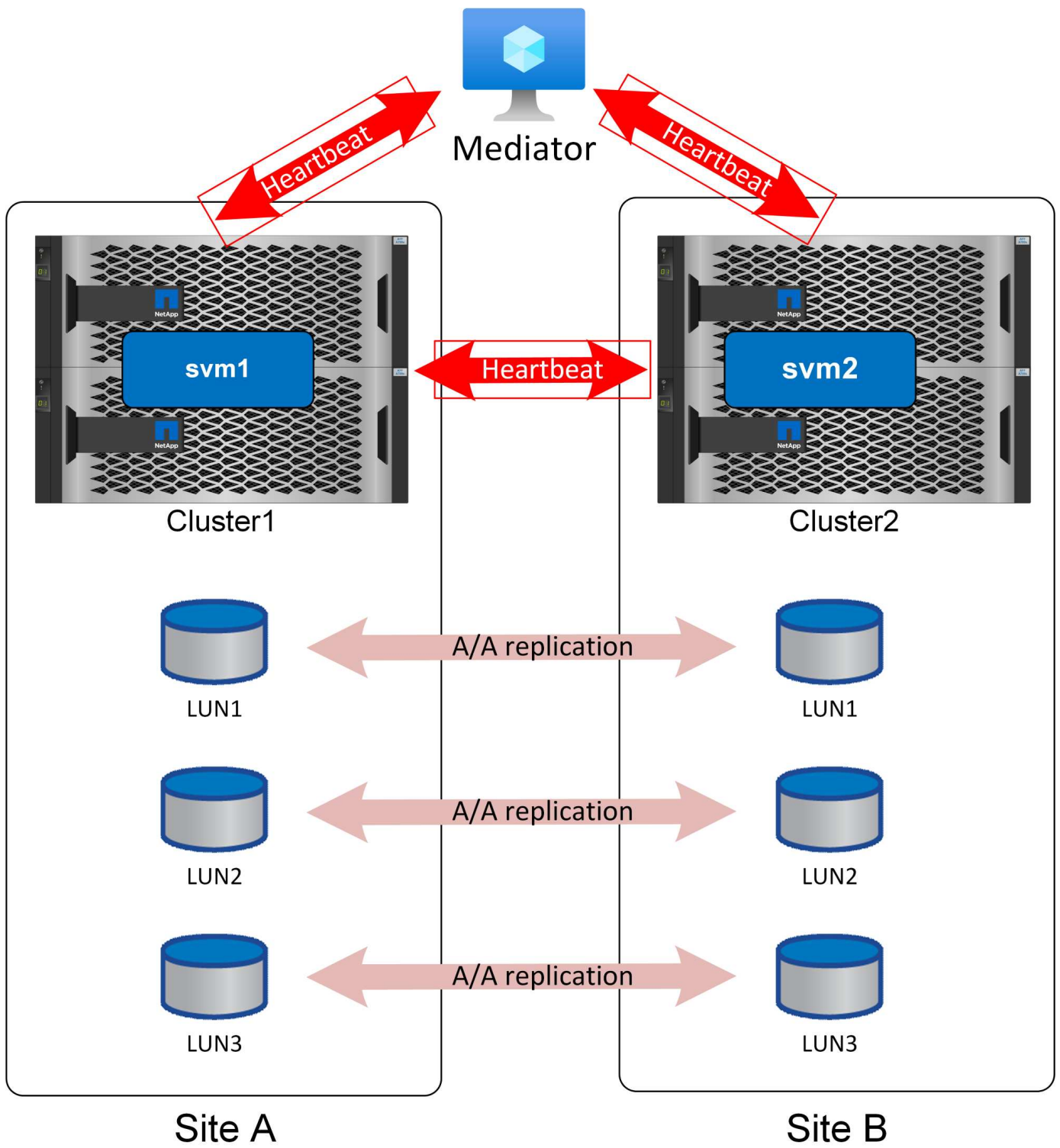
Mediador ONTAP

O Mediador ONTAP é um aplicativo de software que é baixado do suporte do NetApp e normalmente é implantado em uma pequena máquina virtual. O Mediador ONTAP não é um tiebreaker. É um canal de comunicação alternativo para os dois clusters que participam da replicação de sincronização ativa do SnapMirror. As operações automatizadas são orientadas pelo ONTAP com base nas respostas recebidas do parceiro por meio de conexões diretas e por meio do mediador.

Mediador do ONTAP

O mediador é necessário para automatizar o failover com segurança. Idealmente, ele seria colocado em um local 3rd independente, mas ainda pode funcionar para a maioria das necessidades se colocasse em um dos clusters que participam da replicação.

O mediador não é realmente um tiebreaker, embora essa seja efetivamente a função que ele fornece. Ele não realiza nenhuma ação; em vez disso, fornece um canal de comunicação alternativo para comunicação de cluster para cluster.



O desafio nº 1 com failover automatizado é o problema de split-brain, e esse problema surge se os dois locais perderem a conectividade entre si. O que deve acontecer? Você não quer que dois sites diferentes se designem como as cópias sobreviventes dos dados, mas como um único site pode dizer a diferença entre a perda real do site oposto e a incapacidade de se comunicar com o site oposto?

É aqui que o mediador entra na imagem. Se for colocado em um site 3rd e cada site tiver uma conexão de rede separada com esse site, então você terá um caminho adicional para cada site validar a integridade do outro. Olhe para a imagem acima novamente e considere os seguintes cenários.

- O que acontece se o mediador falhar ou não estiver acessível a partir de um ou de ambos os sites?
 - Os dois clusters ainda podem se comunicar entre si pelo mesmo link usado para serviços de replicação.
 - Os dados ainda são servidos com proteção RPO igual a 0
- O que acontece se o Site A falhar?
 - O local B verá ambos os canais de comunicação diminuírem.
 - O local B assumirá os serviços de dados, mas sem o espelhamento RPO igual a 0
- O que acontece se o local B falhar?
 - O local A verá ambos os canais de comunicação diminuírem.
 - O local A assumirá os serviços de dados, mas sem o espelhamento do RPO igual a 0

Há um outro cenário a considerar: Perda do link de replicação de dados. Se o link de replicação entre locais for perdido, o espelhamento RPO 0 obviamente será impossível. O que deve acontecer então?

Isso é controlado pelo status do site preferido. Em uma relação SM-as, um dos locais é secundário ao outro. Isso não tem efeito nas operações normais, e todo o acesso aos dados é simétrico, mas se a replicação for interrompida, o empate terá que ser quebrado para retomar as operações. O resultado é que o local preferido continuará as operações sem espelhamento, e o local secundário interromperá o processamento de e/S até que a comunicação de replicação seja restaurada.

Site preferido

O comportamento de sincronização ativa do SnapMirror é simétrico, com uma exceção importante - configuração de site preferida.

A sincronização ativa do SnapMirror considerará um site a "fonte" e o outro o "destino". Isso implica uma relação de replicação unidirecional, mas isso não se aplica ao comportamento de IO. A replicação é bidirecional e simétrica, e os tempos de resposta de e/S são os mesmos em ambos os lados do espelho.

A `source` designação é controla o local preferido. Se o link de replicação for perdido, os caminhos de LUN na cópia de origem continuarão a servir dados enquanto os caminhos de LUN na cópia de destino ficarão indisponíveis até que a replicação seja restabelecida e o SnapMirror reinsira um estado síncrono. Os caminhos irão então retomar a veiculação de dados.

A configuração de origem/destino pode ser visualizada através do SystemManager:

Relationships

Local destinations Local sources

Search Download Show/hide Filter

Source	Destination	Policy type
▼ jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	Synchronous

Ou na 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
```

O segredo é que a fonte é o SVM no cluster1. Como mencionado acima, os termos "fonte" e "destino" não descrevem o fluxo de dados replicados. Ambos os sites podem processar uma gravação e replicá-la para o site oposto. Com efeito, ambos os clusters são fontes e destinos. O efeito de designar um cluster como uma fonte simplesmente controla qual cluster sobrevive como um sistema de armazenamento de leitura e gravação se o link de replicação for perdido.

Topologia de rede

Acesso uniforme

Rede de acesso uniforme significa que os hosts são capazes de acessar caminhos em ambos os sites (ou domínios de falha dentro do mesmo site).

Um recurso importante do SM-as é a capacidade de configurar os sistemas de storage para saber onde os hosts estão localizados. Quando você mapeia os LUNs para um determinado host, você pode indicar se eles são ou não proximais a um determinado sistema de armazenamento.

Definições de proximidade

Proximidade refere-se a uma configuração por cluster que indica que um determinado host WWN ou ID de iniciador iSCSI pertence a um host local. É uma segunda etapa opcional para configurar o acesso LUN.

O primeiro passo é a configuração usual do igroup. Cada LUN deve ser mapeado para um grupo que contenha as IDs WWN/iSCSI dos hosts que precisam de acesso a esse LUN. Isso controla qual host tem *access* para um LUN.

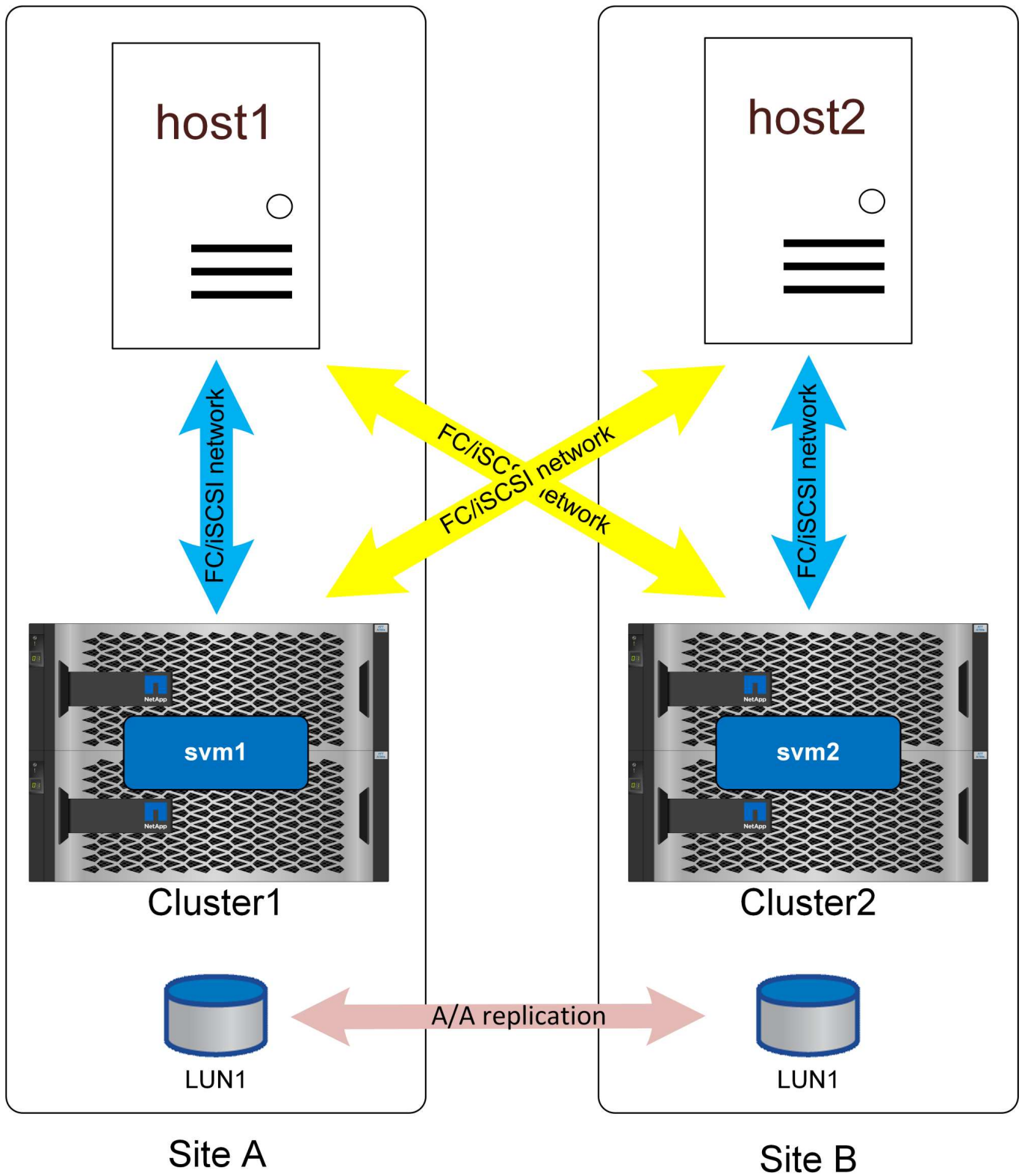
A segunda etapa opcional é configurar a proximidade do host. Isso não controla o acesso, ele controla *Priority*.

Por exemplo, um host no local A pode ser configurado para acessar um LUN que é protegido pela sincronização ativa do SnapMirror e, como a SAN é estendida entre sites, há caminhos disponíveis para esse LUN usando armazenamento no local A ou armazenamento no local B.

Sem configurações de proximidade, esse host usará ambos os sistemas de storage igualmente porque ambos os sistemas de storage anunciarão caminhos ativos/otimizados. Se a latência da SAN e/ou a largura de banda entre locais for limitada, isso pode não ser desejado e você pode querer garantir que, durante a operação normal, cada host utilize preferencialmente caminhos para o sistema de armazenamento local. Isso é configurado adicionando o ID WWN/iSCSI do host ao cluster local como um host proximal. Isso pode ser feito na CLI ou no SystemManager.

AFF

Com um sistema AFF, os caminhos aparecerão como mostrado abaixo quando a proximidade do host for configurada.



Active/Optimized Path

Active Path

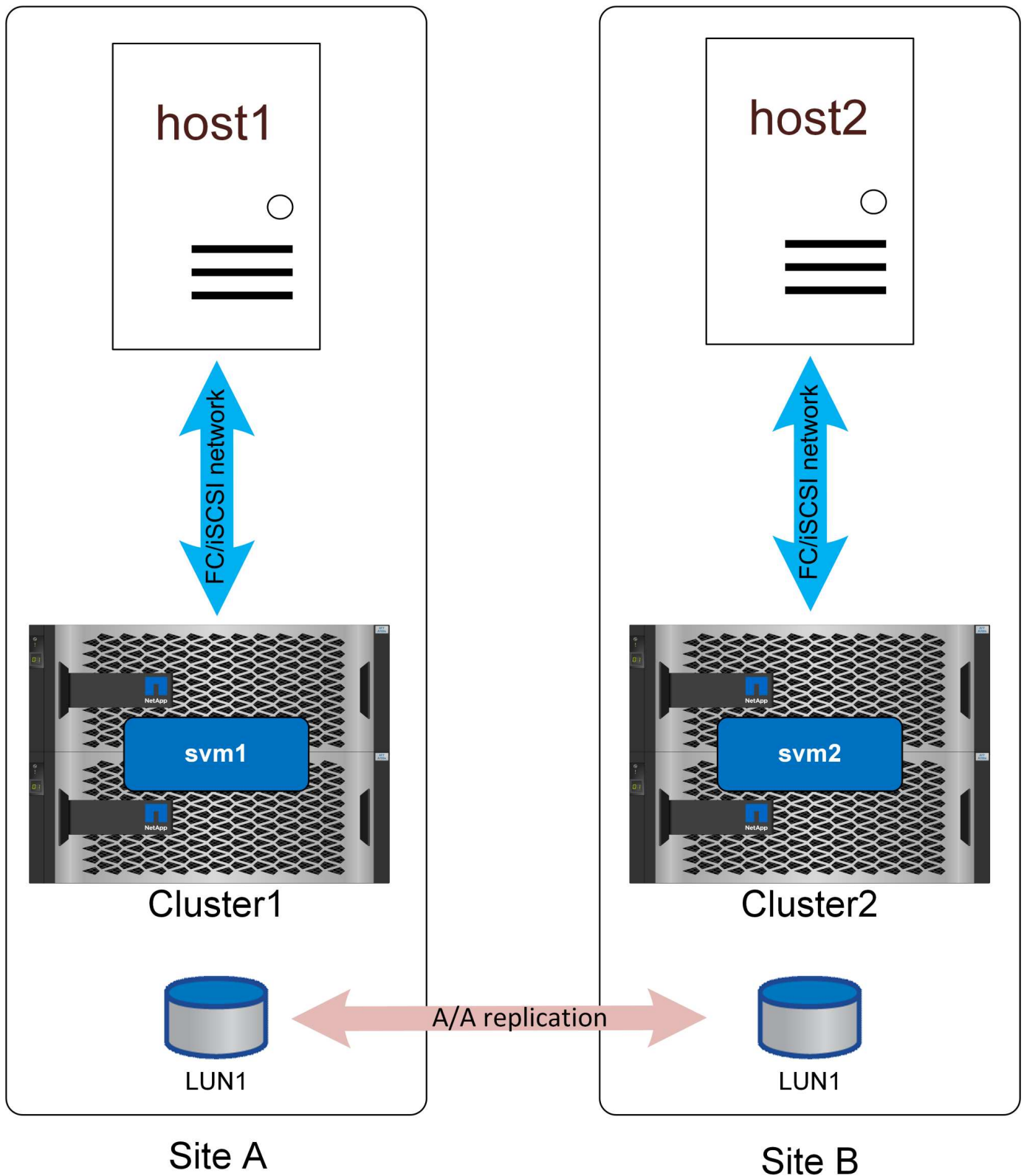
Em operação normal, todo o IO é IO local. Leituras e gravações são atendidas a partir do storage array local. É claro que o write IO também precisará ser replicado pelo controlador local para o sistema remoto antes de ser reconhecido, mas todas as IO de leitura serão atendidas localmente e não incorrerão latência extra ao atravessar o link SAN entre locais.

A única vez que os caminhos não otimizados serão usados é quando todos os caminhos ativos/otimizados forem perdidos. Por exemplo, se todo o array no local perder energia, os hosts no local A ainda poderão acessar caminhos para o array no local B e, portanto, permanecer operacionais, embora estejam com maior latência.

Há caminhos redundantes pelo cluster local que não são mostrados nesses diagramas por uma questão de simplicidade. Os sistemas de storage da ONTAP estão HA, portanto, uma falha da controladora não deve resultar em falha do local. Deve apenas resultar em uma mudança na qual os caminhos locais são usados no site afetado.

ASA

Os sistemas NetApp ASA oferecem multipathing ativo-ativo em todos os caminhos em um cluster. Isso também se aplica às configurações SM-as.



Active/Optimized Path

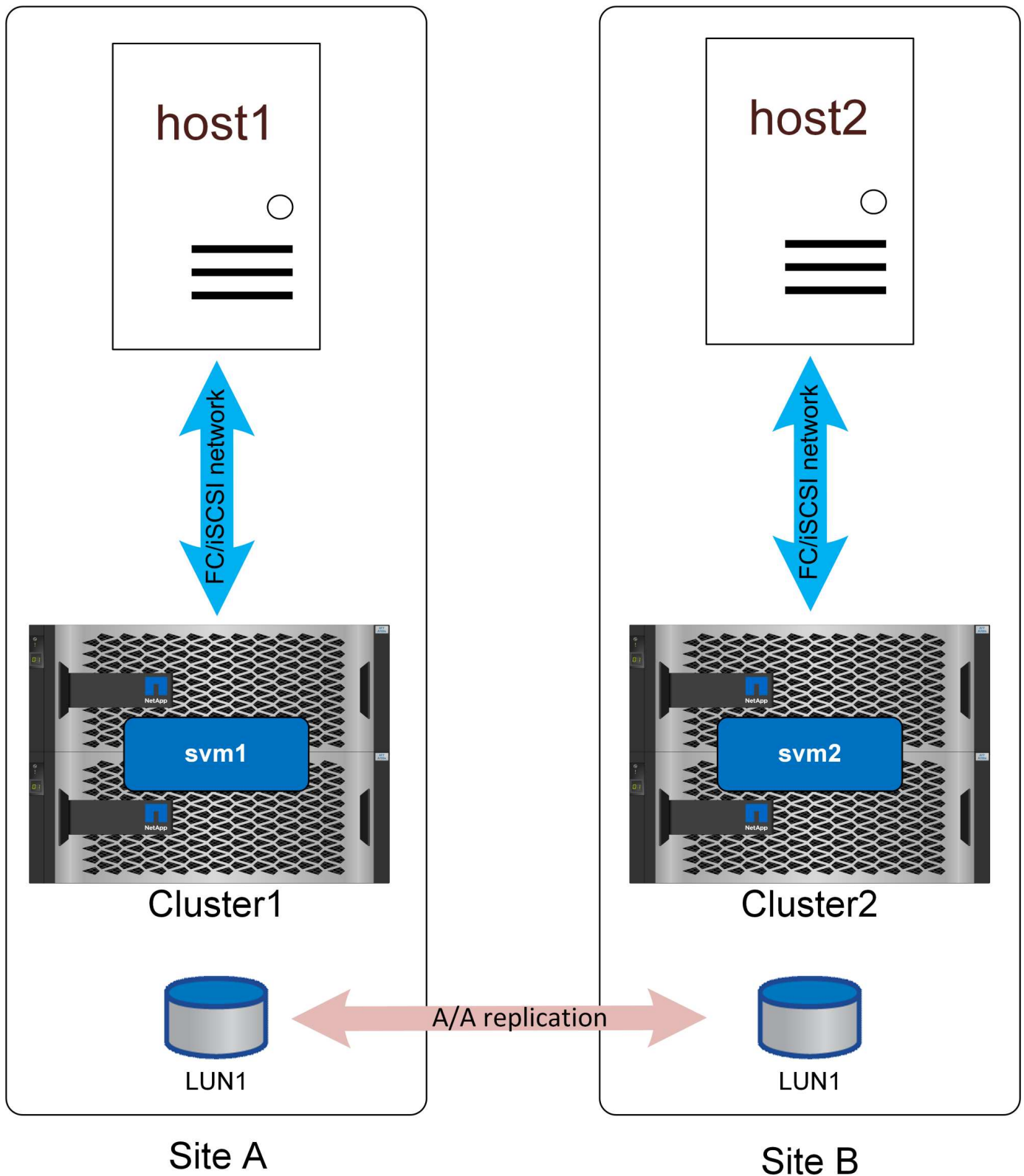
Uma configuração ASA com acesso não uniforme funcionaria em grande parte da mesma forma que faria com o AFF. Com acesso uniforme, o IO estaria atravessando a WAN. Isto pode ou não ser desejável.

Se os dois locais estivessem a 100 metros de distância com conectividade de fibra, não deveria haver latência adicional detetável cruzando a WAN, mas se os locais estivessem a uma distância longa, então o desempenho de leitura sofreria em ambos os locais. Em contraste, com o AFF, esses caminhos de cruzamento de WAN só seriam usados se não houvesse caminhos locais disponíveis e o desempenho do dia-a-dia seria melhor porque todo o IO seria IO local. O ASA com rede de acesso não uniforme seria uma opção para obter os benefícios de custo e recursos do ASA sem incorrer em uma penalidade de acesso à latência entre locais.

O ASA com SM-as em uma configuração de baixa latência oferece dois benefícios interessantes. Primeiro, ele basicamente **dobra** a performance de qualquer host porque a e/S pode ser atendida por duas vezes mais controladoras usando o dobro de caminhos. Em segundo lugar, em um ambiente de local único, ele oferece disponibilidade extrema porque todo um sistema de storage pode ser perdido sem interromper o acesso de host.

Acesso não uniforme

A rede de acesso não uniforme significa que cada host só tem acesso às portas no sistema de storage local. A SAN não é estendida entre sites (ou domínios de falha dentro do mesmo site).



Active/Optimized Path

O principal benefício dessa abordagem é a simplicidade da SAN - você elimina a necessidade de estender uma SAN pela rede. Alguns clientes não têm conectividade de baixa latência suficiente entre locais ou não têm

infraestrutura para túnel do tráfego SAN FC em uma rede entre locais.

A desvantagem para o acesso não uniforme é que certos cenários de falha, incluindo a perda do link de replicação, resultarão em alguns hosts perdendo acesso ao armazenamento. Os aplicativos que são executados como instâncias únicas, como um banco de dados não agrupado, que inerentemente está sendo executado apenas em um único host em qualquer montagem, falharão se a conectividade de armazenamento local for perdida. Os dados ainda seriam protegidos, mas o servidor de banco de dados não teria mais acesso. Ele precisaria ser reiniciado em um local remoto, de preferência através de um processo automatizado. Por exemplo, o VMware HA pode detectar uma situação de todos os caminhos em um servidor e reiniciar uma VM em outro servidor onde os caminhos estão disponíveis.

Em contraste, um aplicativo em cluster, como o Oracle RAC, pode fornecer um serviço que está disponível simultaneamente em dois locais diferentes. Perder um site não significa perda do serviço do aplicativo como um todo. As instâncias ainda estão disponíveis e em execução no local sobrevivente.

Em muitos casos, a sobrecarga de latência adicional de um aplicativo que acessa o storage em um link local a local seria inaceitável. Isso significa que a disponibilidade aprimorada de redes uniformes é mínima, uma vez que a perda de armazenamento em um local levaria à necessidade de encerrar serviços nesse local com falha de qualquer maneira.

Há caminhos redundantes pelo cluster local que não são mostrados nesses diagramas por uma questão de simplicidade. Os sistemas de storage da ONTAP estão HA, portanto, uma falha da controladora não deve resultar em falha do local. Deve apenas resultar em uma mudança na qual os caminhos locais são usados no site afetado.

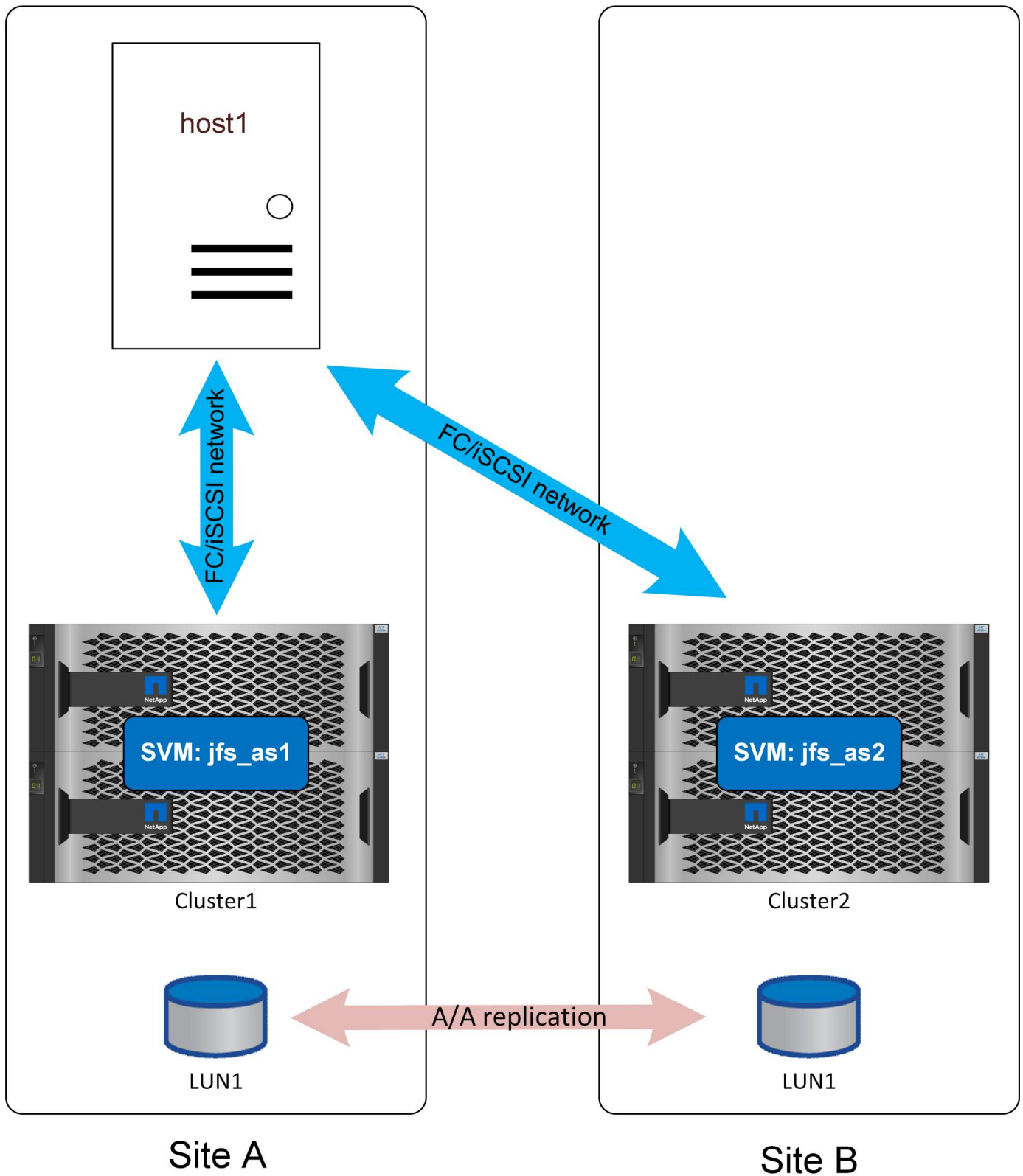
Visão geral

O SQL Server pode ser configurado para trabalhar com a sincronização ativa do SnapMirror de várias maneiras. A resposta certa depende da conectividade de rede disponível, dos requisitos de RPO e dos requisitos de disponibilidade.

Instância autônoma do SQL Server

As práticas recomendadas para layout de arquivo e configuração de servidor são as mesmas recomendadas ["SQL Server no ONTAP"](#) na documentação.

Com uma configuração autônoma, o SQL Server poderia estar sendo executado apenas em um site. Presumivelmente ["uniforme"](#) o acesso seria usado.



Com acesso uniforme, uma falha de armazenamento em qualquer um dos locais não interromperia as operações do banco de dados. Uma falha completa do local no site que incluía o servidor do banco de dados resultaria, naturalmente, em uma falha.

Alguns clientes podem configurar um sistema operacional em execução no site remoto com uma configuração pré-configurada do SQL Server, atualizada com uma versão de compilação equivalente como a da instância

de produção. O failover exigiria a ativação dessa instância autônoma do SQL Server no local alternativo, a descoberta dos LUNS e a inicialização do banco de dados. O processo completo pode ser automatizado com o cmdlet do Windows PowerShell, pois não são necessárias operações do lado do storage.

"Não uniforme" o acesso também poderia ser usado, mas o resultado seria uma interrupção do banco de dados se o sistema de storage em que o servidor de banco de dados estava localizado tivesse falhado porque o banco de dados não teria caminhos disponíveis para o storage. Isso ainda pode ser aceitável em alguns casos. O SnapMirror ativo Sync ainda estaria fornecendo proteção de dados RPO igual a 0 e, em caso de falha do local, a cópia sobrevivente estaria ativa e pronta para retomar as operações usando o mesmo procedimento usado com acesso uniforme, conforme descrito acima.

Um processo de failover simples e automatizado pode ser mais facilmente configurado com o uso de um host Virtualize. Por exemplo, se os arquivos de dados do SQL Server forem replicados sincronamente para o armazenamento secundário, juntamente com um VMDK de inicialização, em caso de desastre, o ambiente completo poderá ser ativado no local alternativo. Um administrador pode ativar manualmente o host no local sobrevivente ou automatizar o processo por meio de um serviço como o VMware HA.

Instância de cluster de failover do SQL Server

As instâncias de failover do SQL Server também podem ser hospedadas em um cluster de failover do Windows executado em um servidor físico ou servidor virtual como sistema operacional convidado. Essa arquitetura de vários hosts oferece resiliência de armazenamento e instância do SQL Server. Essa implantação é útil em ambientes de alta demanda que buscam processos de failover robustos, mantendo o desempenho aprimorado. Em uma configuração de cluster de failover, quando um host ou storage primário é afetado, o SQL Services será failover para o host secundário e, ao mesmo tempo, o storage secundário estará disponível para servir e/S. Nenhum script de automação ou intervenção do administrador é necessário.

Cenários de falha

Planejar uma arquitetura completa de aplicativos de sincronização ativa do SnapMirror requer entender como o SM-as responderá em vários cenários de failover planejados e não planejados.

Para os exemplos a seguir, suponha que o site A esteja configurado como o site preferido.

Perda de conectividade de replicação

Se a replicação SM-as for interrompida, não é possível concluir a e/S de gravação porque seria impossível que um cluster replique alterações no local oposto.

Local A (local preferido)

O resultado da falha do link de replicação no site preferido será uma pausa de aproximadamente 15 segundos no processamento de e/S de gravação, à medida que o ONTAP tenta novamente as operações de gravação replicadas antes de determinar que o link de replicação é genuinamente inacessível. Após os 15 segundos decorridos, o Site Um sistema retoma o processamento de e/S de leitura e escrita. Os caminhos de SAN não serão alterados e os LUNs permanecerão online.

Local B

Como o local B não é o site preferido de sincronização ativa do SnapMirror, seus caminhos de LUN ficarão indisponíveis após cerca de 15 segundos.

Falha do sistema de storage

O resultado de uma falha do sistema de armazenamento é quase idêntico ao resultado da perda do link de replicação. O local sobrevivente deve experimentar uma pausa de IO de aproximadamente 15 segundos. Uma vez decorrido esse período de 15 segundos, o IO será retomado nesse site como de costume.

Perda do mediador

O serviço mediador não controla diretamente as operações de storage. Ele funciona como um caminho de controle alternativo entre clusters. Ele existe principalmente para automatizar o failover sem o risco de um cenário de divisão cerebral. Em operação normal, cada cluster replica alterações em seu parceiro, e cada cluster pode verificar se o cluster do parceiro está on-line e fornecendo dados. Se o link de replicação falhar, a replicação cessaria.

O motivo pelo qual um mediador é necessário para o failover automatizado seguro é porque, de outra forma, seria impossível que um cluster de storage pudesse determinar se a perda de comunicação bidirecional foi o resultado de uma interrupção da rede ou falha real do storage.

O mediador fornece um caminho alternativo para cada cluster para verificar a integridade de seu parceiro. Os cenários são os seguintes:

- Se um cluster puder entrar em Contato diretamente com seu parceiro, os serviços de replicação estarão operacionais. Nenhuma ação necessária.
- Se um site preferido não puder entrar em Contato diretamente com seu parceiro ou por meio do mediador, ele assumirá que ele está realmente indisponível ou foi isolado e que levou seus caminhos LUN off-line. O site preferido continuará lançando o estado RPO/0 e continuará processando e/S de leitura e gravação.
- Se um site não preferencial não puder entrar em Contato diretamente com seu parceiro, mas puder contatá-lo por meio do mediador, ele tomará seus caminhos off-line e aguardará o retorno da conexão de replicação.
- Se um site não preferencial não puder entrar em Contato com seu parceiro diretamente ou por meio de um mediador operacional, ele assumirá que o parceiro está realmente indisponível ou foi isolado e que tomou seus caminhos LUN off-line. O site não preferencial lançará o estado RPO 0 e continuará processando e/S de leitura e gravação. Ele assumirá o papel da fonte de replicação e se tornará o novo site preferido.

Se o mediador não estiver totalmente disponível:

- A falha dos serviços de replicação por qualquer motivo, incluindo a falha do sistema de storage ou local não preferido, resultará no lançamento do estado RPO/0 e no reinício do processamento de e/S de leitura e gravação. O site não preferencial tomará seus caminhos off-line.
- A falha do site preferido resultará em uma falha porque o site não-preferido não será capaz de verificar se o site oposto está realmente off-line e, portanto, não seria seguro para o site não-preferido retomar os serviços.

Restauração de serviços

Depois que uma falha é resolvida, como restaurar a conectividade site-a-site ou ligar um sistema com falha, os pontos de extremidade de sincronização ativa do SnapMirror detetarão automaticamente a presença de uma relação de replicação com defeito e o devolverão ao estado RPO-0. Uma vez que a replicação síncrona for restabelecida, os caminhos com falha ficarão online novamente.

Em muitos casos, os aplicativos em cluster detetarão automaticamente o retorno de caminhos com falha, e esses aplicativos também voltarão online. Em outros casos, pode ser necessária uma análise SAN no nível do host ou os aplicativos podem precisar ser colocados online manualmente. Depende do aplicativo e como ele é

configurado e, em geral, essas tarefas podem ser facilmente automatizadas. O próprio ONTAP é com autorrecuperação e não deve exigir a intervenção do usuário para retomar as operações de storage RPO de 0.

Failover manual

Alterar o local preferido requer uma operação simples. A e/S pausa por um segundo ou dois como autoridade sobre os switches de comportamento de replicação entre clusters, mas a e/S não é afetada.

Segurança dos dados

Proteger um ambiente de banco de dados do SQL Server é um esforço multidimensional que vai além do gerenciamento do próprio banco de dados. O ONTAP oferece vários recursos exclusivos projetados para proteger o aspecto de armazenamento de sua infraestrutura de banco de dados.

Cópias Snapshot

Os instantâneos de armazenamento são réplicas pontuais dos dados de destino. A implementação do ONTAP inclui a capacidade de definir várias políticas e armazenar até 1024 instantâneos por volume. Os snapshots no ONTAP economizam espaço. O espaço só é consumido à medida que o conjunto de dados original muda. Eles também são somente leitura. Um instantâneo pode ser excluído, mas não pode ser alterado.

Em alguns casos, os snapshots podem ser agendados diretamente no ONTAP. Em outros casos, softwares como o SnapCenter podem ser necessários para orquestrar operações de aplicativos ou sistemas operacionais antes de criar snapshots. Qualquer que seja a melhor abordagem para seu workload, uma estratégia de snapshot agressiva pode fornecer segurança dos dados por meio de acesso frequente e fácil a backups de tudo, desde LUNs de inicialização a bancos de dados essenciais.

Nota: Um volume flexível ONTAP, ou mais simplesmente, um volume não é sinônimo de LUN. Os volumes são contêineres de gerenciamento de dados, como arquivos ou LUNs. Por exemplo, um banco de dados pode ser colocado em um conjunto de distribuição de 8 LUN, com todos os LUNs contidos em um único volume.

Para obter mais informações sobre instantâneos, clique em ["aqui."](#)

Snapshots à prova de violações

A partir do ONTAP 9.12.1, os instantâneos não são apenas de leitura, mas também podem ser protegidos contra exclusões acidentais ou intencionais. O recurso é chamado de Snapshots à prova de violações. Um período de retenção pode ser definido e aplicado por meio da política de snapshot. Os instantâneos resultantes não podem ser eliminados até que tenham atingido a data de validade. Não há substituições administrativas ou do centro de suporte.

Isso garante que um intruso, um insider malicioso, ou mesmo um ataque de ransomware não seja capaz de comprometer os backups, mesmo que isso tenha resultado no acesso ao próprio sistema ONTAP. Quando combinado a uma programação de snapshot frequente, o resultado é uma proteção de dados extremamente poderosa com RPO muito baixo.

Para obter mais informações sobre instantâneos à prova de violações, clique em ["aqui."](#)

Replicação SnapMirror

Os instantâneos também podem ser replicados para um sistema remoto. Isso inclui snapshots à prova de violações, onde o período de retenção é aplicado e aplicado no sistema remoto. O resultado são os mesmos benefícios de proteção de dados que os snapshots locais, mas eles ficam em um segundo storage array. Isso garante que a destruição do array original não comprometa os backups.

Um segundo sistema também abre novas opções de segurança administrativa. Por exemplo, alguns clientes do NetApp segregam credenciais de autenticação para os sistemas de storage primário e secundário. Nenhum usuário administrativo tem acesso a ambos os sistemas, o que significa que um administrador mal-intencionado não pode excluir todas as cópias de dados.

Para obter mais informações sobre o SnapMirror, clique em ["aqui."](#)

Máquinas virtuais de armazenamento

Um sistema de armazenamento ONTAP recém-configurado é semelhante a um servidor VMware ESX recém-provisionado porque nenhum deles pode suportar qualquer usuário até que uma máquina virtual seja criada. Com o ONTAP, você cria uma máquina virtual de storage (SVM) que se torna a unidade mais básica de gerenciamento de storage. Cada SVM tem seus próprios recursos de storage, configurações de protocolo, endereços IP e WWNs FCP. Esta é a base da alocação de vários clientes da ONTAP.

Por exemplo, você pode configurar um SVM para workloads de produção críticos e um segundo SVM em um segmento de rede diferente para atividades de desenvolvimento. Dessa forma, você poderia restringir o acesso ao SVM de produção a certos administradores, enquanto concederia aos desenvolvedores um controle mais expansivo sobre os recursos de storage no SVM de desenvolvimento. Você também pode precisar fornecer um terceiro SVM para suas equipes financeiras e de RH para armazenar dados especialmente críticos apenas para os olhos.

Para obter mais informações sobre SVMs, clique em ["aqui."](#)

RBAC administrativo

O ONTAP oferece controle de acesso baseado em funções (RBAC) avançado para logins administrativos. Alguns administradores podem precisar de acesso total ao cluster, enquanto outros podem precisar de acesso a determinados SVMs. O pessoal avançado do helpdesk pode precisar da capacidade de aumentar o tamanho dos volumes. O resultado é que você pode conceder aos usuários administrativos o acesso necessário para desempenhar suas responsabilidades de trabalho, e nada mais. Além disso, você pode proteger esses logins usando PKI de vários fornecedores, restringir o acesso apenas a chaves ssh e impor bloqueios de tentativa de login com falha.

Para obter mais informações sobre controle de acesso administrativo, clique em ["aqui."](#)

Autenticação Multifactor

A ONTAP e alguns outros produtos da NetApp agora oferecem suporte à autenticação multifator (MFA) usando diversos métodos. O resultado é que um nome de usuário/senha comprometidos por si só não é uma thread de segurança sem os dados do segundo fator, como um FOB ou um aplicativo de smartphone.

Para obter mais informações, clique em ["aqui."](#)

API RBAC

A automação requer chamadas de API, mas nem todas as ferramentas exigem acesso administrativo completo. Para ajudar a proteger os sistemas de automação, o RBAC também está disponível no nível da API. Você pode limitar as contas de usuário de automação às chamadas de API necessárias. Por exemplo, o software de monitoramento não precisa alterar o acesso, ele só requer acesso de leitura. Os fluxos de trabalho que provisionam o storage não precisam da capacidade de excluir o storage.

Para saber mais, inicie https://docs.NetApp.com/US-en/ONTAP-Automation/REST/rbac_overview.html[here.]

Verificação multi-admin (MAV)

A autenticação multi "fator" pode ser levada ainda mais longe, exigindo que dois administradores diferentes, cada um com suas próprias credenciais, aprovem determinadas atividades. Isso inclui alterar permissões de login, executar comandos de diagnóstico e excluir dados.

Para obter mais informações sobre a verificação multi-admin (MAV), clique em ["aqui"](#)

MySQL

Visão geral

MySQL e suas variantes, incluindo MariaDB e Percona MySQL, é o banco de dados mais popular do mundo.



Esta documentação sobre o ONTAP e o banco de dados MySQL substitui o banco de dados *TR-4722: MySQL publicado anteriormente sobre as melhores práticas do ONTAP*.

ONTAP é uma plataforma ideal para banco de dados MySQL porque o ONTAP é literalmente projetado para bancos de dados. Vários recursos, como otimizações aleatórias de latência de e/S para qualidade avançada de serviço (QoS) e funcionalidade básica do FlexClone, foram criados especificamente para atender às necessidades dos workloads de banco de dados.

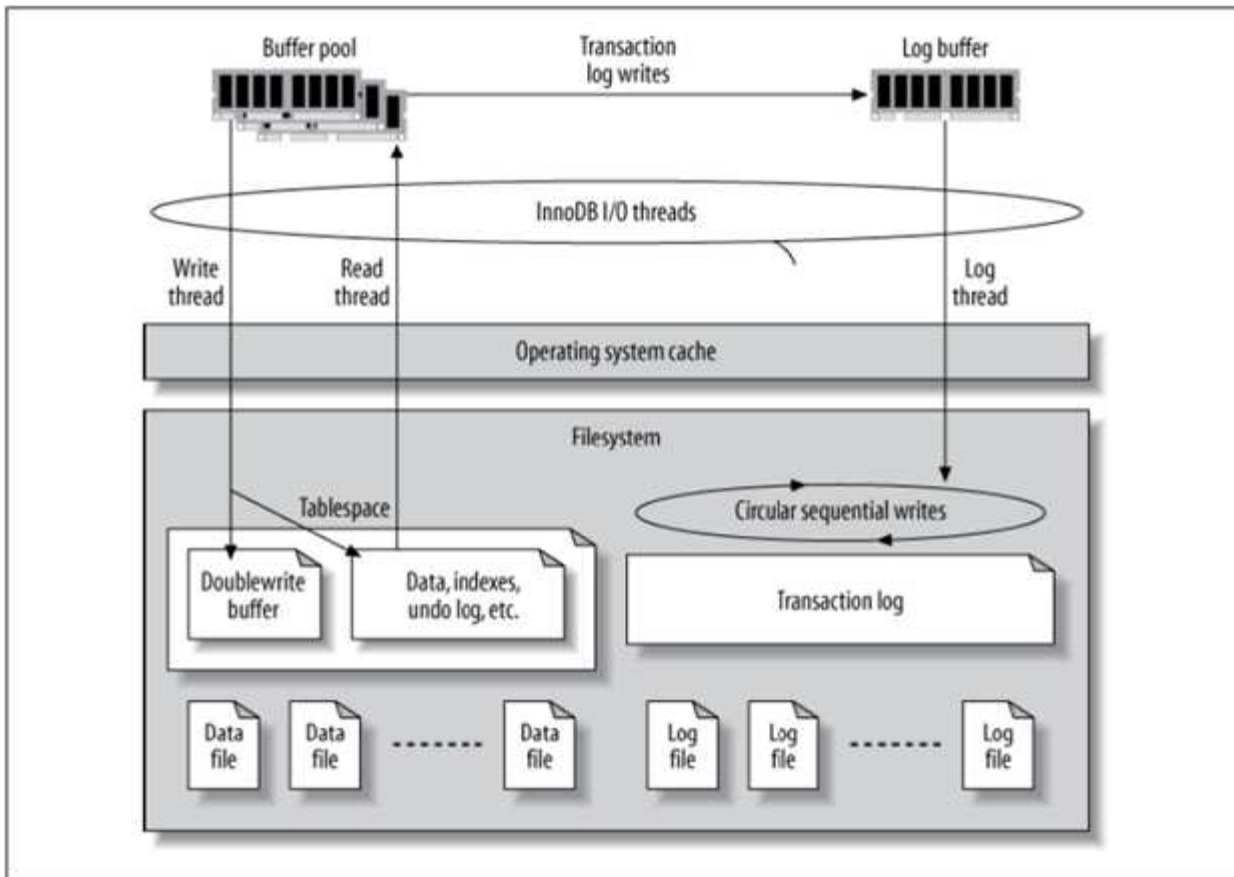
Recursos adicionais, como atualizações sem interrupções, (incluindo substituição de storage) garantem que seus bancos de dados essenciais permaneçam disponíveis. Você também pode ter recuperação instantânea de desastres para ambientes grandes por meio do MetroCluster ou selecionar bancos de dados usando o SnapMirror active Sync.

Mais importante ainda, o ONTAP oferece desempenho incomparável com a capacidade de dimensionar a solução para suas necessidades exclusivas. Nossos sistemas high-end podem fornecer mais de 1M IOPS com latências medidas em microssegundos. No entanto, se você só precisa de 100K IOPS, é possível dimensionar corretamente sua solução de storage com uma controladora menor que ainda executa o mesmo sistema operacional de storage.

Configuração do banco de dados

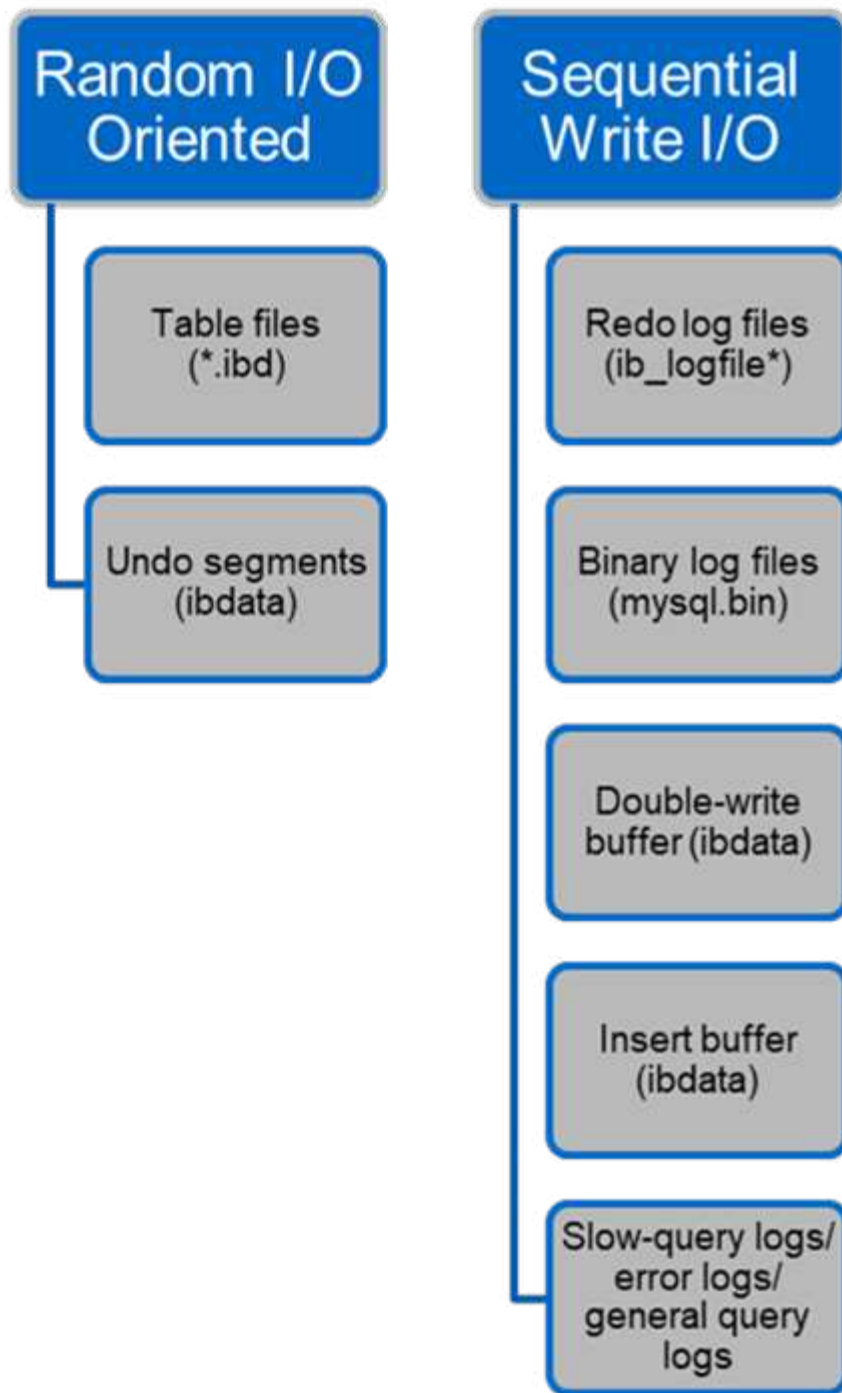
Estrutura de ficheiros

InnoDB atua como a camada intermediária entre o armazenamento e o servidor MySQL, ele armazena os dados nas unidades.



MySQL I/o é categorizado em dois tipos:

- E/S de arquivo aleatório
- E/S de arquivo sequencial



Os arquivos de dados são lidos e substituídos aleatoriamente, o que resulta em IOPS alto. Portanto, o armazenamento SSD é recomendado.

Os arquivos de log de refazer e os arquivos de log binários são logs transacionais. Eles são escritos sequencialmente, para que você possa obter um bom desempenho em HDD com o cache de gravação. Uma leitura sequencial acontece na recuperação, mas raramente causa um problema de desempenho, porque o tamanho do arquivo de log geralmente é menor do que os arquivos de dados, e as leituras sequenciais são mais rápidas do que as leituras aleatórias (que ocorrem em arquivos de dados).

O buffer de gravação dupla é um recurso especial do InnoDB. O InnoDB primeiro grava páginas apagadas no buffer de gravação dupla e, em seguida, grava as páginas em suas posições corretas nos arquivos de dados.

Este processo impede a corrupção de página. Sem o buffer de gravação dupla, a página pode ficar corrompida se ocorrer uma falha de energia durante o processo de gravação em unidades. Como a gravação no buffer de gravação dupla é sequencial, ele é altamente otimizado para HDDs. Leituras sequenciais ocorrem na recuperação.

Como o ONTAP NVRAM já fornece proteção contra gravação, o buffer de gravação dupla não é necessário. O MySQL tem um parâmetro, `skip_innodb_doublewrite`, para desativar o buffer de gravação dupla. Esse recurso pode melhorar substancialmente o desempenho.

O buffer de inserção também é um recurso especial do InnoDB. Se blocos de índice secundários não exclusivos não estiverem na memória, o InnoDB insere entradas no buffer de inserção para evitar operações de e/S aleatórias. Periodicamente, o buffer de inserção é mesclado nas árvores de índice secundárias no banco de dados. O buffer de inserção reduz o número de operações de e/S mesclando solicitações de e/S para o mesmo bloco; operações de e/S aleatórias podem ser sequenciais. O buffer de inserção também é altamente otimizado para HDDs. As gravações e leituras sequenciais ocorrem durante operações normais.

Os segmentos de desfazer são orientados por e/S aleatórias. Para garantir a simultaneidade de múltiplas versões (MVCC), o InnoDB deve Registrar imagens antigas nos segmentos de desfazer. A leitura de imagens anteriores dos segmentos de desfazer requer leituras aleatórias. Se você executar uma transação longa com leituras repetíveis (como `mysqldump` - transação única) ou executar uma consulta longa, leituras aleatórias podem ocorrer. Portanto, armazenar segmentos de desfazer em SSDs é melhor nessa instância. Se você executar apenas transações curtas ou consultas, as leituras aleatórias não são um problema.

A NetApp recomenda o seguinte layout de design de armazenamento devido às características de e/S do InnoDB.



- Um volume para armazenar arquivos aleatórios e sequenciais orientados por e/S do MySQL
- Outro volume para armazenar arquivos puramente sequenciais orientados por e/S do MySQL

Esse layout também ajuda a projetar políticas e estratégias de proteção de dados.

Parâmetros de configuração

O NetApp recomenda alguns parâmetros de configuração importantes do MySQL para obter um desempenho ideal.

Parâmetros	Valores
<code>innodb_log_file_size</code>	256M
<code>innodb_flush_log_at_trx_commit</code>	2
<code>innodb_doublewrite</code>	0
<code>innodb_flush_method</code>	<code>fsync</code>
<code>innodb_buffer_pool_size</code>	11G
<code>innodb_io_capacity</code>	8192
<code>innodb_buffer_pool_instances</code>	8
<code>innodb_lru_scan_depth</code>	8192
<code>open_file_limit</code>	65535

Para definir os parâmetros descritos nesta seção, você deve alterá-los no arquivo de configuração do MySQL (my.cnf). As melhores práticas da NetApp são resultado de testes realizados internamente.

innodb_log_file_size

Selecionar o tamanho certo para o tamanho do arquivo de log InnoDB é importante para as operações de gravação e para ter um tempo de recuperação decente após uma falha do servidor.

Como muitas transações são registradas no arquivo, o tamanho do arquivo de log é importante para operações de gravação. Quando os Registros são modificados, a alteração não é imediatamente gravada de volta para o espaço da tabela. Em vez disso, a alteração é gravada no final do arquivo de log e a página é marcada como suja. InnoDB usa seu log para converter a e/S aleatória em e/S sequencial

Quando o log está cheio, a página suja é escrita para a tablespace em sequência para liberar espaço no arquivo de log. Por exemplo, suponha que um servidor trava no meio de uma transação, e as operações de gravação são registradas apenas no arquivo de log. Antes que o servidor possa voltar ao vivo, ele deve passar por uma fase de recuperação na qual as alterações registradas no arquivo de log são reproduzidas. Quanto mais entradas estiverem no arquivo de log, mais tempo leva para que o servidor se recupere.

Neste exemplo, o tamanho do arquivo de log afeta tanto o tempo de recuperação quanto o desempenho de gravação. Ao escolher o número certo para o tamanho do arquivo de log, equilibre o tempo de recuperação em relação ao desempenho de gravação. Normalmente, qualquer coisa entre 128M e 512M é um bom valor.

innodb_flush_log_at_trx_commit

Quando há uma alteração nos dados, a alteração não é imediatamente gravada no armazenamento.

Em vez disso, os dados são gravados em um buffer de log, que é uma parte da memória que o InnoDB atribui a alterações de buffer que são registradas no arquivo de log. O InnoDB limpa o buffer para o arquivo de log quando uma transação é confirmada, quando o buffer fica cheio, ou uma vez por segundo, qualquer evento acontece primeiro. A variável de configuração que controla este processo é `innodb_flush_log_at_trx_commit`. As opções de valor incluem:

- Quando você define `innodb_flush_log_trx_at_commit=0`, o InnoDB grava os dados modificados (no pool de buffers do InnoDB) no arquivo de log (`ib_logfile`) e limpa o arquivo de log (gravação no armazenamento) a cada segundo. No entanto, não faz nada quando a transação é cometida. Se houver uma falha de energia ou falha do sistema, nenhum dos dados não lavados é recuperável porque não é gravado no arquivo de log ou nas unidades.
- Quando você define `innodb_flush_log_trx_commit=1`, o InnoDB grava o buffer de log no log de transações e passa para o armazenamento durável para cada transação. Por exemplo, para todos os commits de transação, o InnoDB grava no log e depois grava no armazenamento. O armazenamento mais lento afeta negativamente o desempenho; por exemplo, o número de transações InnoDB por segundo é reduzido.
- Quando você define `innodb_flush_log_trx_commit=2`, o InnoDB grava o buffer de log no arquivo de log em cada commit; no entanto, ele não grava dados no armazenamento. O InnoDB limpa os dados uma vez a cada segundo. Mesmo que haja uma falha de energia ou falha do sistema, os dados da opção 2 estão disponíveis no arquivo de log e são recuperáveis.

Se o desempenho for o objetivo principal, defina o valor como 2. Como o InnoDB grava nas unidades uma vez por segundo, não para cada commit de transação, o desempenho melhora drasticamente. Se ocorrer uma

falha de energia ou falha, os dados podem ser recuperados do log de transações.

Se a segurança de dados for o objetivo principal, defina o valor como 1 para que, para cada commit de transação, o InnoDB seja enviado para as unidades. No entanto, o desempenho pode ser afetado.



NetApp recomenda defina o valor `innodb_flush_log_trx_commit` como 2 para um melhor desempenho.

`innodb_doublewrite`

```
`innodb_doublewrite`Quando está ativado (o padrão), o InnoDB armazena todos os dados duas vezes: Primeiro para o buffer de gravação dupla e depois para os arquivos de dados reais.
```

Você pode desativar esse parâmetro com `--skip-innodb_doublewrite` para benchmarks ou quando estiver mais preocupado com o desempenho superior do que com a integridade de dados ou possíveis falhas. InnoDB usa uma técnica de flush de arquivo chamada double-write. Antes de gravar páginas nos arquivos de dados, o InnoDB as grava em uma área contígua chamada buffer de gravação dupla. Depois que a gravação e o flush para o buffer de gravação dupla estiverem completos, o InnoDB grava as páginas em suas posições apropriadas no arquivo de dados. Se o sistema operacional ou um processo `mysqld` falhar durante uma gravação de página, o InnoDB poderá encontrar uma boa cópia da página do buffer de gravação dupla durante a recuperação de falhas.



A NetApp recomenda a desativação do buffer de gravação dupla. O ONTAP NVRAM serve a mesma função. O buffer duplo danificará desnecessariamente o desempenho.

`innodb_buffer_pool_size`

O pool de buffers InnoDB é a parte mais importante de qualquer atividade de ajuste.

O InnoDB depende muito do pool de buffers para armazenar em cache índices e remendar os dados, o índice de hash adaptativo, o buffer de inserção e muitas outras estruturas de dados usadas internamente. O pool de buffers também armazena alterações nos dados para que as operações de gravação não precisem ser executadas imediatamente no armazenamento, melhorando assim o desempenho. O buffer pool é parte integrante do InnoDB e seu tamanho deve ser ajustado em conformidade. Considere os seguintes fatores ao definir o tamanho do pool de buffers:

- Para uma máquina exclusiva InnoDB dedicada, defina o tamanho do pool de buffers para 80% ou mais da RAM disponível.
- Se não for um servidor dedicado MySQL, defina o tamanho para 50% da RAM.

`innodb_flush_method`

O parâmetro `innodb_flush_method` especifica como o InnoDB abre e limpa os arquivos de log e dados.

Otimizações

Na otimização InnoDB, a configuração desse parâmetro ajusta o desempenho do banco de dados quando aplicável.

As seguintes opções são para limpar os arquivos através do InnoDB:

- `fsync`. O InnoDB usa a `fsync()` chamada do sistema para liberar os arquivos de dados e log. Esta opção é a predefinição.
- `O_DSYNC`. InnoDB usa a `O_DSYNC` opção para abrir e lavar os arquivos de log e `fsync()` para lavar os arquivos de dados. InnoDB não usa `O_DSYNC` diretamente, porque houve problemas com ele em muitas variedades de UNIX.
- `O_DIRECT`. O InnoDB usa a `O_DIRECT` opção (ou `directio()` no Solaris) para abrir os arquivos de dados e usa `fsync()` para limpar os arquivos de dados e log. Esta opção está disponível em algumas versões GNU/Linux, FreeBSD e Solaris.
- `O_DIRECT_NO_FSYNC`. O InnoDB usa a `O_DIRECT` opção durante a lavagem de e/S; no entanto, ele ignora a `fsync()` chamada do sistema depois. Essa opção não é adequada para alguns tipos de sistemas de arquivos (por exemplo, XFS). Se você não tiver certeza se o sistema de arquivos requer uma `fsync()` chamada de sistema, por exemplo, para preservar todos os metadados do arquivo, use a `O_DIRECT` opção.

Observação

Nos testes de laboratório do NetApp, a `fsync` opção padrão foi usada em NFS e SAN, e foi um ótimo improvisador de desempenho em comparação com `O_DIRECT`. Ao usar o método `flush` como `O_DIRECT` no ONTAP, observamos que o cliente escreve muitas gravações de byte único na borda do bloco 4096 em modo serial. Essas gravações aumentam a latência na rede e degradam o desempenho.

innodb_io_capacity

No plug-in InnoDB, um novo parâmetro chamado `innodb_io_Capacity` foi adicionado a partir do MySQL 5,7.

Ele controla o número máximo de IOPS que o InnoDB executa (que inclui a taxa de limpeza de páginas sujas, bem como o tamanho do lote do buffer de inserção [`ibuf`]). O parâmetro `innodb_io_Capacity` define um limite superior em IOPS por tarefas de fundo do InnoDB, como limpar páginas do pool de buffers e mesclar dados do buffer de mudança.

Defina o parâmetro `innodb_io_Capacity` para o número aproximado de operações de e/S que o sistema pode executar por segundo. Idealmente, mantenha a configuração o mais baixa possível, mas não tão baixa que as atividades de fundo desacelerem. Se a configuração for muito alta, os dados serão removidos do pool de buffers e inserirão o buffer muito rapidamente para que o armazenamento em cache forneça um benefício significativo.



A NetApp recomenda que, se estiver usando essa configuração em NFS, analise o resultado do teste de IOPS (SysBench/fio) e defina o parâmetro de acordo. Use o menor valor possível para flushing e purging para manter-se a menos que você veja mais páginas modificadas ou sujas do que você deseja no pool de buffers InnoDB.



Não use valores extremos como 20.000 ou mais, a menos que você tenha provado que valores mais baixos não são suficientes para sua carga de trabalho.

O parâmetro `InnoDB_io_Capacity` regula as taxas de lavagem e e/S relacionadas



Você pode prejudicar seriamente o desempenho definindo este parâmetro ou o parâmetro `innodb_io_Capacity_Max` muito alto e desperdiçando operações de e/S com lavagem prematura.

`innodb_lru_scan_depth`

O `innodb_lru_scan_depth` parâmetro influencia os algoritmos e heurísticas da operação de flush para o pool de buffers InnoDB.

Esse parâmetro é do interesse principal dos especialistas em performance que ajustam workloads com uso intenso de e/S. Para cada instância de pool de buffer, este parâmetro especifica o quão abaixo na lista de páginas menos usadas recentemente (LRU) o thread de limpeza de páginas deve continuar a digitalizar, procurando páginas sujas para serem flush. Esta operação em segundo plano é executada uma vez por segundo.

Você pode ajustar o valor para cima ou para baixo para minimizar o número de páginas livres. Não defina o valor muito maior do que o necessário, porque as digitalizações podem ter um custo de desempenho significativo. Além disso, considere ajustar esse parâmetro ao alterar o número de instâncias do pool de buffers, `innodb_lru_scan_depth * innodb_buffer_pool_instances` porque define a quantidade de trabalho executada pelo thread de limpeza de páginas a cada segundo.

Uma configuração menor do que o padrão é adequada para a maioria das cargas de trabalho. Considere aumentar o valor somente se você tiver capacidade de e/S sobressalente em uma carga de trabalho típica. Por outro lado, se uma carga de trabalho com uso intenso de gravação saturar a capacidade de e/S, diminua o valor, especialmente se você tiver um grande pool de buffers.

`open_file_limits`

O `open_file_limits` parâmetro determina o número de arquivos que o sistema operacional permite que o mysqld abra.

O valor deste parâmetro em tempo de execução é o valor real permitido pelo sistema e pode ser diferente do valor especificado na inicialização do servidor. O valor é 0 em sistemas onde o MySQL não pode alterar o número de arquivos abertos. O valor efetivo `open_files_limit` é baseado no valor especificado na inicialização do sistema (se houver) e nos valores de `max_connections` e `table_open_cache` usando estas fórmulas:

- $10 \times \text{max_connections} \times (\text{table_open_cache} \times 2)$
- $\text{max_connections} \times 5$
- Limite do sistema operacional se positivo
- Se o limite do sistema operacional for infinito: `open_files_limit` O valor é especificado na inicialização; 5.000 se nenhum

O servidor tenta obter o número de descritores de arquivo usando o máximo desses quatro valores. Se muitos descritores não puderem ser obtidos, o servidor tentará obter o máximo que o sistema permitir.

Configuração de host

Containerização

Containerização de bancos de dados MySQL está se tornando mais prevalente.

O gerenciamento de contentores de baixo nível é quase sempre realizado através do Docker. As plataformas de gerenciamento de contêineres, como OpenShift e Kubernetes, simplificam ainda mais o gerenciamento de ambientes de contêineres grandes. Os benefícios da Containerização incluem custos mais baixos, porque não há necessidade de licenciar um hypervisor. Além disso, os containers permitem que vários bancos de dados sejam executados isolados uns dos outros enquanto compartilham o mesmo kernel subjacente e sistema operacional. Os contêineres podem ser provisionados em microssegundos.

A NetApp oferece o Astra Trident para fornecer funcionalidades avançadas de gerenciamento de storage. Por exemplo, o Astra Trident permite que um contêiner criado no Kubernetes provisione automaticamente seu storage na camada apropriada, aplique políticas de exportação, defina políticas de snapshot e até clone um contêiner para outro. Para obter informações adicionais, consulte o "[Documentação do Astra Trident](#)".

NFSv3 mesas de slot

O desempenho do NFSv3 no Linux depende de um parâmetro `tcp_max_slot_table_entries` chamado .

As tabelas de slot TCP são equivalentes a NFSv3 mm de profundidade de fila do adaptador de barramento do host (HBA). Essas tabelas controlam o número de operações NFS que podem ficar pendentes de uma só vez. O valor padrão é geralmente 16, o que é muito baixo para um desempenho ideal. O problema oposto ocorre em kernels Linux mais recentes, que podem aumentar automaticamente o limite da tabela de slots TCP para um nível que satura o servidor NFS com solicitações.

Para um desempenho ideal e para evitar problemas de desempenho, ajuste os parâmetros do kernel que controlam as tabelas de slots TCP.

Executar o `sysctl -a | grep tcp.*.slot_table` comando e respeitar os seguintes parâmetros:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Todos os sistemas Linux devem incluir `sunrpc.tcp_slot_table_entries`, mas apenas alguns incluem `sunrpc.tcp_max_slot_table_entries`. Ambos devem ser definidos para 128.

Cuidado

A falha em definir esses parâmetros pode ter efeitos significativos no desempenho. Em alguns casos, o desempenho é limitado porque o sistema operacional linux não está emitindo e/S suficiente. Em outros casos, as latências de e/S aumentam à medida que o sistema operacional linux tenta emitir mais e/S do que pode ser reparado.

Agendadores de e/S.

O kernel Linux permite um controle de baixo nível sobre a maneira como e/S para bloquear dispositivos é agendada.

Os padrões em várias distribuições do Linux variam consideravelmente. O MySQL recomenda que você use NOOP ou um `deadline` agendador de e/S com e/S assíncrona nativa (AIO) no Linux. Em geral, os clientes da NetApp e os testes internos mostram melhores resultados com NoOps.

O mecanismo de armazenamento InnoDB do MySQL usa o subsistema de e/S assíncrono (AIO nativo) no Linux para executar solicitações de leitura e gravação para páginas de arquivos de dados. Esse comportamento é controlado pela `innodb_use_native_aio` opção de configuração, que é ativada por padrão. Com AIO nativo, o tipo de agendador de e/S tem maior influência no desempenho de e/S. Realize benchmarks para determinar qual agendador de e/S fornece os melhores resultados para sua carga de trabalho e ambiente.

Consulte a documentação relevante do Linux e MySQL para obter instruções sobre como configurar o agendador de e/S.

Descritores de arquivo

Para executar, o servidor MySQL precisa de descritores de arquivo e os valores padrão não são suficientes.

Ele os usa para abrir novas conexões, armazenar tabelas no cache, criar tabelas temporárias para resolver consultas complicadas e acessar as persistentes. Se o `mysqld` não for capaz de abrir novos arquivos quando necessário, ele pode parar de funcionar corretamente. Um sintoma comum deste problema é o erro 24, "muitos arquivos abertos". O número de descritores de arquivo que o `mysqld` pode abrir simultaneamente é definido pela `open_files_limit` opção definida no arquivo de configuração (`/etc/my.cnf`). Mas `open_files_limit` também depende dos limites do sistema operacional. Essa dependência torna a configuração da variável mais complicada.

O MySQL não pode definir sua `open_files_limit` opção mais alta do que a especificada em `ulimit 'open files'`. Portanto, você precisa definir explicitamente esses limites no nível do sistema operacional para permitir que o MySQL abra arquivos conforme necessário. Existem duas maneiras de verificar o limite de arquivos no Linux:

- O `ulimit` comando fornece rapidamente uma descrição detalhada dos parâmetros que estão sendo permitidos ou bloqueados. As alterações feitas ao executar este comando não são permanentes e serão apagadas após uma reinicialização do sistema.
- As alterações ao `/etc/security/limit.conf` arquivo são permanentes e não são afetadas pela reinicialização do sistema.

Certifique-se de alterar os limites físicos e macios para o usuário `mysql`. Os seguintes trechos são da configuração:

```
mysql hard nfile 65535
mysql soft nfile 65353
```

Em paralelo, atualize a mesma configuração no `my.cnf` para usar totalmente os limites de arquivos abertos.

Configuração de armazenamento

NFS

A documentação do MySQL recomenda que você use o NFSv4 para implantações nas.

Tamanhos de transferência de NFS do ONTAP

Por padrão, o ONTAP limitará os tamanhos de e/S NFS a 64K. Random IO com um banco de dados MySQL usa um tamanho de bloco muito menor que está bem abaixo do máximo de 64KMB. O IO de bloco grande geralmente é paralelizado, portanto o máximo de 64K também não é uma limitação.

Existem algumas cargas de trabalho em que o máximo 64K cria uma limitação. Em particular, operações de um único processo, como operações de backup de verificação de tabela completa, serão executadas de forma mais rápida e eficiente se o banco de dados puder executar menos, mas maiores. O tamanho ideal de manuseio de e/S para ONTAP com cargas de trabalho de banco de dados é 256K. As opções de montagem NFS listadas para sistemas operacionais específicos abaixo foram atualizadas de 64K para 256K em conformidade.

O tamanho máximo de transferência para um determinado SVM do ONTAP pode ser alterado da seguinte forma:

```
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
```



Nunca diminua o tamanho máximo de transferência permitido no ONTAP abaixo do valor de rsize/wsize dos sistemas de arquivos NFS atualmente instalados. Isso pode criar pendências ou até mesmo corrupção de dados com alguns sistemas operacionais. Por exemplo, se os clientes NFS estiverem atualmente definidos em um rsize/wsize de 65536, o tamanho máximo de transferência do ONTAP poderá ser ajustado entre 65536 e 1048576 sem efeito, porque os próprios clientes são limitados. Reduzir o tamanho máximo de transferência abaixo de 65536 pode danificar a disponibilidade ou os dados.

A NetApp recomenda



Configurando a seguinte configuração NFSv4 fstab (/etc/fstab):

```
nfs4 rw,  
hard,nointr,bg,vers=4,proto=tcp,noatime,rsize=262144,wsizer=262144
```



Um problema comum com o NFSv3 foi os arquivos de log do InnoDB bloqueados após uma falha de energia. Usar o tempo ou trocar arquivos de log resolveu esse problema. No entanto, o NFSv4 tem operações de bloqueio e mantém o controle de arquivos abertos e delegações.

SAN

Existem duas opções para configurar o MySQL com SAN usando o modelo de dois volumes usual.

Bancos de dados menores podem ser colocados em um par de LUNs padrão, desde que as demandas de e/S e capacidade estejam dentro dos limites de um único sistema de arquivos LUN. Por exemplo, um banco de dados que requer aproximadamente 2K IOPS aleatório pode ser hospedado em um único sistema de arquivos em um único LUN. Da mesma forma, um banco de dados com apenas 100GB MB de tamanho caberia em um único LUN sem criar um problema de gerenciamento.

Bancos de dados maiores exigem vários LUNs. Por exemplo, um banco de dados que requer 100K IOPS provavelmente precisará de pelo menos oito LUNs. Um único LUN se tornaria um gargalo devido ao número inadequado de canais SCSI para unidades. Um banco de dados 10TB seria igualmente difícil de gerenciar em um único LUN 10TB. Os gerenciadores lógicos de volume são projetados para unir os recursos de desempenho e capacidade de vários LUNs para melhorar o desempenho e a capacidade de gerenciamento.

Em ambos os casos, um par de volumes ONTAP deve ser suficiente. Com uma configuração simples, o LUN do arquivo de dados seria colocado em um volume dedicado, assim como o LUN de log. Com uma configuração lógica do gerenciador de volumes, todos os LUNs no grupo de volumes de arquivos de dados estariam em um volume dedicado e os LUNs do grupo de volumes de log estariam em um segundo volume dedicado.

A NetApp recomenda usando dois sistemas de arquivos para implantações MySQL na SAN:

- O primeiro sistema de arquivos armazena todos os dados MySQL, incluindo tablespace, dados e índice.
- O segundo sistema de arquivos armazena todos os logs (logs binários, logs lentos e logs de transações).

Existem várias razões para separar dados dessa maneira, incluindo:



- Os padrões de e/S de arquivos de dados e arquivos de log diferem. Separá-los permitiria mais opções com controles de QoS.
- O uso ideal da tecnologia Snapshot requer a capacidade de restaurar os arquivos de dados de forma independente. Commingling arquivos de dados com arquivos de log interfere com a restauração de arquivos de dados.
- A tecnologia NetApp SnapMirror pode ser usada para fornecer uma funcionalidade de recuperação de desastres simples e de baixo RPO para um banco de dados. No entanto, ela requer diferentes programações de replicação para arquivos e logs de dados.



Use esse layout básico de dois volumes para preparar a solução para o futuro, de modo que todos os recursos do ONTAP possam ser usados, se necessário.

A NetApp recomenda a formatação da sua unidade com o sistema de arquivos ext4 devido aos seguintes recursos:



- Abordagem estendida aos recursos de gerenciamento de blocos usados no sistema de arquivos de journaling (JFS) e recursos de alocação atrasada do sistema de arquivos estendido (XFS).
- EXT4 permite sistemas de arquivos de até 1 exbibyte (2⁶⁰ bytes) e arquivos de até 16 tebibytes (16 * 2⁴⁰ bytes). Em contraste, o sistema de arquivos ext3 suporta apenas um tamanho máximo de sistema de arquivos de 16TB e um tamanho máximo de arquivo de 2TB.
- Em sistemas de arquivos ext4, a alocação de vários blocos (mballoc) aloca vários blocos para um arquivo em uma única operação, em vez de alocá-los um por um, como em ext3. Essa configuração reduz a sobrecarga de chamar o alocador de bloco várias vezes e otimiza a alocação de memória.
- Embora o XFS seja o padrão para muitas distribuições Linux, ele gerencia metadados de forma diferente e não é adequado para algumas configurações do MySQL.



A NetApp recomenda usar opções de tamanho de bloco 4K com o utilitário mkfs para alinhar com o tamanho de LUN de bloco existente.

```
mkfs.ext4 -b 4096
```

Os LUNs NetApp armazenam dados em 4KB blocos físicos, o que rende oito blocos lógicos de 512 bytes.

Se você não configurar o mesmo tamanho de bloco, a e/S não será alinhada com os blocos físicos corretamente e poderá gravar em duas unidades diferentes em um grupo RAID, resultando em latência.



É importante alinhar a e/S para operações de leitura/gravação suaves. No entanto, quando a e/S começa em um bloco lógico que não está no início de um bloco físico, a e/S está desalinhada. As operações de e/S são alinhadas somente quando começam em um bloco lógico, o primeiro bloco lógico em um bloco físico.

Banco de dados Oracle

Bancos de dados Oracle no ONTAP

O ONTAP foi projetado para bancos de dados Oracle. Por décadas, o ONTAP foi otimizado para as demandas exclusivas de e/S de banco de dados relacional e vários recursos do ONTAP foram criados especificamente para atender às necessidades dos bancos de dados Oracle e até mesmo a pedido da própria Oracle Inc..



Esta documentação substitui esses relatórios técnicos publicados anteriormente *TR-3633: Bancos de dados Oracle no ONTAP*; *TR-4591: Proteção de dados Oracle: Backup, recuperação, replicação*; *TR-4592: Oracle no MetroCluster*; e *TR-4534: Migração de bancos de dados Oracle para sistemas de armazenamento NetApp*

Além das muitas maneiras possíveis pelas quais o ONTAP agrega valor ao seu ambiente de banco de dados, há também uma grande variedade de requisitos do usuário, incluindo tamanho do banco de dados, requisitos de desempenho e necessidades de proteção de dados. As implantações conhecidas do storage NetApp incluem tudo, desde um ambiente virtualizado de aproximadamente 6.000 bancos de dados executados no VMware ESX até um data warehouse de instância única atualmente dimensionado a 996TB TB e em crescimento. Como resultado, existem algumas práticas recomendadas claras para configurar um banco de dados Oracle no storage NetApp.

Os requisitos para operar um banco de dados Oracle no storage NetApp são abordados de duas maneiras. Em primeiro lugar, quando existir uma boa prática clara, ela será chamada especificamente. Em um alto nível, muitas considerações de design que devem ser abordadas pelos arquitetos de soluções de armazenamento Oracle com base em seus requisitos de negócios específicos serão explicadas.

Configuração ONTAP

RAID

RAID refere-se ao uso de redundância para proteger os dados contra a perda de uma unidade.

Ocasionalmente, surgem perguntas sobre os níveis de RAID na configuração do armazenamento NetApp usado para bancos de dados Oracle e outros aplicativos empresariais. Muitas práticas recomendadas legadas da Oracle em relação à configuração do storage array contêm avisos sobre o uso do espelhamento RAID e/ou a prevenção de certos tipos de RAID. Embora apresentem pontos válidos, essas fontes não se aplicam ao RAID 4 e às tecnologias NetApp RAID DP e RAID-teC usadas no ONTAP.

RAID 4, RAID 5, RAID 6, RAID DP e RAID-teC usam paridade para garantir que a falha da unidade não resulte na perda de dados. Essas opções de RAID oferecem uma utilização do storage muito melhor em comparação com o espelhamento, mas a maioria das implementações de RAID tem uma desvantagem que afeta as operações de gravação. A conclusão de uma operação de gravação em outras implementações de RAID pode exigir várias leituras de unidade para regenerar os dados de paridade, um processo comumente chamado de penalidade de RAID.

O ONTAP, no entanto, não incorre nessa penalidade de RAID. Isso ocorre devido à integração do NetApp WAFL (Write Anywhere File Layout) com a camada RAID. As operações de gravação são Unidas na RAM e preparadas como um stripe RAID completo, incluindo geração de paridade. O ONTAP não precisa executar

uma leitura para concluir uma gravação, o que significa que o ONTAP e o WAFL evitam a penalidade de RAID. A performance de operações críticas à latência, como refazer o log, não é impedida, e as gravações aleatórias de arquivos de dados não incorrem em nenhuma penalidade de RAID resultante da necessidade de gerar novamente a paridade.

Com relação à confiabilidade estatística, até mesmo o RAID DP oferece melhor proteção do que o espelhamento RAID. O principal problema é a demanda feita nas unidades durante uma reconstrução RAID. Com um conjunto RAID espelhado, o risco de perda de dados de uma unidade falhar durante a reconstrução para seu parceiro no conjunto RAID é muito maior do que o risco de uma falha de unidade tripla em um conjunto RAID DP.

Gerenciamento de capacidade

Gerenciar um banco de dados ou outra aplicação empresarial com storage empresarial previsível, gerenciável e de alta performance exige espaço livre nas unidades para gerenciamento de dados e metadados. A quantidade de espaço livre necessária depende do tipo de unidade usada e dos processos de negócios.

O espaço livre é definido como qualquer espaço que não é usado para dados reais e inclui espaço não alocado no próprio agregado e espaço não utilizado dentro dos volumes constituintes. O thin Provisioning também deve ser considerado. Por exemplo, um volume pode conter um LUN de 1TB, dos quais apenas 50% são utilizados por dados reais. Em um ambiente thin Provisioning, isso parece estar consumindo 500GB de espaço corretamente. No entanto, em um ambiente totalmente provisionado, a capacidade total de 1TB parece estar em uso. O 500GB de espaço não alocado está oculto. Este espaço não é utilizado pelos dados reais e, portanto, deve ser incluído no cálculo do espaço livre total.

As recomendações da NetApp para sistemas de storage usados para aplicações empresariais são as seguintes:

Agregados SSD, incluindo sistemas AFF



A NetApp recomenda um mínimo de 10% de espaço livre. Isso inclui todo o espaço não utilizado, incluindo espaço livre dentro do agregado ou de um volume e qualquer espaço livre alocado devido ao uso de provisionamento completo, mas não é usado por dados reais. O espaço lógico não é importante, a questão é quanto espaço físico livre real está disponível para armazenamento de dados.

A recomendação de 10% de espaço livre é muito conservadora. Agregados SSD podem dar suporte a workloads em níveis ainda mais altos de utilização sem afetar a performance. No entanto, à medida que a utilização do agregado aumenta, o risco de ficar sem espaço também aumenta se a utilização não for monitorada com cuidado. Além disso, ao executar um sistema com 99% de capacidade pode não incorrer em uma penalidade de desempenho, mas provavelmente incorreria em esforço de gerenciamento tentando impedi-lo de preencher completamente enquanto hardware adicional é solicitado, e pode levar algum tempo para adquirir e instalar unidades adicionais.

Agregados de HDD, incluindo agregados Flash Pool



A NetApp recomenda um mínimo de 15% de espaço livre quando as unidades giratórias são usadas. Isso inclui todo o espaço não utilizado, incluindo espaço livre dentro do agregado ou de um volume e qualquer espaço livre alocado devido ao uso de provisionamento completo, mas não é usado por dados reais. O desempenho será impactado à medida que o espaço livre se aproxima de 10%.

Máquinas virtuais de armazenamento

O gerenciamento do storage de banco de dados Oracle é centralizado em uma Storage Virtual Machine (SVM).

Um SVM, conhecido como vserver na CLI do ONTAP, é uma unidade funcional básica de storage e é útil comparar um SVM com um convidado em um servidor VMware ESX.

Quando instalado pela primeira vez, o ESX não tem recursos pré-configurados, como hospedar um sistema operacional convidado ou suportar um aplicativo de usuário final. É um recipiente vazio até que uma máquina virtual (VM) seja definida. ONTAP é semelhante. Quando o ONTAP é instalado pela primeira vez, ele não tem funcionalidades de fornecimento de dados até que um SVM seja criado. É a personalidade do SVM que define os serviços de dados.

Assim como em outros aspectos da arquitetura de storage, as melhores opções para SVM e design de interface lógica (LIF) dependem muito dos requisitos de dimensionamento e das necessidades empresariais.

SVMs

Não há prática recomendada oficial para provisionar SVMs para ONTAP. A abordagem certa depende dos requisitos de gerenciamento e segurança.

A maioria dos clientes opera um SVM principal para a maioria de seus requisitos diários, mas cria um pequeno número de SVMs para necessidades especiais. Por exemplo, você pode querer criar:

- SVM para um banco de dados de negócios essencial gerenciado por uma equipe de especialistas
- Um SVM para um grupo de desenvolvimento a quem controle administrativo completo foi dado para que eles possam gerenciar seu próprio storage de forma independente
- SVM para dados empresariais confidenciais, como recursos humanos ou dados de relatórios financeiros, em que a equipe administrativa deve ser limitada

Em um ambiente de alocação a vários clientes, os dados de cada locatário podem receber um SVM dedicado. O limite para o número de SVMs e LIFs por cluster, par de HA e nó depende do protocolo que está sendo usado, do modelo de nó e da versão do ONTAP. Consulte o ["NetApp Hardware Universe"](#) para estes limites.

Gerenciamento de desempenho com QoS ONTAP

O gerenciamento de vários bancos de dados Oracle com segurança e eficiência requer uma estratégia de QoS eficaz. O motivo são os recursos de performance cada vez maiores de um sistema de storage moderno.

Especificamente, o aumento da adoção do storage all-flash possibilitou a consolidação de workloads. Os storage arrays que dependem de Mídia giratória costumavam dar suporte a apenas um número limitado de workloads de e/S com uso intenso de e/S devido aos recursos limitados de IOPS da tecnologia de unidade de rotação mais antiga. Um ou dois bancos de dados altamente ativos saturariam as unidades subjacentes muito antes que as controladoras de storage atingissem seus limites. Isso mudou. A capacidade de desempenho de um número relativamente pequeno de unidades SSD pode saturar até mesmo os controladores de armazenamento mais poderosos. Isso significa que todos os recursos das controladoras podem ser aproveitados sem o medo de colapso repentino da performance na medida em que picos de latência de Mídia giratórios.

Como exemplo de referência, um simples sistema HA AFF A800 de dois nós é capaz de atender a até um milhão de IOPS aleatórios antes que a latência suba acima de um milissegundo. Espera-se que muito poucas

cargas de trabalho individuais alcancem tais níveis. A utilização total desse array de sistema AFF A800 envolverá o hospedar vários workloads e isso com segurança, garantindo a previsibilidade requer controles de QoS.

Há dois tipos de qualidade de serviço (QoS) no ONTAP: IOPS e largura de banda. Controles de QoS podem ser aplicados a SVMs, volumes, LUNs e arquivos.

QoS de IOPS

Um controle de QoS de IOPS é obviamente baseado no total de IOPS de um determinado recurso, mas há vários aspectos da QoS de IOPS que podem não ser intuitivos. Alguns clientes inicialmente ficaram intrigados com o aparente aumento da latência quando um limite de IOPS é atingido. O aumento da latência é o resultado natural da limitação do IOPS. Logicamente, ele funciona de forma semelhante a um sistema de token. Por exemplo, se um determinado volume contendo datafiles tiver um limite de 10K IOPS, cada I/O que chegar deve receber primeiro um token para continuar o processamento. Desde que não mais de 10K tokens tenham sido consumidos em um determinado segundo, nenhum atraso está presente. Se as operações de e/S precisarem esperar para receber o token, essa espera aparecerá como latência adicional. Quanto mais difícil uma carga de trabalho ultrapassar o limite de QoS, mais tempo cada IO deve esperar na fila para que sua vez seja processada, o que parece ao usuário como maior latência.



Tenha cuidado ao aplicar controles de QoS a dados de log de transação/refazer de banco de dados. Embora as demandas de desempenho do Registro de refazer sejam normalmente muito, muito mais baixas do que datafiles, a atividade do log de refazer é bursty. O IO acontece em breves pulsos, e um limite de QoS que parece apropriado para os níveis médios de e/S refazer pode ser muito baixo para os requisitos reais. O resultado pode ser limitações graves de desempenho, pois a QoS interage com cada sequência de log de refazer. Em geral, refazer e arquivar o log não deve ser limitado pela QoS.

QoS de largura de banda

Nem todos os tamanhos de e/S são iguais. Por exemplo, um banco de dados pode estar executando um grande número de leituras de blocos pequenos, o que resultaria no limite de IOPS ser atingido, mas os bancos de dados também podem estar executando uma operação de verificação de tabela completa que consistiria em um número muito pequeno de leituras de blocos grandes, consumindo uma quantidade muito grande de largura de banda, mas relativamente poucos IOPS.

Da mesma forma, um ambiente VMware pode gerar um número muito alto de IOPS aleatório durante a inicialização, mas executaria um IOPS menor, mas maior, durante um backup externo.

Às vezes, o gerenciamento eficaz da performance exige limites de QoS de IOPS ou largura de banda, ou até mesmo ambos.

QoS mínima/garantida

Muitos clientes procuram uma solução que inclua QoS garantida, que é mais difícil de alcançar do que parece e é potencialmente um desperdício. Por exemplo, a colocação de bancos de dados 10 com uma garantia de 10K IOPS exige o dimensionamento de um sistema para um cenário em que todos os bancos de dados 10 estejam sendo executados simultaneamente a 10K IOPS, totalizando 100KK.

A melhor utilização para controles mínimos de QoS é proteger workloads críticos. Por exemplo, considere um controlador ONTAP com um máximo de IOPS possível de 500K K e uma combinação de workloads de produção e desenvolvimento. Você deve aplicar políticas máximas de QoS a workloads de desenvolvimento para impedir que qualquer banco de dados monopolize o controlador. Em seguida, você aplicaria políticas mínimas de QoS a workloads de produção para garantir que eles sempre tenham o IOPS necessário

disponível quando necessário.

QoS adaptável

QoS adaptável refere-se ao recurso ONTAP em que o limite de QoS é baseado na capacidade do objeto de storage. Raramente é usado com bancos de dados porque geralmente não há nenhum link entre o tamanho de um banco de dados e seus requisitos de desempenho. Bancos de dados grandes podem ser quase inertes, enquanto bancos de dados menores podem ser os mais intensivos em IOPS.

A QoS adaptável pode ser muito útil com armazenamentos de dados de virtualização, pois os requisitos de IOPS desses conjuntos de dados tendem a se correlacionar com o tamanho total do banco de dados. Um datastore mais recente contendo 1TB TB de arquivos VMDK provavelmente precisará de cerca de metade do desempenho como um datastore 2TB. A QoS adaptável permite que você aumente os limites de QoS automaticamente à medida que o armazenamento de dados se torna preenchido com dados.

Eficiência

Os recursos de eficiência de espaço do ONTAP são otimizados para bancos de dados Oracle. Em quase todos os casos, a melhor abordagem é deixar os padrões em vigor com todos os recursos de eficiência habilitados.

Os recursos de eficiência de espaço, como compressão, compactação e deduplicação, foram projetados para aumentar a quantidade de dados lógicos compatíveis com uma determinada quantidade de storage físico. O resultado são custos mais baixos e sobrecarga de gerenciamento.

Em um alto nível, a compressão é um processo matemático pelo qual padrões nos dados são detetados e codificados de uma forma que reduz os requisitos de espaço. Em contraste, a deduplicação deteta blocos repetidos reais de dados e remove as cópias externas. A compactação permite que vários blocos lógicos de dados compartilhem o mesmo bloco físico na Mídia.



Veja as seções abaixo sobre provisionamento de thin para uma explicação da interação entre eficiência de armazenamento e reserva fracionária.

Compactação

Antes da disponibilidade de sistemas de storage all-flash, a compactação baseada em array era de valor limitado porque a maioria dos workloads com uso intenso de e/S exigia um número muito grande de fusos para fornecer desempenho aceitável. Os sistemas de armazenamento invariavelmente continham muito mais capacidade do que o necessário como um efeito colateral do grande número de unidades. A situação mudou com o aumento do armazenamento de estado sólido. Não há mais a necessidade de superprovisionamento vastamente de unidades puramente para obter um bom desempenho. O espaço da unidade em um sistema de armazenamento pode ser correspondente às necessidades reais de capacidade.

A maior funcionalidade de IOPS das solid-State drives (SSDs) quase sempre produz economias de custo em comparação com as unidades giratórias, mas a compactação pode gerar mais economias ao aumentar a capacidade efetiva da Mídia de estado sólido.

Existem várias maneiras de compactar dados. Muitos bancos de dados incluem suas próprias capacidades de compressão, mas isso raramente é observado em ambientes de clientes. A razão é geralmente a penalidade de desempenho para uma **mudança** para dados compactados, além disso, com alguns aplicativos, há altos custos de licenciamento para compactação no nível do banco de dados. Finalmente, há as consequências gerais de desempenho para as operações do banco de dados. Faz pouco sentido pagar um alto custo de licença por CPU para uma CPU que executa compactação de dados e descompressão em vez de trabalho

real de banco de dados. Uma opção melhor é descarregar o trabalho de compressão para o sistema de storage.

Compressão adaptável

A compactação adaptável foi totalmente testada com workloads empresariais sem efeito observado no desempenho, mesmo em um ambiente all-flash em que a latência é medida em microssegundos. Alguns clientes até relataram um aumento de desempenho com o uso de compactação porque os dados permanecem compactados no cache, aumentando efetivamente a quantidade de cache disponível em um controlador.

O ONTAP gerencia blocos físicos em 4KB unidades. A compactação adaptável usa um tamanho de bloco de compressão padrão de 8KB, o que significa que os dados são compactados em unidades 8KB. Isso corresponde ao tamanho do bloco 8KB mais frequentemente usado por bancos de dados relacionais. Os algoritmos de compressão tornam-se mais eficientes à medida que mais dados são compactados como uma única unidade. Um tamanho de bloco de compressão 32KB seria mais eficiente em termos de espaço do que uma unidade de bloco de compressão 8KB. Isso significa que a compactação adaptável usando o tamanho padrão de bloco 8KB leva a taxas de eficiência ligeiramente menores, mas também há um benefício significativo para o uso de um tamanho menor de bloco de compressão. As cargas de trabalho de banco de dados incluem uma grande quantidade de atividade de substituição. Substituir um 8KB de um bloco de dados comprimido 32KB requer a leitura de todo o 32KB de dados lógicos, descomprimindo-o, atualizando a região 8KB necessária, recomprimindo e gravando o 32KB inteiro de volta às unidades. Esta é uma operação muito cara para um sistema de storage e é o motivo pelo qual alguns storage arrays concorrentes baseados em tamanhos de bloco de compressão maiores também incorrer em uma penalidade de desempenho significativa nos workloads de banco de dados.



O tamanho do bloco usado pela compressão adaptável pode ser aumentado até 32KBMB. Isso pode melhorar a eficiência do storage e deve ser considerado para arquivos inativos, como logs de transações e arquivos de backup, quando uma quantidade substancial de tais dados é armazenada no array. Em algumas situações, os bancos de dados ativos que usam um tamanho de bloco 16KB ou 32KB também podem se beneficiar do aumento do tamanho de bloco da compactação adaptável a corresponder. Consulte um representante da NetApp ou do parceiro para obter orientação sobre se isso é apropriado para sua carga de trabalho.



Tamanhos de blocos de compactação maiores que 8KBMB não devem ser usados juntamente com a deduplicação em destinos de backup de streaming. A razão é que pequenas alterações nos dados de backup afetam a janela de compressão 32KB. Se a janela mudar, os dados compactados resultantes diferem em todo o arquivo. A deduplicação ocorre após a compactação, o que significa que o mecanismo de deduplicação vê cada backup compactado de forma diferente. Se a deduplicação de backups de streaming for necessária, somente a compactação adaptável de bloco 8KB deve ser usada. A compactação adaptável é preferível, porque funciona em um tamanho de bloco menor e não prejudica a eficiência da deduplicação. Por razões semelhantes, a compactação no lado do host também interfere na eficiência da deduplicação.

Alinhamento da compressão

A compactação adaptável em um ambiente de banco de dados requer alguma consideração do alinhamento do bloco de compressão. Fazer isso é apenas uma preocupação para os dados que estão sujeitos a substituições aleatórias de blocos muito específicos. Essa abordagem é semelhante em conceito ao alinhamento geral do sistema de arquivos, onde o início de um sistema de arquivos deve ser alinhado a um limite de dispositivo 4K e o tamanho de bloco de um sistema de arquivos deve ser um múltiplo de 4K.

Por exemplo, uma gravação 8KBD em um arquivo é compactada somente se ele se alinhar com um limite

8KBD dentro do próprio sistema de arquivos. Este ponto significa que ele deve cair no primeiro 8KB do arquivo, o segundo 8KB do arquivo, e assim por diante. A maneira mais simples de garantir o alinhamento correto é usar o tipo de LUN correto, qualquer partição criada deve ter um deslocamento desde o início do dispositivo que é um múltiplo de 8K, e usar um tamanho de bloco de sistema de arquivos que é um múltiplo do tamanho do bloco de banco de dados.

Dados como backups ou logs de transações são operações sequencialmente escritas que abrangem vários blocos, todos eles compatidos. Portanto, não há necessidade de considerar o alinhamento. O único padrão de e/S de preocupação é as substituições aleatórias de arquivos.

Compactação de dados

A compactação de dados é uma tecnologia que melhora a eficiência da compressão. Como dito anteriormente, a compactação adaptável por si só pode proporcionar, na melhor das hipóteses, economias de 2:1x porque está limitada a armazenar uma e/S de 8KBx em um bloco de 4KB WAFL. Os métodos de compressão com tamanhos de bloco maiores proporcionam melhor eficiência. No entanto, eles não são adequados para dados sujeitos a pequenas substituições de blocos. A descompressão de 32KB unidades de dados, a atualização de uma porção 8KB, a recompressão e a gravação de volta nas unidades cria sobrecarga.

A compactação de dados funciona permitindo que vários blocos lógicos sejam armazenados em blocos físicos. Por exemplo, um banco de dados com dados altamente compressíveis, como texto ou blocos parcialmente completos, pode compactar de 8KB a 1KB. Sem compactação, 1KB PB de dados ainda ocupariam um bloco inteiro de 4KB TB. A compactação de dados in-line permite que 1KB TB de dados compatidos sejam armazenados em apenas 1KB TB de espaço físico, juntamente com outros dados compatidos. Não é uma tecnologia de compressão; é simplesmente uma maneira mais eficiente de alocar espaço nas unidades e, portanto, não deve criar qualquer efeito de desempenho detetável.

O grau de poupança obtido varia. Os dados que já estão compatidos ou criptografados geralmente não podem ser mais compatidos e, portanto, esses conjuntos de dados não se beneficiam com a compactação. Em contraste, datafiles recém-inicializados que contêm pouco mais do que metadados de bloco e zeros compactam até 80:1.

Eficiência de armazenamento sensível à temperatura

A eficiência de armazenamento sensível à temperatura (TSSE) está disponível no ONTAP 9.8 e posterior. Ele depende de mapas de calor de acesso a blocos para identificar blocos acessados com pouca frequência e compactá-los com maior eficiência.

Deduplicação

A deduplicação é a remoção de tamanhos de bloco duplicados de um conjunto de dados. Por exemplo, se o mesmo bloco 4KB existisse em 10 arquivos diferentes, a deduplicação redirecionaria esse bloco 4KB dentro de todos os arquivos 10 para o mesmo bloco físico 4KB. O resultado seria uma melhoria de 10:1 na eficiência para esses dados.

Dados como LUNs de inicialização convidado da VMware geralmente deduplicam muito bem porque consistem em várias cópias dos mesmos arquivos do sistema operacional. A eficiência de 100:1 e maior foi observada.

Alguns dados não contêm dados duplicados. Por exemplo, um bloco Oracle contém um cabeçalho que é globalmente exclusivo para o banco de dados e um trailer que é quase único. Como resultado, a deduplicação de um banco de dados Oracle raramente oferece mais de 1% de economia. A deduplicação com bancos de dados MS SQL é um pouco melhor, mas metadados exclusivos no nível de bloco ainda são uma limitação.

Economia de espaço de até 15% em bancos de dados com 16KB e grandes blocos foram observadas em alguns casos. O 4KB inicial de cada bloco contém o cabeçalho globalmente exclusivo, e o último bloco de 4KB contém o trailer quase único. Os blocos internos são candidatos à deduplicação, embora na prática isso seja quase inteiramente atribuído à deduplicação de dados zerados.

Muitos arrays concorrentes afirmam a capacidade de deduplicar bancos de dados com base na presunção de que um banco de dados é copiado várias vezes. A esse respeito, a deduplicação NetApp também pode ser usada, mas o ONTAP oferece uma opção melhor: A tecnologia NetApp FlexClone. O resultado final é o mesmo; várias cópias de um banco de dados que compartilham a maioria dos blocos físicos subjacentes são criadas. Usar o FlexClone é muito mais eficiente do que ter tempo para copiar arquivos de banco de dados e, em seguida, deduplicá-los. É, na verdade, não duplicação em vez de deduplicação, porque uma duplicata nunca é criada em primeiro lugar.

Eficiência e thin Provisioning

Os recursos de eficiência são formas de thin Provisioning. Por exemplo, um LUN de 100GB GB ocupando um volume de 100GB TB pode ser compactado para 50GB TB. Ainda não há economias reais realizadas porque o volume ainda é 100GB. O volume deve primeiro ser reduzido em tamanho para que o espaço guardado possa ser utilizado noutro local do sistema. Se as alterações posteriores ao LUN 100GBD resultarem em menos compressíveis os dados, o LUN crescerá em tamanho e o volume poderá ser preenchido.

O thin Provisioning é altamente recomendado porque pode simplificar o gerenciamento e fornecer melhorias substanciais na capacidade utilizável com economias de custo associadas. O motivo é simples: Os ambientes de banco de dados geralmente incluem muito espaço vazio, um grande número de volumes e LUNs e dados compressíveis. O provisionamento thick resulta na reserva de espaço no storage para volumes e LUNs, caso eles se tornem 100% cheios e contenham dados 100% não compactáveis. É pouco provável que isso ocorra. O thin Provisioning permite que esse espaço seja recuperado e usado em outros lugares e permite que o gerenciamento de capacidade seja baseado no próprio sistema de storage, em vez de muitos volumes e LUNs menores.

Alguns clientes preferem usar o provisionamento thick, seja para cargas de trabalho específicas ou, geralmente, com base em práticas operacionais e de aquisição estabelecidas.

Atenção: se um volume for provisionado de forma grossa, deve-se ter cuidado para desativar completamente todos os recursos de eficiência para esse volume, incluindo descompressão e remoção de deduplicação usando o `sis undo` comando. O volume não deve aparecer `volume efficiency show` na saída. Se isso acontecer, o volume ainda será parcialmente configurado para recursos de eficiência. Como resultado, as garantias de substituição funcionam de forma diferente, o que aumenta a chance de que a configuração seja ultrapassada fazendo com que o volume fique inesperadamente sem espaço, resultando em erros de e/S do banco de dados.

Práticas recomendadas de eficiência

A NetApp recomenda o seguinte:

Padrões do AFF

Os volumes criados no ONTAP executados em um sistema all-flash AFF são thin Provisioning com todos os recursos de eficiência in-line habilitados. Embora os bancos de dados geralmente não se beneficiem da deduplicação e possam incluir dados não compressíveis, as configurações padrão são, no entanto, apropriadas para quase todas as cargas de trabalho. O ONTAP foi projetado para processar com eficiência todos os tipos de dados e padrões de e/S, resultando ou não em economia. Os padrões só devem ser alterados se os motivos forem totalmente compreendidos e houver um benefício para se desviar.

Recomendações gerais

- Se os volumes e/ou LUNs não forem provisionados de forma fina, você deve desativar todas as configurações de eficiência porque o uso desses recursos não oferece economia e a combinação de provisionamento espesso com eficiência de espaço habilitada pode causar comportamento inesperado, incluindo erros fora do espaço.
- Se os dados não estiverem sujeitos a sobrescritas, como backups ou logs de transação de banco de dados, você poderá obter maior eficiência ativando o TSSE com um período de resfriamento baixo.
- Alguns arquivos podem conter uma quantidade significativa de dados não compressíveis, por exemplo, quando a compactação já está ativada no nível de aplicativo de arquivos são criptografados. Se qualquer um desses cenários for verdadeiro, considere desativar a compactação para permitir uma operação mais eficiente em outros volumes que contenham dados compressíveis.
- Não use a compactação e a deduplicação do 32KB com backups de bancos de dados. Consulte a seção [Compressão adaptável](#) para obter detalhes.

Thin Provisioning

O thin Provisioning para um banco de dados Oracle requer um Planejamento cuidadoso, pois o resultado é configurar mais espaço em um sistema de armazenamento do que necessariamente está fisicamente disponível. Vale muito a pena o esforço porque, quando feito corretamente, o resultado é uma significativa economia de custos e melhorias na capacidade de gerenciamento.

O thin Provisioning vem em várias formas e é parte integrante de muitos recursos que a ONTAP oferece a um ambiente de aplicações empresariais. O provisionamento de thin também está intimamente relacionado às tecnologias de eficiência pelo mesmo motivo: Os recursos de eficiência permitem que mais dados lógicos sejam armazenados do que o que existe tecnicamente no sistema de storage.

Quase qualquer uso de snapshots envolve thin Provisioning. Por exemplo, um banco de dados 10TB típico no storage NetApp inclui cerca de 30 dias de snapshots. Esse arranjo resulta em aproximadamente 10TB GB de dados visíveis no sistema de arquivos ativo e 300TB dedicados a snapshots. O total de 310TB TB de storage geralmente reside em aproximadamente 12TB a 15TB TB de espaço. O banco de dados ativo consome 10TB, e os 300TB restantes de dados requerem apenas 2TB a 5TB GB de espaço porque apenas as alterações aos dados originais são armazenadas.

A clonagem também é um exemplo de thin Provisioning. Um grande cliente da NetApp criou 40 clones de um banco de dados 80TB para uso por desenvolvimento. Se todos os desenvolvedores do 40 que usam esses clones sobrescrevessem cada bloco em cada arquivo de dados, mais de 3,2PB GB de armazenamento seriam necessários. Na prática, a rotatividade é baixa e a exigência de espaço coletivo está mais próxima de 40TB, porque apenas as alterações são armazenadas nas unidades.

Gerenciamento de espaço

Alguns cuidados devem ser tomados com o thin Provisioning de um ambiente de aplicativo porque as taxas de alteração de dados podem aumentar inesperadamente. Por exemplo, o consumo de espaço devido a snapshots pode crescer rapidamente se as tabelas do banco de dados forem reindexadas ou se aplicar patches em larga escala aos convidados VMware. Um backup perdido pode gravar uma grande quantidade de dados em um tempo muito curto. Finalmente, pode ser difícil recuperar alguns aplicativos se um sistema de arquivos ficar sem espaço livre inesperadamente.

Felizmente, esses riscos podem ser resolvidos com uma configuração cuidadosa de `volume-autogrow` políticas e `snapshot-autodelete`. Como os nomes deles implicam, essas opções permitem que um

usuário crie políticas que desobstruam automaticamente o espaço consumido por snapshots ou aumentem um volume para acomodar dados adicionais. Muitas opções estão disponíveis e as necessidades variam de acordo com o cliente.

Consulte o "[documentação de gerenciamento de storage lógico](#)" para obter uma discussão completa sobre esses recursos.

Reservas fracionárias

A reserva fracionária refere-se ao comportamento de um LUN em um volume com relação à eficiência do espaço. Quando a opção `fractional-reserve` é definida para 100%, todos os dados no volume podem experimentar 100% de rotatividade com qualquer padrão de dados sem esgotar espaço no volume.

Por exemplo, considere um banco de dados em um único LUN 250GB em um volume 1TB. Criar um instantâneo resultaria imediatamente na reserva de 250GBMB de espaço adicional no volume para garantir que o volume não fique sem espaço por qualquer motivo. O uso de reservas fracionárias geralmente é desperdiçado porque é extremamente improvável que cada byte no volume do banco de dados precise ser substituído. Não há razão para reservar espaço para um evento que nunca acontece. Ainda assim, se um cliente não puder monitorar o consumo de espaço em um sistema de armazenamento e tiver certeza de que o espaço nunca se esgote, reservas fracionárias de 100% seriam necessárias para usar snapshots.

Compactação e deduplicação

Compactação e deduplicação são ambas formas de thin Provisioning. Por exemplo, um espaço físico de dados de 50TB GB pode ser compactado para 30TB GB, resultando em uma economia de 20TB GB. Para que a compactação possa gerar quaisquer benefícios, alguns desses 20TB precisam ser usados para outros dados, ou o sistema de storage precisa ser adquirido com menos de 50TB TB. O resultado é o armazenamento de mais dados do que o tecnicamente disponível no sistema de armazenamento. Do ponto de vista dos dados, há 50TB TB de dados, embora ocupe apenas 30TB TB nas unidades.

Há sempre a possibilidade de que a compressibilidade de um conjunto de dados mude, o que resultaria em aumento do consumo de espaço real. Esse aumento no consumo significa que a compactação deve ser gerenciada como com outras formas de thin Provisioning em termos de monitoramento e uso `volume-autogrow` e `snapshot-autodelete`.

A compactação e a deduplicação são discutidas em mais detalhes na seção `xref:./oracle/efficiency.html`

Compressão e reservas fracionárias

A compactação é uma forma de thin Provisioning. Reservas fracionárias afetam o uso da compressão, com uma nota importante; o espaço é reservado antes da criação do snapshot. Normalmente, a reserva fracionária só é importante se existir um instantâneo. Se não houver instantâneo, a reserva fracionária não é importante. Este não é o caso da compressão. Se um LUN for criado em um volume com compactação, o ONTAP preservará espaço para acomodar um snapshot. Esse comportamento pode ser confuso durante a configuração, mas é esperado.

Por exemplo, considere um volume de 10GB TB com um LUN de 5GB GB que foi comprimido até 2,5GB TB sem instantâneos. Considere estes dois cenários:

- A reserva fracionária é de 100 resultados na utilização de 7,5GB
- A reserva fracionária é de 0 resultados na utilização de 2,5GB

O primeiro cenário inclui 2,5GB GB de consumo de espaço para dados atuais e 5GB GB de espaço para representar 100% de rotatividade da fonte em antecipação ao uso de instantâneos. O segundo cenário não

reserva espaço extra.

Embora esta situação possa parecer confusa, é improvável que seja encontrada na prática. A compactação implica thin Provisioning, e thin Provisioning em um ambiente LUN requer reservas fracionárias. É sempre possível que os dados compactados sejam sobrescritos por algo não compressível, o que significa que um volume deve ser thin Provisioning para compactação para resultar em qualquer economia.

A NetApp recomenda as seguintes configurações de reserva:



- Defina `fractional-reserve` como 0 quando o monitoramento de capacidade básica estiver em vigor, juntamente com `volume-autogrow` e `snapshot-autodelete`.
- Defina `fractional-reserve` para 100 se não houver capacidade de monitorização ou se for impossível esgotar o espaço em qualquer circunstância.

Espaço livre e alocação de espaço LVM

A eficiência do thin Provisioning de LUNs ativos em um ambiente de sistema de arquivos pode ser perdida com o tempo, à medida que os dados são excluídos. A menos que os dados excluídos sejam sobrescritos com zeros (veja também "ASMRU" ou o espaço seja liberado com recuperação de espaço de recorte/DESMAPEAMENTO, os dados "apagados" ocupam mais e mais espaço em branco não alocado no sistema de arquivos. Além disso, o thin Provisioning de LUNs ativos é de uso limitado em muitos ambientes de banco de dados porque os arquivos de dados são inicializados em seu tamanho total no momento da criação.

Um Planejamento cuidadoso da configuração do LVM pode melhorar a eficiência e minimizar a necessidade de provisionamento de armazenamento e redimensionamento de LUN. Quando um LVM, como Veritas VxVM ou Oracle ASM, é usado, os LUNs subjacentes são divididos em extensões que só são usadas quando necessário. Por exemplo, se um conjunto de dados começar com 2TB TB de tamanho, mas puder crescer para 10TB TB com o tempo, esse conjunto de dados poderá ser colocado em 10TB PB de LUNs provisionados, organizados em um grupo de discos LVM. Ele ocuparia apenas 2TBMB de espaço no momento da criação e só reivindicaria espaço adicional, pois as extensões são alocadas para acomodar o crescimento dos dados. Este processo é seguro, desde que o espaço seja monitorado.

Failover/switchover da ONTAP

É necessário entender as funções de takeover e switchover do storage para garantir que as operações do banco de dados Oracle não sejam interrompidas por essas operações. Além disso, os argumentos usados pelas operações de aquisição e comutação podem afetar a integridade dos dados se usados incorretamente.

- Em condições normais, as gravações recebidas em um determinado controlador são espelhadas de forma síncrona para seu parceiro. Em um ambiente NetApp MetroCluster, as gravações também são espelhadas em um controle remoto. Até que uma gravação seja armazenada em Mídia não volátil em todos os locais, ela não será reconhecida para a aplicação host.
- A Mídia que armazena os dados de gravação é chamada de memória não volátil ou NVMEM. Ele também é por vezes referido à memória de acesso aleatório não volátil (NVRAM), e pode ser considerado como um cache de escrita, embora funcione como um diário. Em uma operação normal, os dados do NVMEM não são lidos; eles são usados apenas para proteger os dados em caso de falha de software ou hardware. Quando os dados são gravados em unidades, os dados são transferidos da RAM no sistema, e não da NVMEM.
- Durante uma operação de takeover, um nó em um par de alta disponibilidade (HA) assume as operações

de seu parceiro. Um switchover é essencialmente o mesmo, mas se aplica às configurações do MetroCluster nas quais um nó remoto assume as funções de um nó local.

Durante as operações de manutenção de rotina, uma operação de aquisição ou comutação de storage deve ser transparente, exceto para uma breve pausa em operações à medida que os caminhos de rede mudam. No entanto, a rede pode ser complicada, e é fácil fazer erros. Por isso, a NetApp recomenda fortemente testar cuidadosamente as operações de takeover e switchover antes de colocar um sistema de storage em produção. Fazer isso é a única maneira de ter certeza de que todos os caminhos de rede estão configurados corretamente. Em um ambiente SAN, verifique cuidadosamente a saída do comando `sanlun lun show -p` para garantir que todos os caminhos primários e secundários esperados estejam disponíveis.

Deve ser tomado cuidado ao emitir uma tomada ou mudança forçada. Forçar uma alteração na configuração de armazenamento com essas opções significa que o estado do controlador que possui as unidades é ignorado e o nó alternativo assume forçosamente o controle das unidades. Forçar uma aquisição incorreta pode resultar em perda ou corrupção de dados. Isso ocorre porque uma tomada ou comutação forçada pode descartar o conteúdo do NVMEM. Após a conclusão do takeover ou switchover, a perda desses dados significa que os dados armazenados nas unidades podem reverter para um estado um pouco mais antigo do ponto de vista do banco de dados.

Uma aquisição forçada com um par de HA normal raramente deve ser necessária. Em quase todos os cenários de falha, um nó desliga e informa o parceiro para que ocorra um failover automático. Existem alguns casos de borda, como uma falha contínua em que a interconexão entre nós é perdida e, em seguida, um controlador é perdido, na qual uma tomada pública forçada é necessária. Em tal situação, o espelhamento entre nós é perdido antes da falha do controlador, o que significa que o controlador sobrevivente não teria mais uma cópia das gravações em andamento. Então, a aquisição precisa ser forçada, o que significa que os dados podem ser perdidos.

A mesma lógica se aplica a um switchover MetroCluster. Em condições normais, uma mudança é quase transparente. No entanto, um desastre pode resultar em uma perda de conectividade entre o local sobrevivente e o local do desastre. Do ponto de vista do site sobrevivente, o problema não poderia ser mais do que uma interrupção na conectividade entre sites, e o site original ainda pode estar processando dados. Se um nó não puder verificar o estado do controlador principal, somente um switchover forçado é possível.

A NetApp recomenda tomar as seguintes precauções:

- Tenha muito cuidado para não forçar acidentalmente uma aquisição ou uma mudança. Normalmente, forçar não deve ser necessário e forçar a mudança pode causar perda de dados.
- Se for necessário um controle ou switchover forçado, certifique-se de que os aplicativos sejam desligados, todos os sistemas de arquivos sejam desmontados e os grupos de volume do gerenciador de volumes lógico (LVM) sejam variados. Os grupos de discos ASM devem ser desmontados.
- No caso de um switchover forçado do MetroCluster, feche o nó com falha de todos os recursos de storage sobreviventes. Para obter mais informações, consulte o Guia de gerenciamento e recuperação de desastres do MetroCluster para obter a versão relevante do ONTAP.



MetroCluster e vários agregados

O MetroCluster é uma tecnologia de replicação síncrona que alterna para o modo assíncrono se a conectividade for interrompida. Essa é a solicitação mais comum dos clientes, pois a replicação síncrona garantida significa que a interrupção na conectividade do local leva a uma interrupção completa da e/S do banco de dados, retirando o banco de dados do serviço.

Com o MetroCluster, agregados ressincronizam rapidamente após a restauração da conectividade. Diferentemente de outras tecnologias de storage, o MetroCluster nunca deve exigir um reespehamento completo após falha do local. Apenas as alterações delta devem ser enviadas.

Em conjuntos de dados que abrangem agregados, existe um pequeno risco de que etapas adicionais de recuperação de dados sejam necessárias em um cenário de desastre contínuo. Especificamente, se (a) a conectividade entre sites for interrompida, (b) a conectividade for restaurada, (c) os agregados atingirem um estado em que alguns são sincronizados e alguns não são, e então (d) o local primário é perdido, o resultado é um local sobrevivente no qual os agregados não são sincronizados uns com os outros. Se isso acontecer, partes do conjunto de dados são sincronizadas entre si e não é possível abrir aplicativos, bancos de dados ou datastores sem recuperação. Se um conjunto de dados abranger agregados, o NetApp recomenda fortemente a utilização de backups baseados em snapshot com uma das muitas ferramentas disponíveis para verificar a capacidade de recuperação rápida neste cenário incomum.

Configuração do banco de dados

Tamanhos de blocos

O ONTAP usa internamente um tamanho de bloco variável, o que significa que os bancos de dados Oracle podem ser configurados com qualquer tamanho de bloco desejado. No entanto, os tamanhos de bloco de sistema de arquivos podem afetar o desempenho e, em alguns casos, um tamanho maior de bloco de refazer pode melhorar o desempenho.

Tamanhos de blocos de arquivos de dados

Alguns sistemas operacionais oferecem uma escolha de tamanhos de bloco de sistema de arquivos. Para sistemas de arquivos que suportam datafiles Oracle, o tamanho do bloco deve ser 8KB quando a compactação é usada. Quando a compressão não é necessária, um tamanho de bloco de 8KB ou 4KB pode ser usado.

Se um arquivo de dados for colocado em um sistema de arquivos com um bloco de 512 bytes, arquivos desalinhados são possíveis. O LUN e o sistema de arquivos podem estar alinhados adequadamente com base nas recomendações do NetApp, mas a e/S do arquivo estaria desalinhada. Tal desalinhamento causaria graves problemas de desempenho.

Os sistemas de arquivos que suportam os logs de refazer devem usar um tamanho de bloco que seja um múltiplo do tamanho do bloco de refazer. Isso geralmente requer que o sistema de arquivos de log refazer e o log refazer usem um tamanho de bloco de 512 bytes.

Refazer tamanhos de bloco

Com taxas de recozimento muito altas, é possível que 4KB tamanhos de bloco tenham um melhor desempenho, pois altas taxas de recozimento permitem que a e/S seja executada em operações cada vez menores e mais eficientes. Se as taxas de refazer forem maiores que 50Mbps, considere testar um tamanho de bloco 4KB.

Alguns problemas de clientes foram identificados com bancos de dados usando logs de refazer com um tamanho de bloco de 512 bytes em um sistema de arquivos com um tamanho de bloco de 4KB e muitas transações muito pequenas. A sobrecarga envolvida na aplicação de várias alterações de 512 bytes a um único bloco de sistema de arquivos 4KB levou a problemas de desempenho que foram resolvidos alterando o sistema de arquivos para usar um tamanho de bloco de 512 bytes.



A NetApp recomenda que você não altere o tamanho do bloco refazer, a menos que seja aconselhado por uma organização de suporte ao cliente ou serviços profissionais relevante ou que a alteração seja baseada na documentação oficial do produto.

db_file_multiblock_read_count

O `db_file_multiblock_read_count` parâmetro controla o número máximo de blocos de banco de dados Oracle que o Oracle lê como uma única operação durante a e/S sequencial

Este parâmetro não afeta, no entanto, o número de blocos que o Oracle lê durante todas e quaisquer operações de leitura, nem afeta e/S aleatórias. Apenas o tamanho de bloco de e/S sequenciais é afetado.

A Oracle recomenda que o usuário deixe esse parâmetro desconfigurado. Isso permite que o software de banco de dados defina automaticamente o valor ideal. Isso geralmente significa que esse parâmetro é definido para um valor que produz um tamanho de e/S de 1MB. Por exemplo, uma leitura 1MB de 8KB blocos exigiria que 128 blocos fossem lidos, e o valor padrão para este parâmetro seria, portanto, 128.

A maioria dos problemas de desempenho do banco de dados observados pelo NetApp nas instalações dos clientes envolve uma configuração incorreta para este parâmetro. Houve razões válidas para alterar esse valor com as versões 8 e 9 do Oracle. Como resultado, o parâmetro pode estar inconscientemente presente em `init.ora` arquivos porque o banco de dados foi atualizado para o Oracle 10 e posterior. Uma configuração legada de 8 ou 16, comparada a um valor padrão de 128, danifica significativamente o desempenho de e/S sequenciais.



NetApp recomenda a configuração `db_file_multiblock_read_count` do parâmetro não deve estar presente no `init.ora` arquivo. O NetApp nunca encontrou uma situação em que a alteração desse parâmetro melhorou o desempenho, mas há muitos casos em que causou danos claros ao throughput sequencial de e/S.

systema de arquivos_options

O parâmetro de inicialização do Oracle `filesystemio_options` controla o uso de e/S assíncrona e direta

Ao contrário da crença comum, a e/S assíncrona e direta não são mutuamente exclusivas. A NetApp observou que esse parâmetro é frequentemente mal configurado em ambientes de clientes, e essa configuração incorreta é diretamente responsável por muitos problemas de desempenho.

E/S assíncrona significa que as operações de e/S Oracle podem ser paralelizadas. Antes da disponibilidade de e/S assíncrona em vários sistemas operacionais, os usuários configuraram vários processos `dbwriter` e alteraram a configuração do processo do servidor. Com e/S assíncrona, o próprio sistema operacional executa e/S em nome do software de banco de dados de maneira altamente eficiente e paralela. Esse processo não coloca os dados em risco, e operações críticas, como o Oracle refazer o log, ainda são executadas de forma síncrona.

A e/S direta ignora o cache do buffer do sistema operacional. E/S em um sistema UNIX normalmente flui através do cache de buffer do sistema operacional. Isso é útil para aplicativos que não mantêm um cache interno, mas o Oracle tem seu próprio cache de buffer dentro do SGA. Em quase todos os casos, é melhor habilitar e/S direta e alocar RAM do servidor para o SGA em vez de confiar no cache de buffer do sistema operacional. O Oracle SGA usa a memória de forma mais eficiente. Além disso, quando a e/S flui pelo buffer do sistema operacional, ela está sujeita a processamento adicional, o que aumenta as latências. As latências

aumentadas são especialmente perceptíveis com e/S de gravação intensa quando a baixa latência é um requisito essencial.

As opções para `filesystemio_options` são:

- **assíncrono.** A Oracle envia solicitações de e/S para o sistema operacional para processamento. Esse processo permite que a Oracle execute outro trabalho em vez de esperar pela conclusão de e/S e, assim, aumenta a paralelização de e/S.
- **directio.** A Oracle executa e/S diretamente contra arquivos físicos em vez de rotear e/S pelo cache do sistema operacional do host.
- **nenhum.** A Oracle usa e/S síncrona e em buffer Nesta configuração, a escolha entre processos de servidor compartilhado e dedicado e o número de dbwriters são mais importantes.
- *** setall.*** A Oracle usa e/S assíncrona e direta Em quase todos os casos, o uso de `setall` é ideal.



O `filesystemio_options` parâmetro não tem efeito em ambientes DNFS e ASM. O uso de DNFS ou ASM resulta automaticamente no uso de e/S assíncrona e direta

Alguns clientes encontraram problemas assíncronos de I/O no passado, especialmente com versões anteriores do Red Hat Enterprise Linux 4 (RHEL4). Alguns conselhos desatualizados na internet ainda sugerem evitar IO assíncrono por causa de informações desatualizadas. E/S assíncrona é estável em todos os sistemas operacionais atuais. Não há razão para desativá-lo, ausente um bug conhecido com o sistema operacional.

Se um banco de dados tiver usado e/S em buffer, um switch para direcionar e/S também pode garantir uma alteração no tamanho do SGA. A desativação da e/S armazenada em buffer elimina o benefício de desempenho que o cache do sistema operacional do host fornece para o banco de dados. Adicionar RAM de volta ao SGA repara este problema. O resultado líquido deve ser uma melhoria no desempenho de e/S.

Embora seja quase sempre melhor usar RAM para o Oracle SGA do que para cache de buffer do SO, pode ser impossível determinar o melhor valor. Por exemplo, pode ser preferível usar e/S em buffer com tamanhos SGA muito pequenos em um servidor de banco de dados com muitas instâncias Oracle ativas intermitentemente. Este arranjo permite o uso flexível da RAM livre restante no sistema operacional por todas as instâncias de banco de dados em execução. Esta é uma situação altamente incomum, mas tem sido observada em alguns sites de clientes.



A NetApp recomenda a configuração `filesystemio_options` para `setall`, mas esteja ciente de que, em algumas circunstâncias, a perda do cache de buffer do host pode exigir um aumento no SGA Oracle.

Tempos limite do RAC

O Oracle RAC é um produto exclusivo com vários tipos de processos internos de heartbeat que monitoram a integridade do cluster.



As informações na "[número de identificação](#)" seção incluem informações críticas para ambientes Oracle RAC que usam armazenamento em rede e, em muitos casos, as configurações padrão do Oracle RAC precisarão ser alteradas para garantir que o cluster RAC sobreviva às alterações do caminho da rede e às operações de failover/switchover de armazenamento.

disktimeout

O parâmetro RAC principal relacionado ao armazenamento é `disktimeout`. Este parâmetro controla o limite dentro do qual a e/S do arquivo de votação deve ser concluída. Se o `disktimeout` parâmetro for excedido, o nó RAC será despejado do cluster. O padrão para este parâmetro é 200. Este valor deve ser suficiente para os procedimentos normais de aquisição de armazenamento e de giveback.

A NetApp recomenda fortemente testar cuidadosamente as configurações do RAC antes de colocá-las em produção, pois muitos fatores afetam uma aquisição ou a giveback. Além do tempo necessário para a conclusão do failover de armazenamento, também é necessário tempo adicional para que as alterações do protocolo de controle de agregação de link (LACP) se propaguem. Além disso, o software de multipathing SAN deve detetar um tempo limite de e/S e tentar novamente em um caminho alternativo. Se um banco de dados estiver extremamente ativo, uma grande quantidade de e/S deve ser enfileirada e tentada novamente antes de o disco de votação ser processado.

Se não for possível executar uma aquisição de armazenamento real ou giveback, o efeito poderá ser simulado com testes de pull de cabo no servidor de banco de dados.

A NetApp recomenda o seguinte:



- Deixando o `disktimeout` parâmetro no valor padrão de 200.
- Sempre teste uma configuração RAC completamente.

número de identificação

O `misscount` parâmetro normalmente afeta apenas o batimento cardíaco da rede entre nós RAC. O padrão é 30 segundos. Se os binários de grade estiverem em um storage array ou a unidade de inicialização do sistema operacional não for local, esse parâmetro pode se tornar importante. Isso inclui hosts com unidades de inicialização localizadas em uma SAN FC, sistemas operacionais iniciados por NFS e unidades de inicialização localizados em datastores de virtualização, como um arquivo VMDK.

Se o acesso a uma unidade de inicialização for interrompido por uma aquisição de armazenamento ou giveback, é possível que a localização binária da grade ou todo o sistema operacional fique temporariamente suspenso. O tempo necessário para o ONTAP concluir a operação de storage e para o sistema operacional alterar caminhos e retomar e/S pode exceder o `misscount` limite. Como resultado, um nó é despejado imediatamente após a conectividade com o LUN de inicialização ou binários de grade ser restaurada. Na maioria dos casos, o despejo e a reinicialização subsequente ocorrem sem mensagens de Registro para indicar o motivo da reinicialização. Nem todas as configurações são afetadas, portanto, teste qualquer host baseado em SAN-boot, NFS-boot ou datastore em um ambiente RAC para que o RAC permaneça estável se a comunicação com a unidade de inicialização for interrompida.

No caso de unidades de inicialização não locais ou binários de hospedagem de sistemas de arquivos não locais `grid`, o `misscount` precisará ser alterado para corresponder ``disktimeout`` ao . Se esse parâmetro for alterado, realize testes adicionais para identificar também quaisquer efeitos no comportamento do RAC, como o tempo de failover do nó.

A NetApp recomenda o seguinte:



- Deixe o `misscount` parâmetro no valor padrão de 30, a menos que uma das seguintes condições se aplique:
 - `grid` Os binários estão localizados em uma unidade conetada à rede, incluindo unidades baseadas em NFS, iSCSI, FC e datastore.
 - O sistema operacional é inicializado pela SAN.
- Nesses casos, avalie o efeito das interrupções da rede que afetam o acesso ao sistema operacional ou `GRID_HOME` aos sistemas de arquivos. Em alguns casos, tais interrupções fazem com que os daemons do Oracle RAC parem, o que pode levar a um `misscount` tempo limite e despejo baseado em -. O tempo limite padrão é 27 segundos, que é o valor de `misscount` menos `reboottime`. Nesses casos, aumente `misscount` para 200 para corresponder `disktimeout`.

Configuração de host

AIX

Tópicos de configuração para banco de dados Oracle no IBM AIX com ONTAP.

E/S simultânea

Alcançar o desempenho ideal no IBM AIX requer o uso de e/S simultâneas. Sem I/O concorrente, as limitações de desempenho provavelmente são porque o AIX executa e/S atômica serializada, o que incorre em sobrecarga significativa.

Originalmente, o NetApp recomendou o uso da `cio` opção de montagem para forçar o uso de e/S concorrente no sistema de arquivos, mas esse processo teve desvantagens e não é mais necessário. Desde a introdução do AIX 5,2 e do Oracle 10gR1, o Oracle no AIX pode abrir arquivos individuais para IO simultâneo, em vez de forçar e/S simultânea em todo o sistema de arquivos.

O melhor método para ativar e/S concorrente é definir o `init.ora` parâmetro `filesystemio_options` como `setall`. Isso permite que a Oracle abra arquivos específicos para uso com e/S concorrente.

Usar `cio` como opção de montagem força o uso de e/S concorrente, o que pode ter consequências negativas. Por exemplo, forçar e/S concorrente desativa o `readahead` em sistemas de arquivos, o que pode danificar o desempenho de e/S que ocorre fora do software de banco de dados Oracle, como copiar arquivos e executar backups de fita. Além disso, produtos como Oracle GoldenGate e SAP BR*Tools não são compatíveis com o uso da `cio` opção de montagem com certas versões do Oracle.

A NetApp recomenda o seguinte:



- Não use a `cio` opção de montagem no nível do sistema de arquivos. Em vez disso, ative a e/S concorrente através do uso `filesystemio_options=setall` do .
- Utilize apenas a `cio` opção de montagem se não for possível definir `filesystemio_options=setall`.

Opções de montagem AIX NFS

A tabela a seguir lista as opções de montagem NFS AIX para bancos de dados de instância única Oracle.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Registros do Redo ControlFiles Datafiles	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,intr</code>

A tabela a seguir lista as opções de montagem NFS AIX para RAC.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Registros do Redo ControlFiles Datafiles	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac</code>
CRS/Voting	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac</code>
Dedicado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Compartilhado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr</code>

A principal diferença entre as opções de montagem de instância única e RAC é a adição `noac` das opções de montagem. Essa adição tem o efeito de desabilitar o cache do sistema operacional do host, que permite que todas as instâncias do cluster RAC tenham uma visão consistente do estado dos dados.

Embora o uso da `cio` opção de montagem e do `init.ora` parâmetro `filesystemio_options=setall` tenha o mesmo efeito de desabilitar o cache do host, ainda é necessário usar `noac`o` . `noac`. É necessário para implantações compartilhadas ORACLE_HOME para facilitar a consistência de arquivos, como arquivos de senha Oracle e `spfile` arquivos de parâmetros. Se cada instância em um cluster RAC tiver um dedicado ORACLE_HOME, esse parâmetro não será necessário.

AIX jfs/jfs2 Opções de montagem

A tabela a seguir lista as opções de montagem do AIX jfs/jfs2.

Tipo de ficheiro	Opções de montagem
ADR Home	Predefinições

Tipo de ficheiro	Opções de montagem
Registros do Redo ControlFiles Datafiles	Predefinições
ORACLE_HOME	Predefinições

Antes de usar dispositivos AIX `hdisk` em qualquer ambiente, incluindo bancos de dados, verifique o parâmetro `queue_depth`. Este parâmetro não é a profundidade da fila HBA; em vez disso, ele se relaciona com a profundidade da fila SCSI do indivíduo `hdisk device`. Depending on how the LUNs are configured, the value for `queue_depth` pode ser muito baixa para um bom desempenho. Os testes mostraram que o valor ideal é 64.

HP-UX

Tópicos de configuração para banco de dados Oracle no HP-UX com ONTAP.

Opções de montagem NFS HP-UX

A tabela a seguir lista as opções de montagem HP-UX NFS para uma única instância.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>
Arquivos de controle Datafiles Redo logs	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,forcedirectio,nointr,suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>

A tabela a seguir lista as opções de montagem HP-UX NFS para RAC.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,noac,suid</code>
Arquivos de controle Datafiles Redo logs	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio,suid</code>
CRS/votação	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio,suid</code>
Dedicado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>

Tipo de ficheiro	Opções de montagem
Compartilhado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,suid</code>

A principal diferença entre as opções de montagem de uma única instância e RAC é a adição `noac` e `forcedirectio` às opções de montagem. Essa adição tem o efeito de desabilitar o cache do sistema operacional do host, o que permite que todas as instâncias do cluster RAC tenham uma visão consistente do estado dos dados. Embora o uso do `init.ora` parâmetro `filesystemio_options=setall` tenha o mesmo efeito de desabilitar o cache do host, ainda é necessário usar `noac` e `forcedirectio`.

O motivo `noac` necessário para implantações compartilhadas `ORACLE_HOME` é facilitar a consistência de arquivos, como arquivos de senha Oracle e `spfiles`. Se cada instância em um cluster RAC tiver um dedicado `ORACLE_HOME`, esse parâmetro não será necessário.

Opções de montagem HP-UX VxFS

Use as seguintes opções de montagem para sistemas de arquivos que hospedam binários Oracle:

```
delaylog,nodatainlog
```

Use as seguintes opções de montagem para sistemas de arquivos que contêm datafiles, logs de refazer, logs de arquivamento e arquivos de controle nos quais a versão do HP-UX não suporta e/S simultânea:

```
nodatainlog,mincache=direct,convosync=direct
```

Quando a e/S simultânea for suportada (VxFS 5.0.1 e posterior, ou com o ServiceGuard Storage Management Suite), utilize estas opções de montagem para sistemas de ficheiros que contenham ficheiros de dados, registos refeitos, registos de arquivo e ficheiros de controlo:

```
delaylog,cio
```



O parâmetro `db_file_multiblock_read_count` é especialmente crítico em ambientes VxFS. A Oracle recomenda que esse parâmetro permaneça desconfigurado no Oracle 10g R1i e posterior, a menos que especificamente direcionado de outra forma. O padrão com um tamanho de bloco Oracle 8KB é 128. Se o valor deste parâmetro for forçado a 16 ou menos, remova a `convosync=direct` opção de montagem porque pode danificar o desempenho sequencial de e/S. Esta etapa prejudica outros aspetos do desempenho e só deve ser tomada se o valor de `db_file_multiblock_read_count` tiver de ser alterado do valor padrão.

Linux

Tópicos de configuração específicos para o sistema operacional Linux.

Tabelas de slots TCP do Linux NFSv3

As tabelas de slot TCP são equivalentes a NFSv3 mm de profundidade de fila do adaptador de barramento do host (HBA). Essas tabelas controlam o número de operações NFS que podem ficar pendentes de uma só vez. O valor padrão é geralmente 16, o que é muito baixo para um desempenho ideal. O problema oposto ocorre em kernels Linux mais recentes, que podem aumentar automaticamente o limite da tabela de slots TCP para um nível que satura o servidor NFS com solicitações.

Para um desempenho ideal e para evitar problemas de desempenho, ajuste os parâmetros do kernel que controlam as tabelas de slots TCP.

Executar o `sysctl -a | grep tcp.*.slot_table` comando e respeitar os seguintes parâmetros:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Todos os sistemas Linux devem incluir `sunrpc.tcp_slot_table_entries`, mas apenas alguns incluem `sunrpc.tcp_max_slot_table_entries`. Ambos devem ser definidos para 128.

Cuidado

A falha em definir esses parâmetros pode ter efeitos significativos no desempenho. Em alguns casos, o desempenho é limitado porque o sistema operacional linux não está emitindo e/S suficiente. Em outros casos, as latências de e/S aumentam à medida que o sistema operacional linux tenta emitir mais e/S do que pode ser reparado.

Opções de montagem em NFS do Linux

A tabela a seguir lista as opções de montagem NFS do Linux para uma única instância.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
Arquivos de controle Datafiles Redo logs	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr</code>

A tabela a seguir lista as opções de montagem NFS do Linux para RAC.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,actimeo=0</code>

Tipo de ficheiro	Opções de montagem
Arquivos de controle arquivos de dados Redo logs	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>
CRS/votação	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,actimeo=0</code>
Dedicado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Compartilhado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>

A principal diferença entre as opções de montagem de instância única e RAC é a adição `actimeo=0` das opções de montagem. Essa adição tem o efeito de desabilitar o cache do sistema operacional do host, o que permite que todas as instâncias do cluster RAC tenham uma visão consistente do estado dos dados. Embora o uso do `init.ora` parâmetro `filesystemio_options=setall` tenha o mesmo efeito de desabilitar o cache do host, ainda é necessário usar ``actimeo=0`` .

O motivo `actimeo=0` necessário para implantações compartilhadas ORACLE_HOME é facilitar a consistência de arquivos, como os arquivos de senha Oracle e `spfiles`. Se cada instância em um cluster RAC tiver um dedicado ORACLE_HOME, esse parâmetro não será necessário.

Geralmente, arquivos que não sejam de banco de dados devem ser montados com as mesmas opções usadas para datafiles de instância única, embora aplicativos específicos possam ter requisitos diferentes. Evite as opções de montagem `noac` e `actimeo=0`, se possível, porque essas opções desativam a leitura e o buffer no nível do sistema de arquivos. Isso pode causar problemas graves de desempenho para processos como extração, tradução e carregamento.

ACESSO e GETATTR

Alguns clientes observaram que um nível extremamente alto de outros IOPS, como O ACCESS e GETATTR, pode dominar suas cargas de trabalho. Em casos extremos, operações como leituras e gravações podem ser tão baixas quanto 10% do total. Este é um comportamento normal com qualquer banco de dados que inclua o uso `actimeo=0` e/ou `noac` no Linux porque essas opções fazem com que o sistema operacional Linux recarregue constantemente metadados de arquivos do sistema de armazenamento. Operações como ACCESS e GETATTR são operações de baixo impactos que são atendidas a partir do cache ONTAP em um ambiente de banco de dados. Não devem ser consideradas IOPS originais, como leituras e gravações, que criem verdadeira demanda em sistemas de storage. No entanto, esses outros IOPS criam alguma carga, especialmente em ambientes RAC. Para resolver esta situação, ative o DNFS, que ignora o cache do buffer do sistema operacional e evita essas operações desnecessárias de metadados.

Linux Direct NFS

Uma opção de montagem adicional, chamada `nosharecache`, é necessária quando (a) o DNFS está ativado e (b) um volume de origem é montado mais de uma vez em um único servidor (c) com uma montagem NFS aninhada. Essa configuração é vista principalmente em ambientes compatíveis com aplicações SAP. Por exemplo, um único volume em um sistema NetApp pode ter um diretório localizado em `/vol/oracle/base` e um segundo em `/vol/oracle/home`. Se `/vol/oracle/base` for montado em `/oracle` e `/vol/oracle/home` for montado em `/oracle/home`, o resultado serão montagens NFS aninhadas que se originam na mesma fonte.

O sistema operacional pode detectar o fato de que `/oracle` e `/oracle/home` residir no mesmo volume, que é o mesmo sistema de arquivos de origem. Em seguida, o SO usa o mesmo identificador de dispositivo para acessar os dados. Isso melhora o uso do cache do sistema operacional e outras operações, mas interfere com o DNFS. Se o DNFS tiver de acessar a um ficheiro, como o `spfile`, ligado `/oracle/home`, poderá tentar, erroneamente, utilizar o caminho errado para os dados. O resultado é uma operação de e/S com falha. Nessas configurações, adicione a `nosharecache` opção de montagem a qualquer sistema de arquivos NFS que compartilhe um volume de origem com outro sistema de arquivos NFS nesse host. Isso força o sistema operacional Linux a alocar um identificador de dispositivo independente para esse sistema de arquivos.

Linux Direct NFS e Oracle RAC

O uso do DNFS tem benefícios especiais de desempenho para o Oracle RAC no sistema operacional Linux porque o Linux não tem um método para forçar e/S direto, o que é necessário com RAC para coerência entre os nós. Como solução alternativa, o Linux requer o uso da `actimeo=0` opção de montagem, que faz com que os dados de arquivo expirem imediatamente do cache do sistema operacional. Essa opção, por sua vez, força o cliente NFS Linux a reler constantemente os dados de atributos, o que danifica a latência e aumenta a carga no controlador de armazenamento.

A ativação do DNFS ignora o cliente NFS do host e evita esse dano. Vários clientes relataram melhorias significativas no desempenho em clusters RAC e reduções significativas na carga do ONTAP (especialmente em relação a outros IOPS) ao habilitar o DNFS.

Linux Direct NFS e arquivo orafstab

Ao usar DNFS no Linux com a opção `multipathing`, várias sub-redes devem ser usadas. Em outros sistemas operacionais, vários canais DNFS podem ser estabelecidos usando as `LOCAL` opções e `DONTRROUTE` para configurar vários canais DNFS em uma única sub-rede. No entanto, isso não funciona corretamente no Linux e problemas de desempenho inesperados podem resultar. Com o Linux, cada NIC usada para o tráfego DNFS deve estar em uma sub-rede diferente.

Programador de e/S.

O kernel Linux permite um controle de baixo nível sobre a maneira como e/S para bloquear dispositivos é agendada. Os padrões em várias distribuições do Linux variam consideravelmente. Testes mostram que o prazo geralmente oferece os melhores resultados, mas ocasionalmente o NOOP foi um pouco melhor. A diferença de desempenho é mínima, mas teste ambas as opções se for necessário extrair o máximo desempenho possível de uma configuração de banco de dados. O CFQ é o padrão em muitas configurações e demonstrou problemas significativos de desempenho com cargas de trabalho de banco de dados.

Consulte a documentação relevante do fornecedor do Linux para obter instruções sobre como configurar o agendador de e/S.

Multipathing

Alguns clientes encontraram falhas durante a interrupção da rede porque o daemon `multipath` não estava sendo executado em seu sistema. Em versões recentes do Linux, o processo de instalação do sistema operacional e do daemon `multipathing` podem deixar esses sistemas operacionais vulneráveis a esse problema. Os pacotes são instalados corretamente, mas não são configurados para inicialização automática após uma reinicialização.

Por exemplo, o padrão para o daemon `multipath` no RHEL5,5 pode aparecer da seguinte forma:


```
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off  1:off  2:off  3:off  4:off  5:off  6:off
```

Isso pode ser corrigido com os seguintes comandos:

```
[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
```

Espelhamento ASM

O espelhamento ASM pode exigir alterações nas configurações de multipath do Linux para permitir que o ASM reconheça um problema e alterne para um grupo de falhas alternativo. A maioria das configurações ASM no ONTAP usa redundância externa, o que significa que a proteção de dados é fornecida pelo array externo e o ASM não espelha dados. Alguns sites usam ASM com redundância normal para fornecer espelhamento bidirecional, normalmente em diferentes sites.

As configurações do Linux mostradas na ["Documentação dos utilitários de host do NetApp"](#) incluem parâmetros multipath que resultam em filas indefinidas de e/S. Isso significa que uma e/S em um dispositivo LUN sem caminhos ativos aguarda o tempo necessário para que a e/S seja concluída. Isso geralmente é desejável porque os hosts Linux esperam que as alterações de caminho SAN sejam concluídas, que os switches FC sejam reiniciados ou que um sistema de storage conclua um failover.

Esse comportamento ilimitado de enfileiramento causa um problema com o espelhamento ASM porque o ASM deve receber uma falha de e/S para que ele tente novamente e/S em um LUN alternativo.

Defina os seguintes parâmetros no arquivo Linux `multipath.conf` para LUNs ASM usados com espelhamento ASM:

```
polling_interval 5
no_path_retry 24
```

Essas configurações criam um tempo limite de 120 segundos para dispositivos ASM. O tempo limite é calculado como `polling_interval * no_path_retry` como segundos. O valor exato pode precisar ser ajustado em algumas circunstâncias, mas um tempo limite de 120 segundos deve ser suficiente para a maioria dos usos. Especificamente, 120 segundos devem permitir que uma tomada de controle ou giveback ocorra sem produzir um erro de e/S que resultaria em que o grupo de falha fosse colocado offline.

Um valor menor `no_path_retry` pode reduzir o tempo necessário para que o ASM alterne para um grupo de falhas alternativo, mas isso também aumenta o risco de um failover indesejado durante atividades de manutenção, como um controle de controle. O risco pode ser atenuado por um monitoramento cuidadoso do estado de espelhamento do ASM. Se ocorrer um failover indesejado, os espelhos podem ser ressynced rapidamente se a ressincronização for executada de forma relativamente rápida. Para obter informações adicionais, consulte a documentação Oracle sobre ASM Fast Mirror Resync para a versão do software Oracle em uso.

Opções de montagem Linux xfs, ext3 e ext4



A NetApp recomenda usando as opções de montagem padrão.

ASMLib/AFD (controlador de filtro ASM)

Tópicos de configuração específicos para o sistema operacional Linux usando AFD e ASMLib

Tamanhos de blocos ASMLib

ASMLib é uma biblioteca de gerenciamento ASM opcional e utilitários associados. Seu valor principal é a capacidade de carimbar um LUN ou um arquivo baseado em NFS como um recurso ASM com uma etiqueta legível por humanos.

Versões recentes do ASMLib detetam um parâmetro LUN chamado Logical Blocks per Physical Block exponent (LBPPBE). Esse valor não foi reportado pelo destino SCSI ONTAP até recentemente. Ele agora retorna um valor que indica que um tamanho de bloco 4KB é preferido. Esta não é uma definição de tamanho de bloco, mas é uma dica para qualquer aplicativo que usa LBPPBE que I/os de um determinado tamanho podem ser manipulados de forma mais eficiente. No entanto, o ASMLib interpreta LBPPBE como um tamanho de bloco e marca persistentemente o cabeçalho ASM quando o dispositivo ASM é criado.

Esse processo pode causar problemas com atualizações e migrações de várias maneiras, tudo com base na incapacidade de misturar dispositivos ASMLib com diferentes tamanhos de bloco no mesmo grupo de discos ASM.

Por exemplo, arrays mais antigos geralmente relataram um valor LBPPBE de 0 ou não relataram esse valor de todo. ASMLib interpreta isso como um tamanho de bloco de 512 bytes. Matrizes mais recentes seriam interpretadas como tendo um tamanho de bloco 4KB. Não é possível misturar dispositivos de 512 bytes e 4KB no mesmo grupo de discos ASM. Isso bloquearia um usuário de aumentar o tamanho do grupo de discos ASM usando LUNs de dois arrays ou utilizando ASM como uma ferramenta de migração. Em outros casos, o RMAN pode não permitir a cópia de arquivos entre um grupo de discos ASM com um tamanho de bloco de 512 bytes e um grupo de discos ASM com um tamanho de bloco de 4KBMB.

A solução preferida é corrigir o ASMLib. O ID do bug Oracle é 13999609, e o patch está presente no oracleasm-support-2, 1.8-1 e superior. Este patch permite que um usuário defina o parâmetro `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` como `true` no `/etc/sysconfig/oracleasm` arquivo de configuração. Isso impede que o ASMLib use o parâmetro LBPPBE, o que significa que os LUNs na nova matriz agora são reconhecidos como dispositivos de bloco de 512 bytes.



A opção não altera o tamanho do bloco em LUNs que foram previamente carimbados pelo ASMLib. Por exemplo, se um grupo de discos ASM com blocos de 512 bytes precisar ser migrado para um novo sistema de armazenamento que relata um bloco 4KB, a opção `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` deve ser definida antes que os novos LUNs sejam carimbados com ASMLib. Se os dispositivos já tiverem sido carimbados por oracleasm, eles devem ser reformatados antes de serem carimbados com um novo tamanho de bloco. Primeiro, desfigure o dispositivo com `oracleasm deletedisk`e`, em seguida, limpe os primeiros 1GB do dispositivo com ``dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`. Por fim, se o dispositivo tiver sido particionado anteriormente, use o `kpartx` comando para remover partições obsoletas ou simplesmente reiniciar o sistema operacional.

Se o ASMLib não puder ser corrigido, o ASMLib pode ser removido da configuração. Esta alteração é disruptiva

e requer a remoção de carimbo de discos ASM e certificar-se de que o `asm_diskstring` parâmetro está definido corretamente. No entanto, essa alteração não requer a migração de dados.

Tamanhos de bloco do Acionamento do filtro ASM (AFD)

O AFD é uma biblioteca de gerenciamento ASM opcional que está se tornando a substituição do ASMLib. Do ponto de vista do armazenamento, ele é muito semelhante ao ASMLib, mas inclui recursos adicionais, como a capacidade de bloquear e/S não-Oracle para reduzir as chances de erros de usuário ou aplicativo que poderiam corromper dados.

Tamanhos de bloco de dispositivos

Como o ASMLib, o AFD também lê o parâmetro LUN blocos lógicos por expoente de bloco físico (LBPPBE) e, por padrão, usa o tamanho do bloco físico, não o tamanho do bloco lógico.

Isso pode criar um problema se AFD for adicionado a uma configuração existente onde os dispositivos ASM já estejam formatados como dispositivos de bloco de 512 bytes. O driver AFD reconheceria o LUN como um dispositivo 4K e a incompatibilidade entre o rótulo ASM e o dispositivo físico impediria o acesso. Da mesma forma, as migrações seriam afetadas porque não é possível misturar dispositivos de 512 bytes e 4KB no mesmo grupo de discos ASM. Isso bloquearia um usuário de aumentar o tamanho do grupo de discos ASM usando LUNs de dois arrays ou utilizando ASM como uma ferramenta de migração. Em outros casos, o RMAN pode não permitir a cópia de arquivos entre um grupo de discos ASM com um tamanho de bloco de 512 bytes e um grupo de discos ASM com um tamanho de bloco de 4KBMB.

A solução é simples - AFD inclui um parâmetro para controlar se usa os tamanhos de blocos lógicos ou físicos. Este é um parâmetro global que afeta todos os dispositivos no sistema. Para forçar o AFD a usar o tamanho do bloco lógico, defina `options oracleafd oracleafd_use_logical_block_size=1` no `/etc/modprobe.d/oracleafd.conf` arquivo.

Tamanhos de transferência multipath

As recentes alterações do kernel do linux impõem restrições de tamanho de e/S enviadas para dispositivos multipath, e o AFD não honra essas restrições. Os I/os são então rejeitados, o que faz com que o caminho LUN fique offline. O resultado é uma incapacidade de instalar o Oracle Grid, configurar ASM ou criar um banco de dados.

A solução é especificar manualmente o comprimento máximo de transferência no arquivo `multipath.conf` para LUNs ONTAP:

```
devices {
    device {
        vendor "NETAPP"
        product "LUN.*"
        max_sectors_kb 4096
    }
}
```



Mesmo que não existam problemas atualmente, este parâmetro deve ser definido se AFD for usado para garantir que uma futura atualização do linux não cause problemas inesperadamente.

Microsoft Windows

Tópicos de configuração para banco de dados Oracle no Microsoft Windows com ONTAP.

NFS

A Oracle suporta o uso do Microsoft Windows com o cliente NFS direto. Esse recurso oferece um caminho para os benefícios de gerenciamento do NFS, incluindo a capacidade de exibir arquivos entre ambientes, redimensionar volumes dinamicamente e utilizar um protocolo IP menos caro. Consulte a documentação oficial da Oracle para obter informações sobre como instalar e configurar um banco de dados no Microsoft Windows usando DNFS. Não existem boas práticas especiais.

SAN

Para uma ótima eficiência de compressão, certifique-se de que o sistema de ficheiros NTFS utilize uma unidade de alocação de 8K GB ou maior. O uso de uma unidade de alocação 4K, que geralmente é o padrão, afeta negativamente a eficiência da compressão.

Solaris

Tópicos de configuração específicos do Solaris os.

Opções de montagem do Solaris NFS

A tabela a seguir lista as opções de montagem do Solaris NFS para uma única instância.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1], roto=tcp, timeo=600, rsize=262144, wsize=262144</code>
Registros do Redo ControlFiles Datafiles	<code>rw,bg,hard,[vers=3,vers=4.1], proto=tcp, timeo=600, rsize=262144, wsize=262144, nointr, llock, suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1], proto=tcp, timeo=600, rsize=262144, wsize=262144, suid</code>

Provou-se que o uso do `llock` melhora significativamente a performance nos ambientes dos clientes, eliminando a latência associada à aquisição e liberação de bloqueios no sistema de storage. Use essa opção com cuidado em ambientes nos quais vários servidores são configurados para montar os mesmos sistemas de arquivos e o Oracle está configurado para montar esses bancos de dados. Embora esta seja uma configuração altamente incomum, ela é usada por um pequeno número de clientes. Se uma instância for iniciada acidentalmente uma segunda vez, a corrupção de dados pode ocorrer porque a Oracle não consegue detetar os arquivos de bloqueio no servidor estrangeiro. Os bloqueios NFS não oferecem proteção de outra forma; como no NFS versão 3, eles são apenas consultivos.

Como `llock` os parâmetros e `forcedirectio` são mutuamente exclusivos, é importante que `filesystemio_options=setall` esteja presente no `init.ora` arquivo para que `directio` seja usado. Sem esse parâmetro, o cache do buffer do sistema operacional do host é usado e o desempenho pode ser afetado negativamente.

A tabela a seguir lista as opções de montagem do Solaris NFS RAC.

Tipo de ficheiro	Opções de montagem
ADR Home	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,noac</code>
Arquivos de controle arquivos de dados Redo logs	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio</code>
CRS/votação	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,forcedirectio</code>
Dedicado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,suid</code>
Compartilhado ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,suid</code>

A principal diferença entre as opções de montagem de uma única instância e RAC é a adição `noac` e `forcedirectio` às opções de montagem. Essa adição tem o efeito de desabilitar o cache do sistema operacional do host, o que permite que todas as instâncias do cluster RAC tenham uma visão consistente do estado dos dados. Embora o uso do `init.ora` parâmetro `filesystemio_options=setall` tenha o mesmo efeito de desabilitar o cache do host, ainda é necessário usar `noac` e `forcedirectio`.

O motivo `actimeo=0` necessário para implantações compartilhadas `ORACLE_HOME` é facilitar a consistência de arquivos, como arquivos de senha Oracle e `spfiles`. Se cada instância em um cluster RAC tiver um dedicado `ORACLE_HOME`, esse parâmetro não será necessário.

Opções de montagem do Solaris UFS

A NetApp recomenda fortemente o uso da opção de montagem de log para que a integridade dos dados seja preservada no caso de uma falha de host do Solaris ou a interrupção da conectividade FC. A opção de montagem de log também preserva a usabilidade dos backups Snapshot.

Solaris ZFS

O Solaris ZFS deve ser instalado e configurado cuidadosamente para oferecer o melhor desempenho.

mvector

O Solaris 11 incluiu uma mudança na forma como processa grandes operações de e/S, o que pode resultar em graves problemas de desempenho em matrizes de armazenamento SAN. O problema está documentado no relatório de bug de rastreamento do NetApp 630173, "regressão de desempenho do Solaris 11 ZFS".

Este não é um bug do ONTAP. É um defeito do Solaris que é rastreado sob os defeitos Solaris 7199305 e 7082975.

Você pode consultar o suporte Oracle para saber se sua versão do Solaris 11 é afetada ou testar a solução alternativa alterando `zfs_mvector_max_size` para um valor menor.

Você pode fazer isso executando o seguinte comando como root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t131072" |mdb -kw
```

Se surgir algum problema inesperado dessa alteração, ela pode ser facilmente revertida executando o seguinte comando como root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t1048576" |mdb -kw
```

Kernel

O desempenho confiável do ZFS requer um kernel Solaris corrigido contra problemas de alinhamento de LUN. A correção foi introduzida com o patch 147440-19 no Solaris 10 e com o SRU 10,5 para Solaris 11. Utilize apenas o Solaris 10 e posterior com o ZFS.

Configuração LUN

Para configurar um LUN, execute as seguintes etapas:

1. Crie um LUN do tipo `solaris`.
2. Instale o Kit de Utilitário do host (HUK) apropriado especificado pelo ["Ferramenta de Matriz de interoperabilidade NetApp \(IMT\)"](#).
3. Siga as instruções no HUK exatamente como descrito. Os passos básicos estão descritos abaixo, mas consulte o ["documentação mais recente"](#) para obter o procedimento adequado.
 - a. Execute o `host_config` utilitário para atualizar o `sd.conf/sdd.conf` arquivo. Isso permite que as unidades SCSI descubram corretamente LUNs ONTAP.
 - b. Siga as instruções dadas pelo `host_config` utilitário para ativar a entrada/saída multipath (MPIO).
 - c. Reinicie. Esta etapa é necessária para que quaisquer alterações sejam reconhecidas em todo o sistema.
4. Particione os LUNs e verifique se eles estão alinhados corretamente. Consulte o "Apêndice B: Verificação do alinhamento do WAFL" para obter instruções sobre como testar e confirmar diretamente o alinhamento.

zpool

Um zpool só deve ser criado após as etapas no ["Configuração LUN"](#) serem executadas. Se o procedimento não for feito corretamente, pode resultar em degradação grave do desempenho devido ao alinhamento de e/S. O desempenho ideal no ONTAP requer que a e/S seja alinhada a um limite de 4K mm numa unidade. Os sistemas de arquivos criados em um zpool usam um tamanho de bloco efetivo que é controlado por meio de um parâmetro `ashift` chamado , que pode ser visualizado executando o comando `zdb -C`.

O valor `ashift` padrão é 9, o que significa 2⁹, ou 512 bytes. Para um desempenho ideal, o `ashift` valor deve ser 12 (2¹² 4K). Esse valor é definido no momento em que o zpool é criado e não pode ser alterado, o que significa que os dados em zpool com `ashift` outros que não 12 devem ser migrados copiando dados para um zpool recém-criado.

Depois de criar um zpool, verifique o valor de `ashift` antes de continuar. Se o valor não for 12, os LUNs não foram detetados corretamente. Destrua o zpool, verifique se todas as etapas mostradas na documentação relevante dos Utilitários do host foram executadas corretamente e recrie o zpool.

Zpools e Solaris LDOMs

Os Solaris LDOMs criam um requisito adicional para garantir que o alinhamento de e/S esteja correto. Embora um LUN possa ser encontrado corretamente como um dispositivo 4K, um dispositivo vdisk virtual em um LDOM não herda a configuração do domínio de e/S. O vdisk baseado nesse LUN retorna para um bloco de 512 bytes.

É necessário um ficheiro de configuração adicional. Primeiro, os LDOM individuais devem ser corrigidos para o bug Oracle 15824910 para habilitar as opções de configuração adicionais. Este patch foi portado para todas as versões usadas atualmente do Solaris. Uma vez que o LDOM é corrigido, ele está pronto para a configuração dos novos LUNs corretamente alinhados da seguinte forma:

1. Identifique o LUN ou LUNs a serem usados no novo zpool. Neste exemplo, é o dispositivo c2d1.

```
[root@LDM1 ~]# 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. Recupere a instância vdc dos dispositivos a serem usados para um pool ZFS:

```
[root@LDM1 ~]# 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. `/platform/sun4v/kernel/drv/vdc.conf` Editar :

```
block-size-list="1:4096";
```

Isso significa que a instância 1 do dispositivo recebe um tamanho de bloco de 4096MB.

Como exemplo adicional, suponha que as instâncias 1 a 6 do vdisk precisem ser configuradas para um tamanho de bloco 4K e `/etc/path_to_inst` lerem da seguinte forma:

```
"/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. O arquivo final `vdc.conf` deve conter o seguinte:

```
block-size-list="1:8192","2:8192","3:8192","4:8192","5:8192","6:8192";
```

Cuidado

O LDOM deve ser reinicializado depois que o `vdc.conf` é configurado e o `vdisk` é criado. Este passo não pode ser evitado. A alteração do tamanho do bloco só entra em vigor após uma reinicialização. prossiga com a configuração do `zpool` e certifique-se de que o `ashift` está corretamente configurado para 12, conforme descrito anteriormente.

Registro intenção ZFS (ZIL)

Geralmente, não há razão para localizar o ZFS Intent Log (ZIL) em um dispositivo diferente. O log pode compartilhar espaço com a piscina principal. O uso principal de um ZIL separado é quando se usa unidades físicas que não têm os recursos de armazenamento em cache de gravação em arrays de armazenamento modernos.

logbias

Defina `logbias` o parâmetro em sistemas de arquivos ZFS que hospedam dados Oracle.

```
zfs set logbias=throughput <filesystem>
```

O uso desse parâmetro reduz os níveis gerais de gravação. Sob os padrões, os dados escritos são comprometidos primeiro com o ZIL e depois para o pool de armazenamento principal. Essa abordagem é apropriada para uma configuração usando uma configuração de unidade simples, que inclui um dispositivo ZIL baseado em SSD e Mídia giratória para o pool de armazenamento principal. Isso ocorre porque permite que um commit ocorra em uma única transação de e/S na Mídia de menor latência disponível.

Ao usar um storage array moderno que inclua sua própria funcionalidade de armazenamento em cache, essa abordagem geralmente não é necessária. Em circunstâncias raras, pode ser desejável submeter uma

gravação com uma única transação ao log, como uma carga de trabalho que consiste em gravações aleatórias altamente concentradas e sensíveis à latência. Há consequências na forma de amplificação de gravação porque os dados registrados são gravados no pool de armazenamento principal, resultando em uma duplicação da atividade de gravação.

E/S direta

Muitas aplicações, incluindo produtos Oracle, podem ignorar o cache de buffer do host habilitando a e/S direta. Esta estratégia não funciona como esperado com sistemas de arquivos ZFS. Embora o cache do buffer do host seja ignorado, o próprio ZFS continua a armazenar dados em cache. Essa ação pode resultar em resultados enganosos ao usar ferramentas como fio ou sio para realizar testes de desempenho, pois é difícil prever se a e/S está chegando ao sistema de armazenamento ou se está sendo armazenada em cache localmente no sistema operacional. Essa ação também torna muito difícil usar esses testes sintéticos para comparar o desempenho do ZFS com outros sistemas de arquivos. Na prática, há pouca ou nenhuma diferença no desempenho do sistema de arquivos em workloads reais do usuário.

Vários zpool

Backups, restaurações, clones e arquivamento baseados em snapshot de dados baseados em ZFS devem ser executados no nível do zpool e, geralmente, exigem vários zpools. Um zpool é análogo a um grupo de discos LVM e deve ser configurado usando as mesmas regras. Por exemplo, um banco de dados provavelmente é melhor definido com os datafiles que residem em `zpool1` e os logs de arquivo, arquivos de controle e logs de refazer que residem em `zpool2`. Essa abordagem permite um hot backup padrão no qual o banco de dados é colocado no modo hot backup, seguido por um snapshot de `zpool1`. O banco de dados é então removido do modo hot backup, o arquivo de log é forçado e um snapshot de `zpool2` é criado. Uma operação de restauração requer a desmontagem dos sistemas de arquivos zfs e a remoção do zpool em sua totalidade, seguindo-se uma operação de restauração do SnapRestore. O zpool pode então ser colocado on-line novamente e o banco de dados recuperado.

sistema de arquivos_options

O parâmetro Oracle `filesystemio_options` funciona de forma diferente com o ZFS. Se `setall` ou `directio` for usado, as operações de gravação são síncronas e ignoram o cache do buffer do sistema operacional, mas as leituras são armazenadas em buffer pelo ZFS. Essa ação causa dificuldades na análise de desempenho, pois às vezes, a e/S é interceptada e atendida pelo cache ZFS, tornando a latência do armazenamento e a e/S total menos do que parece ser.

Configuração de rede

Interfaces lógicas

Os bancos de dados Oracle precisam de acesso ao armazenamento. Interfaces lógicas (LIFs) são o encanamento de rede que conecta uma máquina virtual de armazenamento (SVM) à rede e, por sua vez, ao banco de dados. O design de LIF adequado é necessário para garantir que haja largura de banda suficiente para cada workload de banco de dados, e o failover não resulta em perda de serviços de storage.

Esta seção fornece uma visão geral dos principais princípios de design de LIF. Para obter documentação mais abrangente, consulte o "[Documentação do Gerenciamento de rede da ONTAP](#)". Assim como em outros aspectos da arquitetura de banco de dados, as melhores opções para o design de máquina virtual de storage (SVM, conhecido como vserver na CLI) e interface lógica (LIF) dependem muito dos requisitos de dimensionamento e necessidades de negócios.

Considere os seguintes tópicos principais ao criar uma estratégia de LIF:

- **Desempenho.** A largura de banda da rede é suficiente?
- **Resiliência.** Há algum ponto único de falha no projeto?
- **Capacidade de gerenciamento.** A rede pode ser dimensionada sem interrupções?

Esses tópicos se aplicam à solução completa, desde o host até os switches até o sistema de storage.

Tipos de LIF

Existem vários tipos de LIF. "[Documentação do ONTAP sobre tipos de LIF](#)" Forneça informações mais completas sobre este tópico, mas de uma perspectiva funcional os LIFs podem ser divididos nos seguintes grupos:

- **LIFs de gerenciamento de clusters e nós.** LIFs usadas para gerenciar o cluster de armazenamento.
- **LIFs de gerenciamento de SVM.** Interfaces que permitem o acesso a um SVM por meio da API REST ou ONTAPI (também conhecida como ZAPI) para funções como criação de snapshot ou redimensionamento de volume. Produtos como o SnapManager para Oracle (SMO) precisam ter acesso a um LIF de gerenciamento de SVM.
- **LIFs de dados.** Interfaces para dados FC, iSCSI, NVMe/FC, NVMe/TCP, NFS ou SMB/CIFS.



Um LIF de dados usado para tráfego NFS também pode ser usado para gerenciamento alterando a política de firewall de `data` para `mgmt` ou outra política que permita HTTP, HTTPS ou SSH. Essa alteração pode simplificar a configuração de rede evitando a configuração de cada host para acesso ao LIF de dados NFS e a um LIF de gerenciamento separado. Não é possível configurar uma interface para iSCSI e tráfego de gerenciamento, apesar do fato de ambos usarem um protocolo IP. Um LIF de gerenciamento separado é necessário em ambientes iSCSI.

Design de SAN LIF

O design de LIF em um ambiente SAN é relativamente simples por um motivo: Multipathing. Todas as implementações de SAN modernas permitem que um cliente acesse dados em vários caminhos de rede independentes e selecione o melhor caminho ou caminhos para acesso. Como resultado, o desempenho em relação ao design de LIF é mais simples de abordar, pois os clientes SAN equilibram automaticamente a e/S de carga nos melhores caminhos disponíveis.

Se um caminho ficar indisponível, o cliente selecionará automaticamente um caminho diferente. A simplicidade resultante do design torna os LIFs SAN geralmente mais gerenciáveis. Isso não significa que um ambiente SAN sempre seja gerenciado com mais facilidade, porque há muitos outros aspectos do storage SAN que são muito mais complicados do que o NFS. Isso significa simplesmente que o design de SAN LIF é mais fácil.

Desempenho

A consideração mais importante com o desempenho de LIF em um ambiente SAN é a largura de banda. Por exemplo, um cluster ONTAP AFF de dois nós com duas portas FC de 16GB GB por nó permite até 32GB Gbps de largura de banda de/para cada nó.

Resiliência

OS LIFs DE SAN não fazem failover em um sistema de storage da AFF. Se um LIF SAN falhar devido ao failover de controladora, o software multipathing do cliente detecta a perda de um caminho e redireciona e/S

para um LIF diferente. Com os sistemas de armazenamento ASA, os LIFs serão falhados após um curto atraso, mas isso não interrompe a e/S porque já existem caminhos ativos no outro controlador. O processo de failover ocorre para restaurar o acesso do host em todas as portas definidas.

Capacidade de gerenciamento

A migração de LIF é uma tarefa muito mais comum em um ambiente NFS, pois a migração de LIF geralmente é associada à realocação de volumes no cluster. Não é necessário migrar um LIF em um ambiente SAN quando os volumes são transferidos para o par de HA. Isso ocorre porque, após a conclusão da movimentação de volume, o ONTAP envia uma notificação à SAN sobre uma alteração nos caminhos e os clientes SAN reotimizam automaticamente. A migração de LIF com SAN está associada principalmente a grandes alterações de hardware físico. Por exemplo, se uma atualização sem interrupções dos controladores for necessária, um LIF SAN é migrado para o novo hardware. Se uma porta FC estiver defeituosa, um LIF pode ser migrado para uma porta não utilizada.

Recomendações de design

A NetApp faz as seguintes recomendações:

- Não crie mais caminhos do que o necessário. O número excessivo de caminhos torna o gerenciamento geral mais complicado e pode causar problemas com failover de caminho em alguns hosts. Além disso, alguns hosts têm limitações de caminho inesperadas para configurações como inicialização por SAN.
- Poucas configurações devem exigir mais de quatro caminhos para um LUN. O valor de ter mais de dois nós anunciando caminhos para LUNs é limitado porque o host agregado de um LUN fica inacessível se o nó proprietário do LUN e seu parceiro de HA falhar. A criação de caminhos em nós que não o par de HA principal não é útil em tal situação.
- Embora o número de caminhos de LUN visíveis possa ser gerenciado selecionando quais portas estão incluídas nas zonas FC, geralmente é mais fácil incluir todos os pontos de destino potenciais na zona FC e controlar a visibilidade de LUN no nível ONTAP.
- No ONTAP 8,3 e posterior, o recurso de mapeamento de LUN seletivo (SLM) é o padrão. Com o SLM, qualquer novo LUN é automaticamente anunciado a partir do nó que é proprietário do agregado subjacente e do parceiro de HA do nó. Esse arranjo evita a necessidade de criar conjuntos de portas ou configurar o zoneamento para limitar a acessibilidade de portas. Cada LUN está disponível no número mínimo de nós necessário para ter a melhor performance e resiliência. *Caso um LUN precise ser migrado para fora dos dois controladores, os nós adicionais podem ser adicionados com o `lun mapping add-reporting-nodes` comando para que os LUNs sejam anunciados nos novos nós. Isso cria caminhos SAN adicionais para os LUNs para migração de LUN. No entanto, o host deve executar uma operação de descoberta para usar os novos caminhos.
- Não se preocupe excessivamente com o tráfego indireto. É melhor evitar o tráfego indireto em um ambiente com uso intenso de e/S para o qual cada microssegundo de latência é crítico, mas o efeito de performance visível é insignificante para workloads típicos.

Design de LIF de NFS

Em contraste com os protocolos SAN, o NFS tem uma capacidade limitada de definir vários caminhos para os dados. As extensões NFS paralelas (pNFS) para NFSv4 abordam essa limitação, mas como as velocidades ethernet atingiram 100GBMbps e além raramente há valor na adição de caminhos adicionais.

Performance e resiliência

Embora a medição do desempenho do SAN LIF seja principalmente uma questão de calcular a largura de banda total de todos os caminhos principais, determinar o desempenho do NFS LIF requer uma visão mais detalhada da configuração exata da rede. Por exemplo, duas portas 10Gb podem ser configuradas como

portas físicas brutas ou podem ser configuradas como um grupo de interfaces LACP (Link Aggregation Control Protocol). Se eles estiverem configurados como um grupo de interfaces, várias políticas de balanceamento de carga estarão disponíveis que funcionam de forma diferente, dependendo se o tráfego é comutado ou roteado. Por fim, o Oracle Direct NFS (DNFS) oferece configurações de balanceamento de carga que não existem em nenhum cliente NFS do sistema operacional no momento.

Diferentemente dos protocolos SAN, os sistemas de arquivos NFS exigem resiliência na camada do protocolo. Por exemplo, um LUN é sempre configurado com multipathing habilitado, o que significa que vários canais redundantes estão disponíveis para o sistema de storage, cada um dos quais usa o protocolo FC. Um sistema de arquivos NFS, por outro lado, depende da disponibilidade de um único canal TCP/IP que só pode ser protegido na camada física. Esse arranjo é o motivo pelo qual existem opções como failover de portas e agregação de portas LACP.

Em um ambiente NFS, a performance e a resiliência são fornecidas na camada de protocolo de rede. Como resultado, ambos os tópicos estão interligados e devem ser discutidos juntos.

Vincular LIFs a grupos de portas

Para vincular um LIF a um grupo de portas, associe o endereço IP de LIF a um grupo de portas físicas. O método principal para agregar portas físicas em conjunto é o LACP. A capacidade de tolerância a falhas do LACP é bastante simples; cada porta em um grupo LACP é monitorada e removida do grupo de portas em caso de mau funcionamento. Existem, no entanto, muitos equívocos sobre como o LACP funciona em relação ao desempenho:

- O LACP não requer a configuração no switch para corresponder ao endpoint. Por exemplo, o ONTAP pode ser configurado com balanceamento de carga baseado em IP, enquanto um switch pode usar balanceamento de carga baseado em MAC.
- Cada endpoint que usa uma conexão LACP pode escolher independentemente a porta de transmissão de pacotes, mas não pode escolher a porta usada para recebimento. Isso significa que o tráfego de ONTAP para um destino específico está vinculado a uma porta específica, e o tráfego de retorno pode chegar em uma interface diferente. No entanto, isso não causa problemas.
- O LACP não distribui uniformemente o tráfego o tempo todo. Em um ambiente grande com muitos clientes NFS, o resultado é normalmente até mesmo o uso de todas as portas em uma agregação LACP. No entanto, qualquer sistema de arquivos NFS no ambiente é limitado à largura de banda de apenas uma porta, e não a agregação inteira.
- Embora as políticas LACP robin estejam disponíveis no ONTAP, essas políticas não abordam a conexão de um switch para um host. Por exemplo, uma configuração com um tronco LACP de quatro portas em um host e um tronco LACP de quatro portas no ONTAP ainda é capaz de ler apenas um sistema de arquivos usando uma única porta. Embora o ONTAP possa transmitir dados através das quatro portas, não há tecnologias de switch disponíveis que sejam enviadas do switch para o host através das quatro portas. Apenas um é usado.

A abordagem mais comum em ambientes maiores que consistem em muitos hosts de banco de dados é construir um agregado LACP de um número apropriado de interfaces 10Gb (ou mais rápidas) usando o balanceamento de carga IP. Essa abordagem permite que o ONTAP forneça uso uniforme de todas as portas, desde que existam clientes suficientes. O balanceamento de carga é interrompido quando há menos clientes na configuração porque o entroncamento LACP não redistribui dinamicamente a carga.

Quando uma conexão é estabelecida, o tráfego em uma determinada direção é colocado em apenas uma porta. Por exemplo, um banco de dados que executa uma verificação de tabela completa em um sistema de arquivos NFS conectado por meio de um tronco LACP de quatro portas lê dados através de apenas uma placa de interface de rede (NIC). Se apenas três servidores de banco de dados estiverem em tal ambiente, é possível que todos os três estejam lendo da mesma porta, enquanto as outras três portas estiverem ociosas.

Vincule LIFs a portas físicas

Vincular um LIF a uma porta física resulta em um controle mais granular sobre a configuração de rede, pois um determinado endereço IP em um sistema ONTAP está associado a apenas uma porta de rede de cada vez. A resiliência é então realizada por meio da configuração de grupos de failover e políticas de failover.

Políticas de failover e grupos de failover

O comportamento dos LIFs durante a interrupção da rede é controlado por políticas de failover e grupos de failover. As opções de configuração foram alteradas com as diferentes versões do ONTAP. Consulte o ["Documentação de gerenciamento de rede ONTAP para grupos e políticas de failover"](#) para obter detalhes específicos sobre a versão do ONTAP que está sendo implantado.

O ONTAP 8,3 e superior permitem o gerenciamento de failover de LIF com base em domínios de broadcast. Portanto, um administrador pode definir todas as portas que têm acesso a uma determinada sub-rede e permitir que o ONTAP selecione um LIF de failover apropriado. Essa abordagem pode ser usada por alguns clientes, mas tem limitações em um ambiente de rede de storage de alta velocidade devido à falta de previsibilidade. Por exemplo, um ambiente pode incluir ambas as portas 1GB para acesso de rotina ao sistema de arquivos e portas 10Gb para e/S de arquivo de dados. Se ambos os tipos de portas existirem no mesmo domínio de broadcast, o failover de LIF pode resultar na movimentação de e/S de um arquivo de dados de uma porta 10Gb para uma porta 1GB.

Em resumo, considere as seguintes práticas:

1. Configurar um grupo de failover conforme definido pelo usuário.
2. Preencha o grupo de failover com portas na controladora de parceiro de failover de storage (SFO) para que as LIFs sigam os agregados durante um failover de storage. Isso evita a criação de tráfego indireto.
3. Use portas de failover com características de desempenho correspondentes ao LIF original. Por exemplo, um LIF em uma única porta 10Gb física deve incluir um grupo de failover com uma única porta 10Gb. Um LIF LACP de quatro portas deve falhar para outro LIF LACP de quatro portas. Essas portas seriam um subconjunto das portas definidas no domínio de broadcast.
4. Defina a política de failover como somente parceiro SFO. Isso garante que o LIF siga o agregado durante o failover.

Reversão automática

Defina `auto-revert` o parâmetro conforme desejado. A maioria dos clientes prefere definir este parâmetro para que `true` o LIF reverta para sua porta inicial. No entanto, em alguns casos, os clientes definiram isso como "falso" que um failover inesperado pode ser investigado antes de retornar um LIF à sua porta inicial.

Relação LIF-volume

Um equívoco comum é que deve haver uma relação do 1:1 entre volumes e LIFs NFS. Embora essa configuração seja necessária para mover um volume em qualquer lugar em um cluster, sem nunca criar tráfego de interconexão adicional, ela não é categoricamente um requisito. O tráfego entre clusters deve ser considerado, mas a mera presença de tráfego entre clusters não cria problemas. Muitos dos benchmarks publicados criados para o ONTAP incluem predominantemente I/O. indireto

Por exemplo, um projeto de banco de dados contendo um número relativamente pequeno de bancos de dados críticos ao desempenho que exigiam apenas um total de 40 volumes pode garantir um volume 1:1 para a estratégia LIF, um arranjo que exigiria 40 endereços IP. Qualquer volume poderia então ser movido para qualquer lugar do cluster junto com o LIF associado, e o tráfego sempre seria direto, minimizando cada fonte de latência, mesmo nos níveis de microssegundos.

Como um exemplo de contador, um ambiente grande e hospedado pode ser mais facilmente gerenciado com um relacionamento 1:1 entre clientes e LIFs. Com o tempo, um volume pode precisar ser migrado para um nó diferente, o que causaria algum tráfego indireto. No entanto, o efeito de desempenho deve ser indetetável, a menos que as portas de rede no switch de interconexão estejam saturando. Se houver problema, um novo LIF pode ser estabelecido em nós adicionais e o host pode ser atualizado na próxima janela de manutenção para remover o tráfego indireto da configuração.

Configuração TCP/IP e ethernet

Muitos clientes do Oracle no ONTAP usam ethernet, o protocolo de rede de NFS, iSCSI, NVMe/TCP e, especialmente, a nuvem.

Configurações do sistema operacional do host

A maioria da documentação do fornecedor de aplicativos inclui configurações específicas de TCP e ethernet destinadas a garantir que o aplicativo esteja funcionando de forma ideal. Essas mesmas configurações geralmente são suficientes para oferecer também um desempenho ideal de storage baseado em IP.

Controle de fluxo Ethernet

Esta tecnologia permite que um cliente solicite que um remetente pare temporariamente a transmissão de dados. Isso geralmente é feito porque o recetor não consegue processar os dados de entrada com rapidez suficiente. Ao mesmo tempo, solicitar que um remetente cessasse a transmissão era menos disruptivo do que ter um recetor descartar pacotes porque os buffers estavam cheios. Este não é mais o caso com as pilhas TCP usadas nos sistemas operacionais atuais. Na verdade, o controle de fluxo causa mais problemas do que resolve.

Os problemas de desempenho causados pelo controle de fluxo Ethernet têm aumentado nos últimos anos. Isso ocorre porque o controle de fluxo Ethernet opera na camada física. Se uma configuração de rede permitir que qualquer sistema operacional host envie uma solicitação de controle de fluxo Ethernet para um sistema de armazenamento, o resultado será uma pausa em e/S para todos os clientes conetados. Como um número crescente de clientes é servido por um único controlador de armazenamento, a probabilidade de um ou mais desses clientes enviar solicitações de controle de fluxo aumenta. O problema tem sido visto frequentemente em sites de clientes com ampla virtualização de SO.

Uma NIC em um sistema NetApp não deve receber solicitações de controle de fluxo. O método utilizado para alcançar este resultado varia consoante o fabricante do comutador de rede. Na maioria dos casos, o controle de fluxo em um switch Ethernet pode ser definido como `receive desired receive on` ou `,`, o que significa que uma solicitação de controle de fluxo não é encaminhada para o controlador de armazenamento. Em outros casos, a conexão de rede no controlador de armazenamento pode não permitir a desativação do controle de fluxo. Nesses casos, os clientes devem ser configurados para nunca enviar solicitações de controle de fluxo, seja alterando para a configuração da NIC no próprio servidor host ou para as portas de switch às quais o servidor host está conetado.



A NetApp recomenda certificando-se de que os controladores de armazenamento NetApp não recebam pacotes de controle de fluxo Ethernet. Isso geralmente pode ser feito definindo as portas do switch às quais o controlador está conetado, mas alguns hardwares do switch têm limitações que podem exigir alterações no lado do cliente.

Tamanhos de MTU

O uso de quadros jumbo tem sido mostrado para oferecer alguma melhoria de desempenho em redes 1GBG, reduzindo a sobrecarga de CPU e rede, mas o benefício geralmente não é significativo.



A NetApp recomenda a implementação de quadros jumbo quando possível, tanto para obter quaisquer benefícios potenciais de desempenho quanto para preparar a solução para o futuro.

Usar quadros jumbo em uma rede 10GbG é quase obrigatório. Isso ocorre porque a maioria das implementações 10Gb atingem um limite de pacotes por segundo sem quadros jumbo antes que atinjam a marca 10Gb. O uso de quadros jumbo melhora a eficiência no processamento TCP/IP porque permite que o sistema operacional, servidor, NICs e o sistema de armazenamento processem menos pacotes, mas maiores. A melhoria do desempenho varia de NIC para NIC, mas é significativa.

Para implementações de quadros jumbo, existe a crença comum, mas incorreta, de que todos os dispositivos conectados devem suportar quadros jumbo e que o tamanho da MTU deve corresponder de ponta a ponta. Em vez disso, os dois pontos de extremidade da rede negociam o tamanho de quadro mais alto mutuamente aceitável ao estabelecer uma conexão. Em um ambiente típico, um switch de rede é definido para um tamanho MTU de 9216, o controlador NetApp é definido como 9000 e os clientes são definidos como uma combinação de 9000 e 1514. Os clientes que podem suportar um MTU de 9000 podem usar quadros jumbo, e os clientes que só podem suportar 1514 podem negociar um valor mais baixo.

Problemas com este arranjo são raros em um ambiente completamente comutado. No entanto, tenha cuidado em um ambiente roteado que nenhum roteador intermediário é forçado a fragmentar quadros jumbo.

A NetApp recomenda configurando o seguinte:



- Os quadros jumbo são desejáveis, mas não são necessários com Ethernet de 1GB GB (GbE).
- Os quadros Jumbo são necessários para o máximo desempenho com 10GbE e mais rápido.

Parâmetros TCP

Três configurações geralmente são mal configuradas: Carimbos de data/hora TCP, reconhecimento seletivo (SACK) e escala de janela TCP. Muitos documentos desatualizados na Internet recomendam desativar um ou mais desses parâmetros para melhorar o desempenho. Havia algum mérito a esta recomendação há muitos anos, quando os recursos da CPU eram muito menores e havia um benefício para reduzir a sobrecarga no processamento TCP sempre que possível.

No entanto, com os sistemas operacionais modernos, desabilitar qualquer um desses recursos TCP geralmente resulta em nenhum benefício detectável, ao mesmo tempo em que potencialmente danifica o desempenho. Os danos no desempenho são especialmente prováveis em ambientes de rede virtualizados, pois esses recursos são necessários para o manuseio eficiente da perda de pacotes e alterações na qualidade da rede.



A NetApp recomenda a ativação de timestamps TCP, SACK e escala de janelas TCP no host, e todos esses três parâmetros devem estar ativados por padrão em qualquer sistema operacional atual.

Configuração de FC SAN

A configuração do FC SAN para bancos de dados Oracle trata principalmente de seguir as práticas recomendadas diárias de SAN.

Isso inclui medidas de Planejamento típicas, como garantir a existência de largura de banda suficiente na SAN entre o host e o sistema de storage, verificar se todos os caminhos de SAN existem entre todos os

dispositivos necessários, usar as configurações de porta FC exigidas pelo fornecedor do switch FC, evitar a contenção de ISL e usar o monitoramento adequado da malha SAN.

Zoneamento

Uma zona FC nunca deve conter mais de um iniciador. Tal arranjo pode parecer funcionar inicialmente, mas a conversa cruzada entre iniciadores eventualmente interfere com o desempenho e a estabilidade.

As zonas de destino múltiplo são geralmente consideradas seguras, embora em raras circunstâncias o comportamento das portas de destino FC de diferentes fornecedores tenha causado problemas. Por exemplo, evite incluir as portas de destino de um storage array NetApp e não NetApp na mesma zona. Além disso, é ainda mais provável que colocar um sistema de storage NetApp e um dispositivo de fita na mesma zona cause problemas.

Rede de conexão direta

Às vezes, os administradores de storage preferem simplificar suas infraestruturas removendo switches de rede da configuração. Isso pode ser suportado em alguns cenários.

ISCSI e NVMe/TCP

Um host usando iSCSI ou NVMe/TCP pode ser conectado diretamente a um sistema de storage e operar normalmente. A razão é pathing. Conexões diretas a dois controladores de storage diferentes resultam em dois caminhos independentes para o fluxo de dados. A perda de caminho, porta ou controlador não impede que o outro caminho seja usado.

NFS

O armazenamento NFS com conexão direta pode ser usado, mas com uma limitação significativa - o failover não funcionará sem um esforço significativo de script, o que seria da responsabilidade do cliente.

O motivo pelo qual o failover sem interrupções é complicado com o storage NFS com conexão direta é o roteamento que ocorre no sistema operacional local. Por exemplo, suponha que um host tenha um endereço IP de 192.168.1.1/24 e esteja conectado diretamente a um controlador ONTAP com um endereço IP de 192.168.1.50/24. Durante o failover, esse endereço 192.168.1.50 pode fazer failover para a outra controladora e estará disponível para o host, mas como o host detecta sua presença? O endereço 192.168.1.1 original ainda existe na NIC host que não se conecta mais a um sistema operacional. O tráfego destinado a 192.168.1.50 continuaria a ser enviado para uma porta de rede inoperável.

A segunda NIC do SO poderia ser configurada como 192.168.1.2 e seria capaz de se comunicar com o endereço 192.168.1.50 com falha, mas as tabelas de roteamento local teriam um padrão de usar um endereço **e apenas um** para se comunicar com a sub-rede 192.168.1.0/24. Um sysadmin poderia criar uma estrutura de script que detectaria uma conexão de rede com falha e alteraria as tabelas de roteamento local ou colocaria interfaces para cima e para baixo. O procedimento exato dependeria do SO em uso.

Na prática, os clientes da NetApp têm NFS com conexão direta, mas normalmente apenas para workloads em que as pausas de e/S durante failovers são aceitáveis. Quando os suportes rígidos são usados, não deve haver nenhum erro de e/S durante essas pausas. O IO deve travar até que os serviços sejam restaurados, seja por um failback ou intervenção manual para mover endereços IP entre NICs no host.

Conexão direta FC

Não é possível conectar diretamente um host a um sistema de storage ONTAP usando o protocolo FC. A razão é o uso de NPIV. A WWN que identifica uma porta ONTAP FC para a rede FC usa um tipo de virtualização chamado NPIV. Qualquer dispositivo conectado a um sistema ONTAP deve ser capaz de reconhecer um NPIV WWN. Não há fornecedores atuais de HBA que ofereçam um HBA que possa ser instalado em um host que possa suportar um destino NPIV.

Configuração de armazenamento

FC SAN

Alinhamento LUN

Alinhamento LUN refere-se a otimizar e/S em relação ao layout do sistema de arquivos subjacente.

Diretiva não resolvida no <stdin> - incluir::lun-alignment.adoc[]

Consulte também a discussão sobre o alinhamento do bloco de compressão na "[Eficiência](#)" seção . Qualquer layout que esteja alinhado com os limites do bloco de compressão 8KBD também está alinhado com os limites 4KBD.

Avisos de desalinhamento

Diretiva não resolvida no <stdin> - include::database-misalignment-warnings.adoc[]

O alinhamento em ambientes Solaris é mais complicado. "[Configuração do host SAN ONTAP](#)" Consulte para obter mais informações.

Cuidado

Nos ambientes Solaris x86, tenha cuidado adicional com o alinhamento adequado, pois a maioria das configurações tem várias camadas de partições. Os cortes de partição do Solaris x86 geralmente existem em cima de uma tabela de partição de Registro de inicialização principal padrão.

Dimensionamento de LUN e contagem de LUN

Selecionar o tamanho ideal de LUN e o número de LUNs a serem usados é essencial para obter o desempenho e a capacidade de gerenciamento ideais dos bancos de dados Oracle.

Um LUN é um objeto virtualizado no ONTAP que existe em todas as unidades no agregado de hospedagem. Como resultado, o desempenho do LUN não é afetado pelo seu tamanho porque o LUN se baseia no potencial de desempenho total do agregado, independentemente do tamanho escolhido.

Por uma questão de conveniência, os clientes podem querer usar um LUN de um tamanho específico. Por exemplo, se um banco de dados for construído em um grupo de discos LVM ou Oracle ASM composto por dois LUNs de 1TB cada, esse grupo de discos deve ser aumentado em incrementos de 1TB. Pode ser preferível construir o grupo de discos a partir de oito LUNs de 500GB cada, para que o grupo de discos possa ser aumentado em incrementos menores.

A prática de estabelecer um tamanho de LUN padrão universal é desencorajada porque isso pode complicar a

capacidade de gerenciamento. Por exemplo, um tamanho de LUN padrão de 100GB pode funcionar bem quando um banco de dados ou datastore está no intervalo de 1TB a 2TB, mas um banco de dados ou datastore de 20TB GB de tamanho exigiria 200 LUNs. Isso significa que os tempos de reinicialização do servidor são mais longos, há mais objetos para gerenciar nas várias UIs e produtos como SnapCenter devem executar a descoberta em muitos objetos. O uso de menos LUNs maiores evita esses problemas.

- A contagem de LUN é mais importante do que o tamanho do LUN.
- O tamanho do LUN é controlado principalmente pelos requisitos de contagem de LUN.
- Evite criar mais LUNs do que o necessário.

Contagem de LUN

Ao contrário do tamanho do LUN, a contagem de LUN afeta o desempenho. O desempenho da aplicação geralmente depende da capacidade de executar e/S paralela pela camada SCSI. Como resultado, dois LUNs oferecem melhor performance do que um único LUN. Usar um LVM como Veritas VxVM, Linux LVM2 ou Oracle ASM é o método mais simples para aumentar o paralelismo.

Os clientes da NetApp geralmente experimentaram o mínimo de benefícios ao aumentar o número de LUNs além de dezasseis, embora o teste de ambientes 100% SSD com e/S aleatória muito pesada tenha demonstrado melhorias adicionais em até 64 LUNs.

A NetApp recomenda o seguinte:



Em geral, quatro a dezesseis LUNs são suficientes para atender às necessidades de e/S de qualquer workload de banco de dados. Menos de quatro LUNs podem criar limitações de desempenho devido às limitações nas implementações de SCSI de host.

Colocação de LUN

O posicionamento ideal de LUNs de banco de dados nos volumes do ONTAP depende principalmente de como vários recursos do ONTAP serão usados.

Volumes

Um ponto comum de confusão com os clientes novos no ONTAP é o uso de FlexVols, comumente chamado de simplesmente "volumes".

Um volume não é um LUN. Esses termos são usados de forma sinônima com muitos outros produtos de fornecedores, incluindo provedores de nuvem. O ONTAP volumes é simplesmente volumes de gerenciamento. Eles não servem dados por si mesmos, nem ocupam espaço. Eles são contentores para arquivos ou LUNs e existem para melhorar e simplificar a capacidade de gerenciamento, especialmente em escala.

Volumes e LUNs

Os LUNs relacionados normalmente são colocados em um único volume. Por exemplo, um banco de dados que requer 10 LUNs normalmente teria todos os 10 LUNs colocados no mesmo volume.



- Usar uma proporção de 1:1:1, de LUNs para volumes, ou seja, um LUN por volume, é **não** uma prática recomendada formal.
- Em vez disso, os volumes devem ser vistos como contêineres para workloads ou conjuntos de dados. Pode haver um único LUN por volume, ou pode haver muitos. A resposta certa depende dos requisitos de gerenciabilidade.
- Espalhar LUNs em um número desnecessário de volumes pode levar a problemas adicionais de sobrecarga e agendamento para operações, como operações de snapshot, números excessivos de objetos exibidos na IU e resultar em alcançar limites de volume da plataforma antes que o limite de LUN seja atingido.

Volumes, LUNs e instantâneos

As políticas e programações de snapshot são colocadas no volume, não no LUN. Um conjunto de dados que consiste em 10 LUNs exigiria apenas uma única política de snapshot quando os LUNs estiverem colocalizados no mesmo volume.

Além disso, a localização conjunta de todos os LUNs relacionados para um determinado conjunto de dados em um único volume proporciona operações de snapshot atômico. Por exemplo, um banco de dados que residia em 10 LUNs ou um ambiente de aplicativo baseado em VMware que consiste em 10 sistemas operacionais diferentes pode ser protegido como um único objeto consistente se os LUNs subjacentes forem todos colocados em um único volume. Se forem colocados em volumes diferentes, os instantâneos podem ou não estar 100% sincronizados, mesmo que programados ao mesmo tempo.

Em alguns casos, um conjunto relacionado de LUNs pode precisar ser dividido em dois volumes diferentes devido aos requisitos de recuperação. Por exemplo, um banco de dados pode ter quatro LUNs para datafiles e dois LUNs para logs. Nesse caso, um volume de arquivo de dados com 4 LUNs e um volume de log com 2 LUNs podem ser a melhor opção. A razão é recuperabilidade independente. Por exemplo, o volume do arquivo de dados poderia ser restaurado seletivamente para um estado anterior, o que significa que todos os quatro LUNs seriam revertidos para o estado do instantâneo, enquanto o volume de log com seus dados críticos não seria afetado.

Volumes, LUNs e SnapMirror

As políticas e operações do SnapMirror são, como operações de snapshot, executadas no volume, e não no LUN.

A realocação de LUNs relacionados em um único volume permite criar uma única relação do SnapMirror e atualizar todos os dados contidos com uma única atualização. Tal como acontece com instantâneos, a atualização também será uma operação atômica. O destino do SnapMirror teria a garantia de ter uma réplica pontual das LUNs de origem. Se os LUNs foram espalhados por vários volumes, as réplicas podem ou não ser consistentes umas com as outras.

Volumes, LUNs e QoS

Embora a QoS possa ser aplicada seletivamente a LUNs individuais, geralmente é mais fácil defini-la no nível do volume. Por exemplo, todos os LUNs usados pelos convidados em um determinado servidor ESX podem ser colocados em um único volume e, em seguida, uma política de QoS adaptável do ONTAP pode ser aplicada. O resultado é um limite de IOPS por TB com dimensionamento automático que se aplica a todos os LUNs.

Da mesma forma, se um banco de dados exigisse 100K IOPS e 10 LUNs ocupados, seria mais fácil definir um único limite de 100K IOPS em um único volume do que definir 10 limites de 10K IOPS individuais, um em cada LUN.

Esquemas de vários volumes

Há alguns casos em que a distribuição de LUNs em vários volumes pode ser benéfica. O principal motivo é o striping do controlador. Por exemplo, um sistema de storage de HA pode hospedar um único banco de dados em que o potencial de processamento e armazenamento em cache completo de cada controlador é necessário. Nesse caso, um design típico seria colocar metade dos LUNs em um único volume no controlador 1 e a outra metade dos LUNs em um único volume no controlador 2.

Da mesma forma, o striping do controlador pode ser usado para balanceamento de carga. Um sistema de HA que hospedava 100 bancos de dados de 10 LUNs cada um pode ser projetado no local em que cada banco de dados recebe um volume de 5 LUN em cada uma das duas controladoras. O resultado é o carregamento simétrico garantido de cada controlador à medida que bancos de dados adicionais são provisionados.

No entanto, nenhum desses exemplos envolve uma proporção de volume para LUN de 1:1:1. O objetivo continua a ser otimizar a capacidade de gerenciamento, colocando LUNs relacionados em volumes.

Um exemplo em que uma taxa de 1:1 LUN para volume faz sentido é a Containerização, onde cada LUN pode realmente representar uma única carga de trabalho e precisa ser gerenciado individualmente. Nesses casos, uma proporção de 1:1 pode ser ótima.

Redimensionamento LUN e redimensionamento LVM

Quando um sistema de arquivos baseado em SAN atingiu seu limite de capacidade, há duas opções para aumentar o espaço disponível:

- Aumente o tamanho dos LUNs
- Adicione um LUN a um grupo de volumes existente e aumente o volume lógico contido

Embora o redimensionamento LUN seja uma opção para aumentar a capacidade, geralmente é melhor usar um LVM, incluindo o Oracle ASM. Uma das principais razões pelas quais LVMs existem é evitar a necessidade de um redimensionamento LUN. Com uma LVM, vários LUNs são colados em um pool virtual de storage. Os volumes lógicos esculpidos fora deste pool são gerenciados pela LVM e podem ser facilmente redimensionados. Um benefício adicional é evitar hotspots em uma determinada unidade, distribuindo um determinado volume lógico em todos os LUNs disponíveis. A migração transparente geralmente pode ser realizada usando o gerenciador de volumes para realocar as extensões subjacentes de um volume lógico para novos LUNs.

LVM striping (distribuição LVM)

A distribuição de LVM refere-se à distribuição de dados entre vários LUNs. O resultado é uma melhoria significativa na performance de muitos bancos de dados.

Antes da era das unidades flash, a distribuição foi usada para ajudar a superar as limitações de desempenho das unidades giratórias. Por exemplo, se um sistema operacional precisar executar uma operação de leitura 1MB, ler que 1MB TB de dados de uma única unidade exigiria muita busca e leitura de cabeça de unidade, pois o 1MB é transferido lentamente. Se esse 1MB TB de dados fosse distribuído em 8 LUNs, o sistema operacional poderia emitir oito operações de leitura 128K em paralelo e reduzir o tempo necessário para concluir a transferência 1MB.

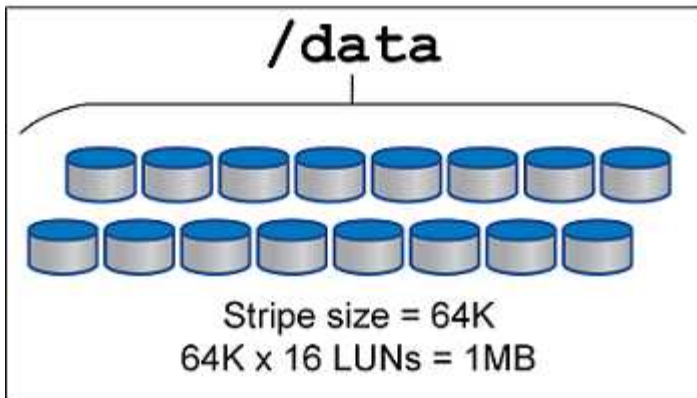
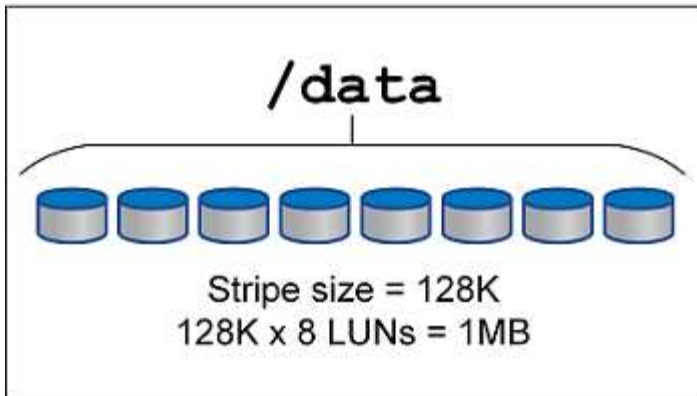
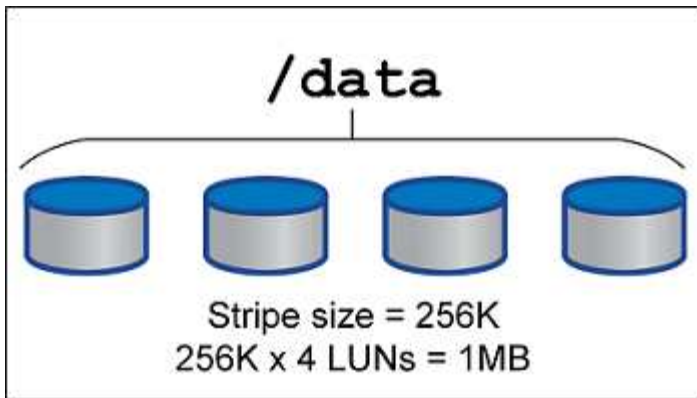
Riscar com unidades giratórias foi mais difícil porque o padrão de e/S tinha que ser conhecido com antecedência. Se o striping não foi ajustado corretamente para os padrões de e/S verdadeiros, as configurações listradas podem danificar o desempenho. Com os bancos de dados Oracle e, especialmente, com configurações all-flash, a distribuição é muito mais fácil de configurar e, comprovadamente, aumentou significativamente o desempenho.

Gerenciadores de volume lógicos, como o Oracle ASM stripe por padrão, mas o LVM de SO nativo não. Alguns deles unem vários LUNs como um dispositivo concatenado, o que resulta em datafiles que existem em um e apenas um dispositivo LUN. Isso causa pontos quentes. Outras implementações de LVM são padrão para extensões distribuídas. Isso é semelhante ao striping, mas é mais grosseiro. Os LUNs no grupo de volumes são cortados em pedaços grandes, chamados de extensões e normalmente medidos em muitos megabytes, e os volumes lógicos são então distribuídos por essas extensões. O resultado é a e/S aleatória contra um arquivo deve ser bem distribuída entre LUNs, mas as operações de e/S sequenciais não são tão eficientes quanto poderiam ser.

A e/S de aplicativos intensivos quase sempre é (a) em unidades do tamanho básico do bloco ou (b) um megabyte.

O principal objetivo de uma configuração de distribuição é garantir que a e/S de arquivo único possa ser executada como uma única unidade, e as e/S de vários blocos, que devem ter 1MB MB de tamanho, possam ser paralelizadas uniformemente em todos os LUNs no volume distribuído. Isso significa que o tamanho do stripe não deve ser menor que o tamanho do bloco do banco de dados e o tamanho do stripe multiplicado pelo número de LUNs deve ser 1MB.

A figura a seguir mostra três opções possíveis para o tamanho da faixa e ajuste de largura. O número de LUNs é selecionado para atender aos requisitos de performance conforme descrito acima, mas, em todos os casos, o total de dados em um único stripe é de 1MB.



NFS

Visão geral

A NetApp fornece storage NFS de nível empresarial há mais de 30 anos, e seu uso está crescendo cada vez mais, levando em consideração a simplicidade de suas infraestruturas baseadas em nuvem.

O protocolo NFS inclui várias versões com requisitos variados. Para obter uma descrição completa da configuração NFS com o ONTAP, ["TR-4067 NFS nas práticas recomendadas da ONTAP"](#) consulte . As seções a seguir abrangem alguns dos requisitos mais críticos e erros comuns do usuário.

Versões de NFS

O cliente do sistema operacional NFS deve ter suporte do NetApp.

- O NFSv3 é suportado com os sistemas operacionais que seguem o padrão NFSv3.

- O NFSv3 é compatível com o cliente Oracle DNFS.
- O NFSv4 é compatível com todos os sistemas operacionais que seguem o padrão NFSv4.
- NFSv4,1 e NFSv4,2 requerem suporte específico ao SO. Consulte o "[NetApp IMT](#)" para ver os SO suportados.
- O suporte ao Oracle DNFS para NFSv4,1 requer o Oracle 12.2.0.2 ou superior.



"[Matriz de suporte NetApp](#)"O para NFSv3 e NFSv4 não inclui sistemas operacionais específicos. Todos os SO que obedecem ao RFC são geralmente suportados. Ao pesquisar no IMT on-line para suporte a NFSv3 ou NFSv4, não selecione um sistema operacional específico porque não haverá correspondências exibidas. Todos os SO são implicitamente suportados pela política geral.

Tabelas de slots TCP do Linux NFSv3

As tabelas de slot TCP são equivalentes a NFSv3 mm de profundidade de fila do adaptador de barramento do host (HBA). Essas tabelas controlam o número de operações NFS que podem ficar pendentes de uma só vez. O valor padrão é geralmente 16, o que é muito baixo para um desempenho ideal. O problema oposto ocorre em kernels Linux mais recentes, que podem aumentar automaticamente o limite da tabela de slots TCP para um nível que satura o servidor NFS com solicitações.

Para um desempenho ideal e para evitar problemas de desempenho, ajuste os parâmetros do kernel que controlam as tabelas de slots TCP.

Executar o `sysctl -a | grep tcp.*.slot_table` comando e respeitar os seguintes parâmetros:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Todos os sistemas Linux devem incluir `sunrpc.tcp_slot_table_entries`, mas apenas alguns incluem `sunrpc.tcp_max_slot_table_entries`. Ambos devem ser definidos para 128.

Cuidado

A falha em definir esses parâmetros pode ter efeitos significativos no desempenho. Em alguns casos, o desempenho é limitado porque o sistema operacional linux não está emitindo e/S suficiente. Em outros casos, as latências de e/S aumentam à medida que o sistema operacional linux tenta emitir mais e/S do que pode ser reparado.

ADR e NFS

Alguns clientes relataram problemas de desempenho resultantes de uma quantidade excessiva de e/S nos dados ADR no local. O problema geralmente não ocorre até que muitos dados de desempenho tenham sido acumulados. A razão para a e/S excessiva é desconhecida, mas este problema parece ser um resultado de processos Oracle repetidamente verificando o diretório de destino para mudanças.

A remoção `noac` das opções e/ou `actimeo=0` montagem permite que o armazenamento em cache do sistema operacional do host ocorra e reduz os níveis de e/S de storage.



A NetApp recomenda não colocar ADR dados em um sistema de arquivos com `noac` ou `actimeo=0` porque problemas de desempenho são prováveis. Separe ADR os dados em um ponto de montagem diferente, se necessário.

nfs-rootonly e mount-rootonly

O ONTAP inclui uma opção NFS chamada `nfs-rootonly` que controla se o servidor aceita conexões de tráfego NFS de portas altas. Como medida de segurança, apenas o usuário raiz tem permissão para abrir conexões TCP/IP usando uma porta de origem abaixo de 1024, porque essas portas são normalmente reservadas para uso do sistema operacional, não para processos de usuário. Essa restrição ajuda a garantir que o tráfego NFS seja de um cliente NFS do sistema operacional real e não de um processo mal-intencionado emulando um cliente NFS. O cliente Oracle DNFS é um driver de espaço de usuário, mas o processo é executado como `root`, portanto geralmente não é necessário alterar o valor de `nfs-rootonly`. As conexões são feitas a partir de portas baixas.

A `mount-rootonly` opção só se aplica a NFSv3. Ele controla se a chamada DE MONTAGEM RPC será aceita a partir de portas maiores que 1024. Quando o DNFS é usado, o cliente está novamente executando como `root`, para que ele possa abrir portas abaixo de 1024. Este parâmetro não tem efeito.

Os processos de abertura de conexões com DNFS em NFS versões 4,0 e superiores não são executados como raiz e, portanto, exigem portas acima de 1024. O `nfs-rootonly` parâmetro deve ser definido como desativado para que o DNFS conclua a conexão.

Se `nfs-rootonly` o estiver ativado, o resultado será uma parada durante a fase de montagem abrindo conexões DNFS. A saída `sqlplus` é semelhante a esta:

```
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
```

O parâmetro pode ser alterado da seguinte forma:

```
Cluster01::> nfs server modify -nfs-rootonly disabled
```



Em situações raras, você pode precisar alterar tanto `nfs-rootonly` quanto `mount-rootonly` para desabilitado. Se um servidor estiver gerenciando um número extremamente grande de conexões TCP, é possível que nenhuma porta abaixo de 1024 esteja disponível e o sistema operacional seja forçado a usar portas mais altas. Esses dois parâmetros ONTAP precisariam ser alterados para permitir que a conexão fosse concluída.

Políticas de exportação de NFS: Superusuário e setuid

Se os binários Oracle estiverem localizados em um compartilhamento NFS, a política de exportação deverá incluir permissões de superusuário e `setuid`.

As exportações de NFS compartilhadas usadas para serviços de arquivos genéricos, como diretórios home do usuário, geralmente esmagam o usuário raiz. Isso significa que uma solicitação do usuário root em um host que montou um sistema de arquivos é remapeada como um usuário diferente com Privileges inferior. Isso ajuda a proteger os dados, impedindo que um usuário root em um servidor específico acesse dados no servidor compartilhado. O bit setuid também pode ser um risco de segurança em um ambiente compartilhado. O bit setuid permite que um processo seja executado como um usuário diferente do usuário que invoca o comando. Por exemplo, um script shell que era de propriedade do root com o bit setuid é executado como root. Se esse script shell pudesse ser alterado por outros usuários, qualquer usuário que não seja root poderá emitir um comando como root atualizando o script.

Os binários Oracle incluem arquivos de propriedade do root e usam o bit setuid. Se os binários Oracle estiverem instalados em um compartilhamento NFS, a política de exportação deverá incluir as permissões de superusuário e setuid apropriadas. No exemplo abaixo, a regra inclui tanto `allow-suid` e permite `superuser` o acesso (raiz) para clientes NFS usando autenticação de 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
```

Configuração NFSv4/4,1

Para a maioria das aplicações, há muito pouca diferença entre NFSv3 e NFSv4. A e/S da aplicação geralmente é muito simples e não se beneficia significativamente de alguns dos recursos avançados disponíveis no NFSv4. Versões mais altas do NFS não devem ser vistas como uma "atualização" da perspectiva do storage de banco de dados, mas sim como versões do NFS que incluem recursos adicionais. Por exemplo, se a segurança de ponta a ponta do modo de privacidade Kerberos (krb5p) for necessária, então NFSv4 será necessário.



A NetApp recomenda usar o NFSv4,1 se forem necessários recursos do NFSv4. Há algumas melhorias funcionais no protocolo NFSv4 em NFSv4,1 que melhoram a resiliência em certos casos de borda.

Mudar para NFSv4 é mais complicado do que simplesmente mudar as opções de montagem de `vers-3` para `vers-4,1`. Uma explicação mais completa da configuração do NFSv4 com o ONTAP, incluindo orientações sobre a configuração do sistema operacional, "[TR-4067 NFS nas práticas recomendadas da ONTAP](#)" consulte . As secções seguintes deste TR explicam alguns dos requisitos básicos para a utilização do NFSv4.

Domínio NFSv4

Uma explicação completa da configuração NFSv4/4,1 está além do escopo deste documento, mas um problema comumente encontrado é uma incompatibilidade no mapeamento de domínio. De um ponto de vista `sysadmin`, os sistemas de arquivos NFS parecem se comportar normalmente, mas os aplicativos relatam erros sobre permissões e/ou setuid em determinados arquivos. Em alguns casos, os administradores concluíram incorretamente que as permissões dos binários do aplicativo foram danificadas e executaram comandos `chown` ou `chmod` quando o problema real era o nome do domínio.

O nome de domínio NFSv4 é definido no ONTAP SVM:

```
Cluster01::> nfs server show -fields v4-id-domain
vserver    v4-id-domain
-----
vserver1   my.lab
```

O nome de domínio NFSv4 no host é definido em `/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
```

Os nomes de domínio devem corresponder. Se não o fizerem, erros de mapeamento semelhantes aos seguintes aparecem em `/var/log/messages`:

```
Apr 12 11:43:08 host1 nfsidmap[16298]: nss_getpwnam: name 'root@my.lab'
does not map into domain 'default.com'
```

Binários de aplicativos, como binários de banco de dados Oracle, incluem arquivos de propriedade do root com o bit setuid, o que significa que uma incompatibilidade nos nomes de domínio NFSv4 causa falhas na inicialização do Oracle e um aviso sobre a propriedade ou permissões de um arquivo chamado `oradism`, que está localizado no `$ORACLE_HOME/bin` diretório. Deve aparecer da seguinte forma:

```
[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 este arquivo aparecer com a propriedade de ninguém, pode haver um problema de mapeamento de domínio NFSv4.

```
[root@host1 bin]# ls -l oradism
-rwsr-x--- 1 nobody oinstall 147848 Apr 17 2019 oradism
```

Para corrigir isso, verifique o `/etc/idmap.cfg` arquivo na configuração `v4-id-domain` no ONTAP e certifique-se de que eles sejam consistentes. Se não estiverem, faça as alterações necessárias, execute `nfsidmap -c` e aguarde um momento para que as alterações se propaguem. A propriedade do arquivo deve então ser devidamente reconhecida como raiz. Se um usuário tivesse tentado executar `chown root` esse arquivo antes que a configuração dos domínios NFS fosse corrigida, talvez seja necessário executar `chown root` novamente.

Oracle Direct NFS (DNFS)

Os bancos de dados Oracle podem usar o NFS de duas maneiras.

Primeiro, ele pode usar um sistema de arquivos montado usando o cliente NFS nativo que faz parte do sistema operacional. Isso às vezes é chamado de kernel NFS, ou kNFS. O sistema de arquivos NFS é montado e usado pelo banco de dados Oracle exatamente o mesmo que qualquer outro aplicativo usaria um sistema de arquivos NFS.

O segundo método é o Oracle Direct NFS (DNFS). Esta é uma implementação do padrão NFS no software de banco de dados Oracle. Ele não altera a maneira como os bancos de dados Oracle são configurados ou gerenciados pelo DBA. Desde que o próprio sistema de armazenamento tenha as configurações corretas, o uso do DNFS deve ser transparente para o grupo DBA e para os usuários finais.

Um banco de dados com o recurso DNFS habilitado ainda tem os sistemas de arquivos NFS usuais montados. Uma vez que o banco de dados está aberto, o banco de dados Oracle abre um conjunto de sessões TCP/IP e executa operações NFS diretamente.

NFS direto

O principal valor do NFS direto da Oracle é ignorar o cliente NFS do host e executar operações de arquivos NFS diretamente em um servidor NFS. A ativação da TI requer apenas a alteração da biblioteca do Oracle Disk Manager (ODM). As instruções para este processo são fornecidas na documentação da Oracle.

O uso do DNFS resulta em uma melhoria significativa no desempenho de e/S e diminui a carga no host e no sistema de armazenamento, pois a e/S é realizada da maneira mais eficiente possível.

Além disso, o Oracle DNFS inclui uma **opção** para multipathing de interface de rede e tolerância a falhas. Por exemplo, duas interfaces 10Gb podem ser Unidas para oferecer 20Gb Gbps de largura de banda. Uma falha de uma interface faz com que a I/O seja tentada novamente na outra interface. A operação geral é muito semelhante ao multipathing FC. Multipathing era comum anos atrás, quando 1GB ethernet era o padrão mais comum. Uma NIC 10Gb é suficiente para a maioria das cargas de trabalho Oracle, mas se for necessário mais, 10Gb NICs podem ser colados.

Quando o DNFS é usado, é fundamental que todos os patches descritos no Oracle Doc 1495104,1 sejam instalados. Se um patch não puder ser instalado, o ambiente deve ser avaliado para garantir que os bugs descritos nesse documento não causem problemas. Em alguns casos, a incapacidade de instalar os patches necessários impede o uso do DNFS.

Não use DNFS com qualquer tipo de resolução de nome de round-robin, incluindo DNS, DDNS, NIS ou qualquer outro método. Isso inclui o recurso de balanceamento de carga DNS disponível no ONTAP. Quando um banco de dados Oracle usando DNFS resolve um nome de host para um endereço IP, ele não deve ser alterado em pesquisas subsequentes. Isso pode resultar em falhas de banco de dados Oracle e possível corrupção de dados.

Ativar DNFS

O Oracle DNFS pode trabalhar com o NFSv3 sem necessidade de configuração além de ativar a biblioteca DNFS (consulte a documentação Oracle para o comando específico necessário), mas se o DNFS não conseguir estabelecer conectividade, ele pode reverter silenciosamente para o cliente NFS do kernel. Se isso acontecer, o desempenho pode ser gravemente afetado.

Se você deseja usar a multiplexação DNFS em várias interfaces, com NFSv4.X, ou usar criptografia, você deve configurar um arquivo oranfstab. A sintaxe é extremamente rigorosa. Pequenos erros no arquivo podem resultar em suspensão de inicialização ou ignorar o arquivo oranfstab.

No momento da escrita, o multipathing DNFS não funciona com o NFSv4,1 com versões recentes do Oracle Database. Um arquivo oranfstab que especifica NFSv4,1 como um protocolo só pode usar uma instrução de caminho único para uma determinada exportação. O motivo é que o ONTAP não oferece suporte ao entroncamento ClientID. Patches do banco de dados Oracle para resolver essa limitação podem estar disponíveis no futuro.

A única maneira de ter certeza de que o DNFS está operando como esperado é consultar as tabelas dnfs.

Abaixo está um exemplo de arquivo oranfstab localizado em /etc. Este é um dos vários locais que um arquivo oranfstab pode ser colocado.

```
[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
```

O primeiro passo é verificar se o DNFS está operacional para os sistemas de arquivos especificados:

```
SQL> select dirname,nfsversion from v$dnfs_servers;

DIRNAME
-----
NFSVERSION
-----
/logs
NFSv3.0

/dbf
NFSv3.0
```

Essa saída indica que o DNFS está em uso com esses dois sistemas de arquivos, mas ele não significa que o oranfstab esteja operacional. Se um erro estivesse presente, o DNFS teria descoberto automaticamente os sistemas de arquivos NFS do host e você ainda poderá ver a mesma saída deste comando.

Multipathing pode ser verificado da seguinte forma:

```
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.

Estas são as conexões que o DNFS está usando. Dois caminhos e canais são visíveis para cada entrada SVRNAME. Isso significa que o multipathing está funcionando, o que significa que o arquivo oranfstab foi reconhecido e processado.

Acesso direto ao NFS e ao sistema de arquivos do host

O uso do DNFS pode ocasionalmente causar problemas para aplicativos ou atividades do usuário que dependem dos sistemas de arquivos visíveis montados no host porque o cliente DNFS acessa o sistema de arquivos fora da banda do sistema operacional do host. O cliente DNFS pode criar, excluir e modificar arquivos sem o conhecimento do sistema operacional.

Quando as opções de montagem para bancos de dados de instância única são usadas, elas permitem o armazenamento em cache de atributos de arquivo e diretório, o que também significa que o conteúdo de um diretório é armazenado em cache. Portanto, o DNFS pode criar um arquivo, e há um curto atraso antes que o sistema operacional releia o conteúdo do diretório e o arquivo se torne visível para o usuário. Isso geralmente não é um problema, mas, em raras ocasiões, utilitários como SAP BR*Tools podem ter problemas. Se isso acontecer, solucione o problema alterando as opções de montagem para usar as recomendações do Oracle RAC. Essa alteração resulta na desativação de todo o cache do host.

Altere apenas as opções de montagem quando (a) DNFS for usado e (b) um problema resulta de um atraso na visibilidade do arquivo. Se o DNFS não estiver em uso, o uso das opções de montagem do Oracle RAC em um banco de dados de instância única resulta em desempenho degradado.



Veja a nota sobre `nosharecache` o in ["Opções de montagem em NFS do Linux"](#) para um problema DNFS específico do Linux que pode produzir resultados incomuns.

Loações e bloqueios de NFS

O NFSv3 está sem monitoração de estado. Isso efetivamente significa que o servidor NFS (ONTAP) não controla quais sistemas de arquivos são montados, por quem, ou quais bloqueios estão realmente no lugar.

O ONTAP tem alguns recursos que gravarão tentativas de montagem para que você tenha uma ideia de quais clientes podem estar acessando dados, e pode haver bloqueios de consultoria presentes, mas essa informação não é garantida para ser 100% completa. Ele não pode ser concluído, porque o rastreamento do estado do cliente NFS não faz parte do padrão NFSv3.

NFSv4 declaração

Em contraste, NFSv4 é stateful. O servidor NFSv4 rastreia quais clientes estão usando quais sistemas de arquivos, quais arquivos existem, quais arquivos e/ou regiões de arquivos estão bloqueados, etc. isso significa que precisa haver comunicação regular entre um servidor NFSv4 para manter os dados de estado atuais.

Os estados mais importantes que estão sendo gerenciados pelo servidor NFS são NFSv4 bloqueios e NFSv4 arrendamentos, e eles estão muito interligados. Você precisa entender como cada um funciona por si mesmo,

e como eles se relacionam uns com os outros.

NFSv4 fechaduras

Com o NFSv3, os bloqueios são consultivos. Um cliente NFS ainda pode modificar ou excluir um arquivo "bloqueado". Um bloqueio NFSv3 não expira por si só, ele deve ser removido. Isso cria problemas. Por exemplo, se você tiver um aplicativo em cluster que crie NFSv3 bloqueios e um dos nós falhar, o que você faz? Você pode codificar o aplicativo nos nós sobreviventes para remover os bloqueios, mas como você sabe que isso é seguro? Talvez o nó "falhou" esteja operacional, mas não esteja se comunicando com o restante do cluster?

Com NFSv4, os bloqueios têm uma duração limitada. Desde que o cliente que detém os bloqueios continue a efetuar o check-in com o servidor NFSv4, nenhum outro cliente tem permissão para adquirir esses bloqueios. Se um cliente não conseguir efetuar o check-in com o NFSv4, os bloqueios eventualmente serão revogados pelo servidor e outros clientes poderão solicitar e obter bloqueios.

NFSv4 arrendamentos

NFSv4 bloqueios estão associados a um leasing de NFSv4. Quando um cliente NFSv4 estabelece uma conexão com um servidor NFSv4, ele recebe um leasing. Se o cliente obtém um bloqueio (existem muitos tipos de bloqueios), então o bloqueio está associado ao leasing.

Este leasing tem um tempo limite definido. Por padrão, o ONTAP definirá o valor de tempo limite para 30 segundos:

```
Cluster01::*> nfs server show -vserver vserver1 -fields v4-lease-seconds

vserver    v4-lease-seconds
-----
vserver1   30
```

Isso significa que um cliente NFSv4 precisa fazer o check-in com o servidor NFSv4 a cada 30 segundos para renovar suas locações.

A locação é renovada automaticamente por qualquer atividade, portanto, se o cliente estiver fazendo trabalho, não há necessidade de realizar operações de adição. Se um aplicativo ficar quieto e não estiver fazendo trabalho real, ele precisará executar uma espécie de operação keep-alive (chamada de SEQUÊNCIA). É essencialmente apenas dizer "Eu ainda estou aqui, por favor, atualize meus arrendamentos."

```
*Question:* What happens if you lose network connectivity for 31 seconds?
O NFSv3 está sem monitoração de estado. Não está à espera de comunicação
dos clientes. O NFSv4 tem estado monitorado e, uma vez decorrido esse
período de locação, o leasing expira e os bloqueios são revogados e os
arquivos bloqueados são disponibilizados para outros clientes.
```

Com o NFSv3, você pode mover cabos de rede, reinicializar switches de rede, fazer alterações de configuração e ter certeza de que nada de ruim aconteceria. Os aplicativos normalmente apenas esperariam pacientemente para que a conexão de rede funcione novamente.

Com NFSv4, você tem 30 segundos (a menos que você tenha aumentado o valor desse parâmetro dentro do

ONTAP) para concluir seu trabalho. Se você exceder isso, suas locações expiram. Normalmente, isso resulta em falhas no aplicativo.

Como exemplo, se você tiver um banco de dados Oracle e tiver uma perda de conectividade de rede (às vezes chamada de "partição de rede") que exceda o tempo limite de concessão, você irá travar o banco de dados.

Aqui está um exemplo do que acontece no log de alerta Oracle se isso acontecer:

```
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 você olhar para os syslogs, você deve ver vários desses erros:

```
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!
```

As mensagens de log são geralmente o primeiro sinal de um problema, além do congelamento do aplicativo. Normalmente, você não vê nada durante a interrupção da rede porque os processos e o próprio sistema operacional estão bloqueados tentando acessar o sistema de arquivos NFS.

Os erros aparecem depois que a rede estiver operacional novamente. No exemplo acima, uma vez que a conectividade foi restabelecida, o sistema operacional tentou readquirir os bloqueios, mas era tarde demais. O arrendamento expirou e os bloqueios foram removidos. Isso resulta em um erro que se propaga até a camada Oracle e causa a mensagem no log de alerta. Você pode ver variações desses padrões dependendo da versão e configuração do banco de dados.

Em resumo, o NFSv3 tolera a interrupção da rede, mas o NFSv4 é mais sensível e impõe um período de locação definido.

E se um tempo limite de 30 segundos não for aceitável? E se você gerenciar uma rede que muda dinamicamente onde os switches são reinicializados ou os cabos são realocados e o resultado é a interrupção ocasional da rede? Você pode optar por estender o período de locação, mas se você quiser fazer isso requer uma explicação de NFSv4 períodos de carência.

NFSv4 períodos de carência

Se um servidor NFSv3 for reinicializado, ele estará pronto para servir o IO quase instantaneamente. Não estava mantendo qualquer tipo de estado sobre os clientes. O resultado é que uma operação de aquisição da ONTAP geralmente parece estar próxima de instantânea. No momento em que um controlador estiver pronto para começar a fornecer dados, ele enviará um ARP para a rede que sinaliza a alteração na topologia. Os clientes normalmente detectam isso quase instantaneamente e os dados continuam fluindo.

NFSv4, no entanto, irá produzir uma breve pausa. É apenas parte de como o NFSv4 funciona.



As seções a seguir são atuais a partir do ONTAP 9.15.1, mas o comportamento de leasing e bloqueio, bem como as opções de ajuste podem mudar de versão para versão. Se você precisar ajustar os tempos de leasing/bloqueio NFSv4, consulte o suporte da NetApp para obter as informações mais recentes.

Os servidores NFSv4 precisam rastrear os arrendamentos, bloqueios e quem está usando quais dados. Se um servidor NFS picar e reiniciar, ou perder energia por um momento, ou for reiniciado durante a atividade de manutenção, o resultado será o leasing/bloqueio e outras informações do cliente serão perdidas. O servidor precisa descobrir qual cliente está usando quais dados antes de retomar as operações. É aqui que entra o período de carência.

Se você de repente ligar o seu servidor NFSv4. Quando ele voltar, os clientes que tentam retomar o IO receberão uma resposta que diz essencialmente: "Perdi informações de leasing/bloqueio. Pretende registrar novamente os seus bloqueios?" Esse é o início do período de carência. O padrão é 45 segundos no ONTAP:

```
Cluster01::> nfs server show -vserver vserver1 -fields v4-grace-seconds

vserver    v4-grace-seconds
-----
vserver1   45
```

O resultado é que, após uma reinicialização, um controlador irá pausar o IO enquanto todos os clientes recuperam seus arrendamentos e bloqueios. Quando o período de carência terminar, o servidor retomará as operações de e/S.

Esse período de carência controla a recuperação de leasing durante alterações na interface de rede, mas há um segundo período de carência que controla a recuperação durante o failover de storage, `locking.grace_lease_seconds`. Esta é uma opção de nível de nó.

```
cluster01::> node run [node names or *] options
locking.grace_lease_seconds
```

Por exemplo, se você precisasse frequentemente executar failovers de LIF e precisasse reduzir o período de carência, você mudaria `v4-grace-seconds`. Se você quisesse melhorar o tempo de retomada de e/S durante o failover da controladora, seria necessário alterar ``locking.grace_lease_seconds``o .

Apenas altere estes valores com cautela e depois de compreender completamente os riscos e consequências. As pausas de e/S envolvidas com operações de failover e migração com o NFSv4.X não podem ser totalmente evitadas. Os períodos de bloqueio, leasing e carência fazem parte da RFC do NFS. Para muitos clientes, o NFSv3 é preferível porque os tempos de failover são mais rápidos.

Prazos de concessão vs períodos de carência

O período de carência e o período de leasing estão ligados. Como mencionado acima, o tempo limite de leasing padrão é de 30 segundos, o que significa que NFSv4 clientes devem fazer check-in com o servidor pelo menos a cada 30 segundos ou perder seus arrendamentos e, por sua vez, seus bloqueios. O período de carência existe para permitir que um servidor NFS reconstrua dados de concessão/bloqueio e o padrão é de 45 segundos. O período de carência deve ser superior ao período de locação. Isso garante que um ambiente cliente NFS projetado para renovar contratos de arrendamento pelo menos a cada 30 segundos terá a capacidade de fazer check-in com o servidor após uma reinicialização. Um período de carência de 45 segundos garante que todos os clientes que esperam renovar seus arrendamentos pelo menos a cada 30 segundos definitivamente tenham a oportunidade de fazê-lo.

Se um tempo limite de 30 segundos não for aceitável, você pode optar por estender o período de locação.

Se você quiser aumentar o tempo limite de leasing para 60 segundos para suportar uma interrupção de rede de 60 segundos, você também terá que aumentar o período de carência. Isso significa que você terá pausas de e/S mais longas durante o failover de controladora.

Isso normalmente não deve ser um problema. Os usuários típicos atualizam somente as controladoras ONTAP uma ou duas vezes por ano, e o failover não planejado devido a falhas de hardware é extremamente raro. Além disso, se você tivesse uma rede em que uma interrupção de rede de 60 segundos era uma possibilidade preocupante, e você precisasse do tempo limite da concessão para 60 segundos, provavelmente você não obteria o failover raro do sistema de storage, resultando em uma pausa de 61 segundos também. Você já reconheceu que tem uma rede que está pausando por mais de 60 segundos com bastante frequência.

Armazenamento em cache NFS

A presença de qualquer uma das seguintes opções de montagem faz com que o cache do host seja desativado:

```
cio, actimeo=0, noac, forcedirectio
```

Essas configurações podem ter um efeito negativo grave na velocidade de instalação do software, patches e operações de backup/restauração. Em alguns casos, especialmente com aplicações em cluster, essas opções são necessárias como resultado inevitável da necessidade de fornecer coerência de cache em todos os nós do cluster. Em outros casos, os clientes usam erroneamente esses parâmetros e o resultado é danos desnecessários no desempenho.

Muitos clientes removem temporariamente essas opções de montagem durante a instalação ou o patch dos binários da aplicação. Essa remoção pode ser realizada com segurança se o usuário verificar que nenhum outro processo está usando ativamente o diretório de destino durante o processo de instalação ou patch.

Tamanhos de transferência NFS

Por padrão, o ONTAP limita os tamanhos de e/S de NFS a 64K.

A e/S aleatória com a maioria dos aplicativos e bancos de dados usa um tamanho de bloco muito menor que está bem abaixo do máximo 64K. A e/S de bloco grande geralmente é paralelizada, portanto o máximo de 64K GB também não é uma limitação para obter largura de banda máxima.

Existem algumas cargas de trabalho em que o máximo 64K cria uma limitação. Em particular, operações de um único processo, como operação de backup ou recuperação ou uma verificação de tabela completa de banco de dados, são executadas de forma mais rápida e eficiente se o banco de dados puder executar

menos, mas maiores I/os. O tamanho ideal de manuseio de e/S para ONTAP é 256K.

O tamanho máximo de transferência para um determinado SVM do ONTAP pode ser alterado da seguinte forma:

```
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::*>
```

Cuidado

Nunca diminua o tamanho máximo de transferência permitido no ONTAP abaixo do valor de rsize/wsize dos sistemas de arquivos NFS atualmente instalados. Isso pode criar pendências ou até mesmo corrupção de dados com alguns sistemas operacionais. Por exemplo, se os clientes NFS estiverem atualmente definidos em um rsize/wsize de 65536, o tamanho máximo de transferência do ONTAP poderá ser ajustado entre 65536 e 1048576 sem efeito porque os próprios clientes são limitados. Reduzir o tamanho máximo de transferência abaixo de 65536 pode danificar a disponibilidade ou os dados.

NVFAIL

O NVFAIL é um recurso do ONTAP que garante a integridade durante cenários de failover catastróficos.

Os bancos de dados estão vulneráveis a corrupção durante eventos de failover de storage porque mantêm grandes caches internos. Se um evento catastrófico exigir forçar um failover de ONTAP ou forçar o switchover de MetroCluster, independentemente da integridade da configuração geral, o resultado será confirmado anteriormente que as alterações podem ser efetivamente descartadas. O conteúdo da matriz de armazenamento salta para trás no tempo, e o estado do cache do banco de dados não reflete mais o estado dos dados no disco. Essa inconsistência resulta em corrupção de dados.

O cache pode ocorrer na camada de aplicativo ou servidor. Por exemplo, uma configuração do Oracle Real Application Cluster (RAC) com servidores ativos em um site primário e remoto armazena dados em cache no Oracle SGA. Uma operação de comutação forçada que resultou em dados perdidos colocaria o banco de dados em risco de corrupção porque os blocos armazenados no SGA podem não corresponder aos blocos no disco.

Um uso menos óbvio do cache está na camada do sistema de arquivos do sistema operacional. Blocos de um sistema de arquivos NFS montado podem ser armazenados em cache no sistema operacional. Como alternativa, um sistema de arquivos em cluster baseado em LUNs localizados no site principal pode ser montado em servidores no local remoto e, mais uma vez, os dados podem ser armazenados em cache. Uma falha do NVRAM ou de uma tomada forçada ou comutação forçada nessas situações pode resultar em corrupção do sistema de arquivos.

O ONTAP protege bancos de dados e sistemas operacionais desse cenário com o NVFAIL e suas configurações associadas.

Utilitário de reclusão ASMRU (ASMRU)

O ONTAP remove com eficiência blocos zerados gravados em um arquivo ou LUN quando a compactação in-line está ativada. Utilitários como o Oracle ASM Reclamation Utility (ASRU) funcionam escrevendo zeros em extensões ASM não utilizadas.

Isso permite que os DBAs recuperem espaço na matriz de armazenamento após os dados serem excluídos. O ONTAP interceta os zeros e desaloca o espaço do LUN. O processo de recuperação é extremamente rápido porque nenhum dado está sendo gravado no sistema de storage.

Do ponto de vista do banco de dados, o grupo de discos ASM contém zeros, e a leitura dessas regiões dos LUNs resultaria em um fluxo de zeros, mas o ONTAP não armazena os zeros nas unidades. Em vez disso, são feitas mudanças simples de metadados que marcam internamente as regiões zeroadas do LUN como vazias de qualquer dado.

Por motivos semelhantes, os testes de desempenho envolvendo dados isolados não são válidos, pois blocos de zeros não são realmente processados como gravações no storage array.



Ao usar o ASRU, certifique-se de que todos os patches recomendados pela Oracle estejam instalados.

Virtualização

A virtualização de bancos de dados com VMware, Oracle OLVM ou KVM é uma escolha cada vez mais comum para os clientes da NetApp que escolheram a virtualização até mesmo para seus bancos de dados mais críticos.

Capacidade de suporte

Existem muitos equívocos sobre as políticas de suporte Oracle para virtualização, especialmente para produtos VMware. Não é incomum ouvir que o Oracle outright não suporta virtualização. Essa noção está incorreta e leva a oportunidades perdidas para se beneficiar da virtualização. O Oracle Doc ID 249212,1 discute os requisitos reais e raramente é considerado pelos clientes como uma preocupação.

Se ocorrer um problema em um servidor virtualizado e esse problema for anteriormente desconhecido para o suporte Oracle, o cliente poderá ser solicitado a reproduzir o problema em hardware físico. Um cliente Oracle que executa uma versão de ponta a ponta de um produto pode não querer usar a virtualização devido ao potencial de problemas de capacidade de suporte, mas essa situação não tem sido um mundo real para clientes de virtualização que usam versões de produtos Oracle geralmente disponíveis.

Apresentação de armazenamento

Os clientes que consideram a virtualização de seus bancos de dados devem basear suas decisões de storage em suas necessidades de negócios. Embora esta seja uma declaração geralmente verdadeira para todas as DECISÕES DE TI, é especialmente importante para projetos de banco de dados, porque o tamanho e o escopo dos requisitos variam consideravelmente.

Existem três opções básicas para apresentação de armazenamento:

- LUNs virtualizados em datastores do hipervisor
- iSCSI LUNs gerenciados pelo iniciador iSCSI na VM, não pelo hipervisor

- Sistemas de arquivos NFS montados pela VM (não em um datastore baseado em NFS)
- Mapeamentos diretos de dispositivos. Os RDMs VMware são desfavorecidos pelos clientes, mas os dispositivos físicos ainda são frequentemente mapeados diretamente com a virtualização KVM e OLVM.

Desempenho

O método de apresentar armazenamento a um convidado virtualizado geralmente não afeta o desempenho. Os sistemas operacionais de host, drivers de rede virtualizados e implementações de armazenamento de dados do hipervisor são altamente otimizados e geralmente podem consumir toda a largura de banda de rede FC ou IP disponível entre o hipervisor e o sistema de storage, desde que sejam seguidas as práticas recomendadas básicas. Em alguns casos, obter desempenho ideal pode ser um pouco mais fácil usando uma abordagem de apresentação de storage em comparação com outra, mas o resultado final deve ser comparável.

Capacidade de gerenciamento

O fator chave para decidir como apresentar o storage a um convidado virtualizado é a capacidade de gerenciamento. Não há método certo ou errado. A melhor abordagem depende das necessidades operacionais, habilidades e preferências DE TI.

Os fatores a considerar incluem:

- **Transparência.** Quando uma VM gerencia seus sistemas de arquivos, é mais fácil para um administrador de banco de dados ou um administrador de sistema identificar a origem dos sistemas de arquivos para seus dados. Os sistemas de arquivos e LUNs são acessados de forma diferente do que com um servidor físico.
- **Consistência.** Quando uma VM possui seus sistemas de arquivos, o uso ou não uso de uma camada de hipervisor afeta a gerenciabilidade. Os mesmos procedimentos de provisionamento, monitoramento, proteção de dados etc. podem ser usados em todo o estado, incluindo ambientes virtualizados e não virtualizados.

Por outro lado, em um data center 100% virtualizado, pode ser preferível também usar o armazenamento baseado em datastore em toda a área ocupada com a mesma lógica mencionada acima - consistência - a capacidade de usar os mesmos procedimentos para provisionamento, proteção, monitoramento e proteção de dados.

- **Estabilidade e resolução de problemas.** Quando uma VM é proprietária de seus sistemas de arquivos, fornecer performance boa e estável e solucionar problemas fica mais simples porque toda a pilha de storage está presente na VM. A única função do hipervisor é transportar quadros FC ou IP. Quando um datastore é incluído em uma configuração, complica a configuração introduzindo outro conjunto de timeouts, parâmetros, arquivos de log e possíveis bugs.
- * Portabilidade.* Quando uma VM possui seus sistemas de arquivos, o processo de mover um ambiente Oracle se torna muito mais simples. Os sistemas de arquivos podem ser movidos facilmente entre convidados virtualizados e não virtualizados.
- **Aprisionamento do fornecedor.** depois que os dados são colocados em um datastore, usar um hypervisor diferente ou tirar os dados do ambiente virtualizado torna-se completamente difícil.
- **Capacitação de snapshot.** Os procedimentos tradicionais de backup em um ambiente virtualizado podem se tornar um problema devido à largura de banda relativamente limitada. Por exemplo, um tronco 10GbE de quatro portas pode ser suficiente para suportar as necessidades diárias de desempenho de muitos bancos de dados virtualizados, mas esse tronco seria insuficiente para executar backups usando RMAN ou outros produtos de backup que exigem o streaming de uma cópia de tamanho completo dos dados. O resultado é que um ambiente virtualizado cada vez mais consolidado precisa executar backups por meio

de snapshots de storage. Isso evita a necessidade de sobrecompilar a configuração do hipervisor apenas para suportar os requisitos de largura de banda e CPU na janela de backup.

O uso de sistemas de arquivos de propriedade de convidados às vezes facilita a utilização de backups e restaurações baseados em snapshot, porque os objetos de storage que precisam de proteção podem ser segmentados com mais facilidade. No entanto, há um número cada vez maior de produtos de proteção de dados de virtualização que se integram bem aos armazenamentos de dados e snapshots. A estratégia de backup deve ser totalmente considerada antes de tomar uma decisão sobre como apresentar o storage a um host virtualizado.

Drivers paravirtualizados

Para um desempenho ideal, o uso de drivers de rede paravirtualizados é fundamental. Quando um datastore é usado, um driver SCSI paravirtualizado é necessário. Um driver de dispositivo paravirtualizado permite que um convidado se integre mais profundamente no hypervisor, em vez de um driver emulado no qual o hypervisor passa mais tempo de CPU imitando o comportamento do hardware físico.

Sobrecarga da RAM

Overcommit RAM significa configurar mais RAM virtualizada em vários hosts do que existe no hardware físico. Isso pode causar problemas inesperados de desempenho. Ao virtualizar um banco de dados, os blocos subjacentes do Oracle SGA não devem ser trocados pelo hipervisor para storage. Isso causa resultados de desempenho altamente instáveis.

Striping do datastore

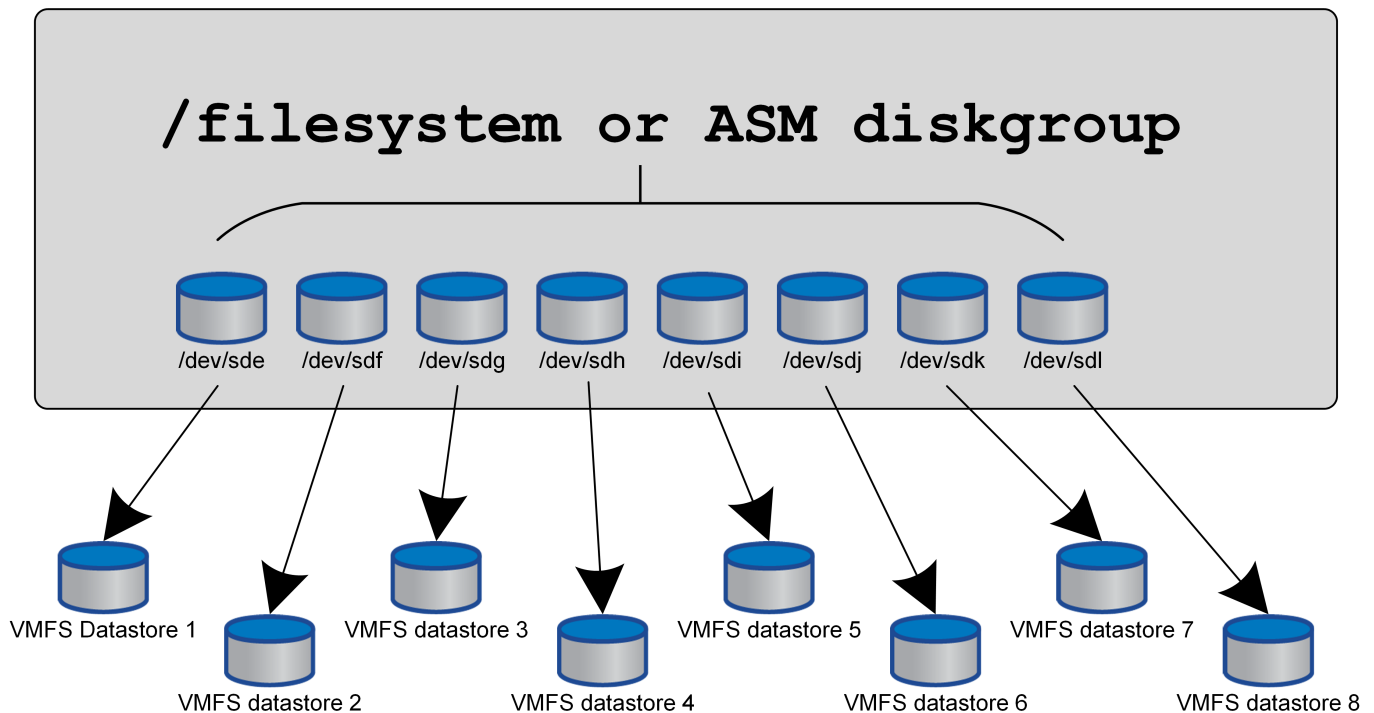
Ao usar bancos de dados com datastores, há um fator crítico a considerar em relação ao desempenho - striping.

As tecnologias de armazenamento de dados, como o VMFS, são capazes de abranger vários LUNs, mas não são dispositivos distribuídos. Os LUNs são concatenados. O resultado final pode ser pontos quentes LUN. Por exemplo, um banco de dados Oracle típico pode ter um grupo de discos ASM de 8 LUN. Todos os 8 LUNs virtualizados podem ser provisionados em um datastore VMFS de 8 LUN, mas não há garantia sobre quais LUNs os dados residirão. A configuração resultante pode ser todo 8 LUN virtualizado ocupando um único LUN dentro do datastore VMFS. Isso se torna um gargalo de desempenho.

Geralmente é necessário riscar. Com alguns hypervisors, incluindo KVM, é possível criar um datastore usando o striping LVM como descrito ["aqui"](#). Com a VMware, a arquitetura parece um pouco diferente. Cada LUN virtualizado precisa ser colocado em um datastore VMFS diferente.

Por exemplo:

Virtualized host



O driver principal para essa abordagem não é o ONTAP, é devido à limitação inerente do número de operações que uma única VM ou LUN de hipervisor pode servir em paralelo. Um único LUN de ONTAP geralmente suporta muito mais IOPS do que um host pode solicitar. O limite de desempenho de LUN único é quase universalmente um resultado do sistema operacional do host. Assim, a maioria dos bancos de dados precisa de 4 a 8 LUNs para atender às suas necessidades de performance.

As arquiteturas VMware precisam planejar suas arquiteturas cuidadosamente para garantir que as máximas de armazenamento de dados e/ou caminho LUN não sejam encontradas com essa abordagem. Additionally, não há nenhum requisito para um conjunto exclusivo de armazenamentos de dados VMFS para cada banco de dados. A principal necessidade é garantir que cada host tenha um conjunto limpo de 4-8 caminhos de e/S, desde os LUNs virtualizados até os LUNs de back-end no próprio sistema de storage. Em raras ocasiões, ainda mais datastores podem ser benéficos para demandas de desempenho verdadeiramente extremas, mas 4-8 LUNs geralmente são suficientes para 95% de todos os bancos de dados. Um único volume ONTAP contendo 8 LUNs pode suportar até 250.000 IOPS de bloco Oracle aleatório com uma configuração típica de os/ONTAP/rede.

Disposição em camadas

Visão geral

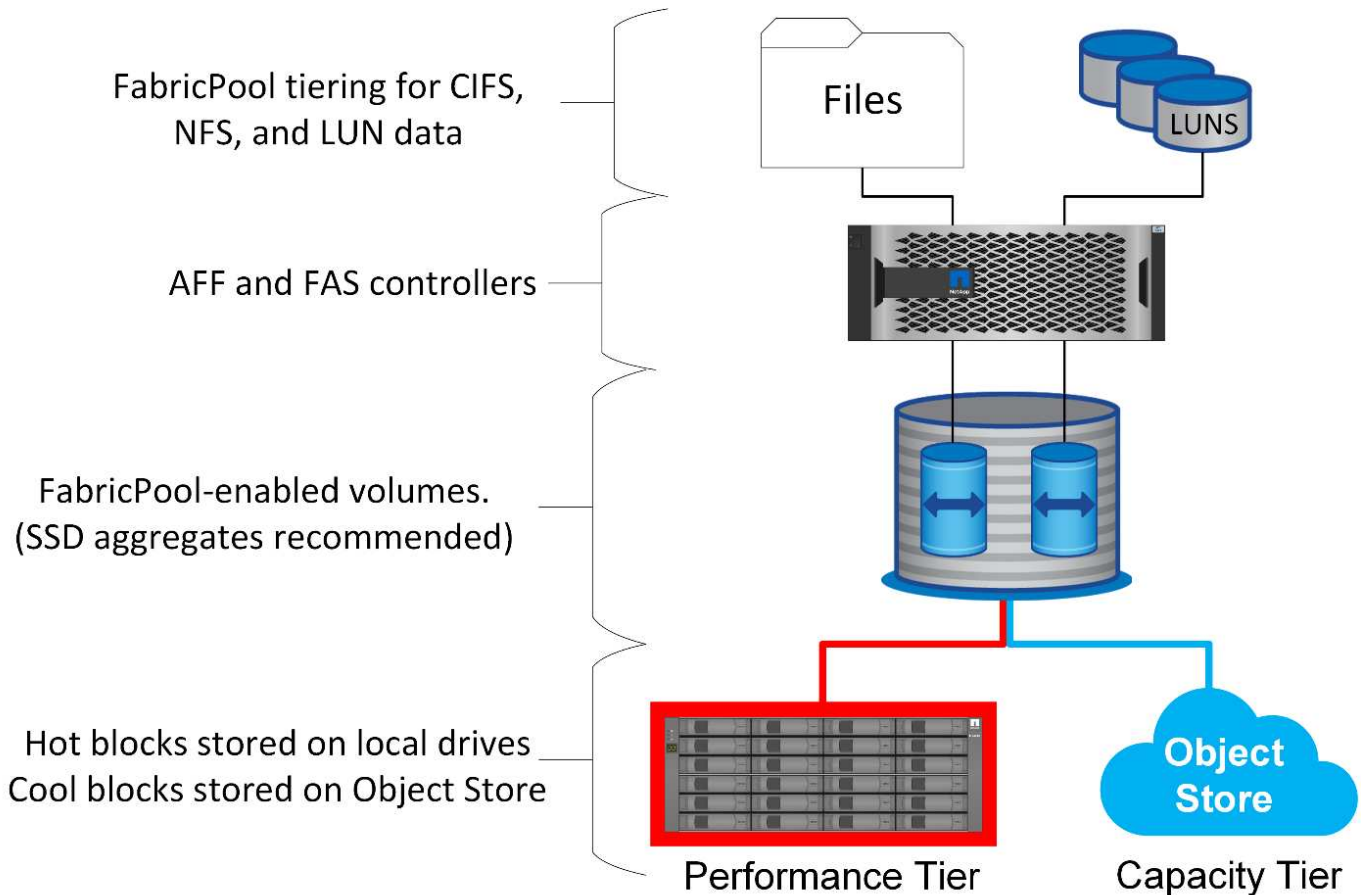
Entender como a disposição em camadas do FabricPool afeta a Oracle e outros bancos de dados requer um entendimento da arquitetura FabricPool de baixo nível.

Arquitetura

O FabricPool é uma tecnologia em camadas que classifica os blocos como ativos ou inativos e os coloca na camada mais apropriada de storage. A camada de performance fica na maioria das vezes localizada no storage SSD e hospeda os blocos de dados ativos. A camada de capacidade fica em um armazenamento de objetos e hospeda os blocos de dados inativos. O suporte ao storage de objetos inclui NetApp StorageGRID,

ONTAP S3, storage Microsoft Azure Blob, serviço de storage de objetos Alibaba Cloud, IBM Cloud Object Storage, storage do Google Cloud e Amazon AWS S3.

Várias políticas de disposição em camadas estão disponíveis para controlar a classificação dos blocos como ativo ou inativo. Além disso, as políticas podem ser definidas por volume e alteradas conforme necessário. Somente os blocos de dados são movidos entre as categorias de performance e capacidade. Os metadados que definem a estrutura do sistema de arquivos e LUN sempre permanecem no nível de performance. Como resultado, o gerenciamento é centralizado no ONTAP. Os arquivos e LUNs não são diferentes dos dados armazenados em qualquer outra configuração do ONTAP. O controlador NetApp AFF ou FAS aplica as políticas definidas para mover dados para o nível apropriado.



Provedores de armazenamento de objetos

Os protocolos de armazenamento de objetos usam solicitações HTTP ou HTTPS simples para armazenar um grande número de objetos de dados. O acesso ao armazenamento de objetos deve ser confiável, pois o acesso aos dados do ONTAP depende do atendimento imediato das solicitações. As opções incluem as opções de acesso padrão e pouco frequentes do Amazon S3, além do Microsoft Azure Hot and Cool Blob Storage, IBM Cloud e Google Cloud. As opções de arquivamento, como o Amazon Glacier e o Amazon Archive, não são suportadas porque o tempo necessário para recuperar dados pode exceder as tolerâncias dos sistemas operacionais e aplicativos host.

O NetApp StorageGRID também é compatível e é uma solução de classe empresarial ideal. Ele é um sistema de storage de objetos de alto desempenho, dimensionável e altamente seguro que pode fornecer redundância geográfica para dados do FabricPool, bem como outras aplicações de armazenamento de objetos que provavelmente farão parte dos ambientes de aplicativos empresariais.

O StorageGRID também pode reduzir custos evitando as cobranças de saída impostas por muitos

fornecedores de nuvem pública pela leitura de dados de seus serviços.

Dados e metadados

Observe que o termo "dados" aqui se aplica aos blocos de dados reais, não aos metadados. Apenas os blocos de dados são dispostos em camadas, enquanto os metadados permanecem na camada de performance. Além disso, o status de um bloco como quente ou frio só é afetado pela leitura do bloco de dados real. A simples leitura do nome, carimbo de data/hora ou metadados de propriedade de um arquivo não afeta a localização dos blocos de dados subjacentes.

Backups

Embora o FabricPool possa reduzir significativamente o espaço físico do storage, ele não é por si só uma solução de backup. Os metadados do NetApp WAFL sempre permanecem no nível de performance. Se um desastre catastrófico destruir o nível de performance, não será possível criar um novo ambiente usando os dados na categoria de capacidade porque não contém metadados do WAFL.

No entanto, o FabricPool pode se tornar parte de uma estratégia de backup. Por exemplo, o FabricPool pode ser configurado com a tecnologia de replicação NetApp SnapMirror. Cada metade do espelho pode ter sua própria conexão com um destino de armazenamento de objetos. O resultado são duas cópias independentes dos dados. A cópia primária consiste em blocos na camada de performance e blocos associados na camada de capacidade, e a réplica é um segundo conjunto de blocos de performance e capacidade.

Políticas de disposição em camadas

Políticas de disposição em camadas

Quatro políticas estão disponíveis no ONTAP, que controlam como os dados da Oracle no nível de desempenho se tornam um candidato a ser relocado para o nível de capacidade.

Apenas Snapshot

O `snapshot-only tiering-policy` aplica-se apenas a blocos que não são compartilhados com o sistema de ficheiros ativo. Essencialmente, isso resulta em camadas de backups de bancos de dados. Os blocos se tornam candidatos à disposição em camadas depois que um snapshot é criado e o bloco é substituído, resultando em um bloco que existe apenas no snapshot. O atraso antes de um `snapshot-only` bloco ser considerado frio é controlado pela `tiering-minimum-cooling-days` definição do volume. A gama a partir de ONTAP 9.8 é de 2 a 183 dias.

Muitos conjuntos de dados têm taxas de alteração baixas, o que resulta em economias mínimas com essa política. Por exemplo, um banco de dados típico observado no ONTAP tem uma taxa de alteração inferior a 5% por semana. Os logs de arquivo de banco de dados podem ocupar um espaço extenso, mas eles geralmente continuam a existir no sistema de arquivos ativo e, portanto, não seriam candidatos para a disposição em camadas sob esta política.

Auto

A `auto` política de disposição em camadas estende a disposição em camadas para blocos específicos de snapshot e blocos no sistema de arquivos ativo. O atraso antes de um bloco ser considerado frio é controlado pela `tiering-minimum-cooling-days` definição do volume. A gama a partir de ONTAP 9.8 é de 2 a 183 dias.

Essa abordagem permite opções de disposição em categorias que não estão disponíveis na `snapshot-only` política. Por exemplo, uma política de proteção de dados pode exigir 90 dias de retenção de determinados arquivos de log. Definir um período de resfriamento de 3 dias resulta em todos os arquivos de log com mais de 3 dias para serem dispostos na camada de desempenho. Essa ação libera espaço substancial na categoria de performance e ainda permite que você visualize e gerencie todos os 90 dias de dados.

Nenhum

A `none` política de disposição em categorias impede que blocos adicionais sejam dispostos em camadas na camada de storage, mas todos os dados ainda na camada de capacidade permanecem na camada de capacidade até que sejam lidos. Se o bloco for então lido, ele será puxado para trás e colocado no nível de desempenho.

O principal motivo para usar a `none` política de disposição em camadas é impedir que blocos sejam dispostos em camadas, mas pode se tornar útil alterar as políticas ao longo do tempo. Por exemplo, digamos que um conjunto de dados específico seja amplamente categorizado na camada de capacidade, mas surge uma necessidade inesperada de recursos completos de performance. A política pode ser alterada para impedir qualquer disposição em camadas adicional e para confirmar que todos os blocos de leitura de volta à medida que a IO aumenta permanecem na categoria de performance.

Tudo

A `all` política de disposição em categorias substitui a `backup` política a partir do ONTAP 9.6. `backup`A` política aplicada somente a volumes de proteção de dados, o que significa um destino do SnapMirror ou do NetApp SnapVault. A ``all` política funciona da mesma forma, mas não se restringe a volumes de proteção de dados.

Com essa política, os blocos são imediatamente considerados frios e elegíveis para serem dispostos na camada de capacidade imediatamente.

Essa política é especialmente apropriada para backups de longo prazo. Ele também pode ser usado como uma forma de Gerenciamento de armazenamento hierárquico (HSM). No passado, o HSM era comumente usado para categorizar os blocos de dados de um arquivo em fita, mantendo o próprio arquivo visível no sistema de arquivos. Um volume FabricPool com a `all` política permite armazenar arquivos em um ambiente visível e gerenciável, mas não consome praticamente espaço na camada de storage local.

Políticas de recuperação

As políticas de disposição em camadas controlam quais blocos de banco de dados Oracle são dispostos da camada de performance para a camada de capacidade. As políticas de recuperação controlam o que acontece quando um bloco que foi categorizado é lido.

Padrão

Todos os volumes do FabricPool são inicialmente definidos como `default`, o que significa que o comportamento é controlado pela política de recuperação de nuvem. "O comportamento exato depende da política de disposição em camadas usada.

- `auto`— apenas recuperar dados lidos aleatoriamente
- `snapshot-only`— recuperar todos os dados lidos sequencialmente ou aleatoriamente
- `none`— recuperar todos os dados lidos sequencialmente ou aleatoriamente

- `all`– não recupere dados do nível de capacidade

Na leitura

A configuração `cloud-retrieval-policy` como `On-read` substitui o comportamento padrão para que uma leitura de qualquer dado em camadas resulte no retorno desses dados ao nível de desempenho.

Por exemplo, um volume pode ter sido levemente usado por um longo tempo sob `auto` a política de disposição em camadas e a maioria dos blocos agora está dividida.

Se uma mudança inesperada nas necessidades de negócios exigir que alguns dos dados sejam verificados repetidamente para preparar um determinado relatório, pode ser desejável alterar o `cloud-retrieval-policy` para `on-read` para garantir que todos os dados lidos sejam devolvidos ao nível de desempenho, incluindo dados de leitura sequencial e aleatória. Isso melhoraria o desempenho de e/S sequenciais em relação ao volume.

Promover

O comportamento da política `promover` depende da política de disposição em camadas. Se a política de disposição em categorias for `auto`, a configuração do `cloud-retrieval-policy` para `promote` recuperará todos os blocos da camada de capacidade na próxima verificação de disposição em categorias.

Se a política de disposição em categorias for `snapshot-only`, os únicos blocos retornados serão os blocos associados ao sistema de arquivos ativo. Normalmente, isso não teria nenhum efeito porque os únicos blocos dispostos sob `snapshot-only` a política seriam blocos associados exclusivamente a snapshots. Não haveria blocos em camadas no sistema de arquivos ativo.

No entanto, se os dados em um volume tiverem sido restaurados por uma operação de clone de arquivo ou `SnapRestore` de volume a partir de um `snapshot`, alguns dos blocos que foram dispostos em camadas por estarem associados apenas a snapshots podem agora ser exigidos pelo sistema de arquivos ativo. Pode ser desejável alterar temporariamente a `cloud-retrieval-policy` política para `promote` recuperar rapidamente todos os blocos necessários localmente.

Nunca

Não recupere blocos da camada de capacidade.

Estratégias de disposição em camadas

Disposição completa de arquivos em categorias

Embora a disposição em camadas do `FabricPool` opere no nível de bloco, em alguns casos, ela pode ser usada para fornecer a disposição em camadas no nível do arquivo.

Muitos conjuntos de dados de aplicações são organizados por data, e esses dados geralmente têm cada vez menos probabilidade de serem acessados à medida que envelhecem. Por exemplo, um banco pode ter um repositório de arquivos PDF que contém cinco anos de declarações de clientes, mas apenas os últimos meses estão ativos. O `FabricPool` pode ser usado para realocar arquivos de dados mais antigos para o nível de capacidade. Um período de resfriamento de 14 dias garantiria que os 14 dias mais recentes de arquivos PDF permaneçam no nível de desempenho. Além disso, os arquivos que são lidos pelo menos a cada 14 dias permanecerão ativos e, portanto, permanecerão no nível de desempenho.

Políticas

Para implementar uma abordagem de disposição em camadas baseada em arquivos, você precisa ter arquivos gravados e não modificados posteriormente. A `tiering-minimum-cooling-days` política deve ser definida suficientemente alta para que os arquivos que você possa precisar permaneçam no nível de desempenho. Por exemplo, um conjunto de dados para o qual os 60 dias de dados mais recentes são necessários com garantias de desempenho ideais, definindo `tiering-minimum-cooling-days` o período como 60. Resultados semelhantes também podem ser alcançados com base nos padrões de acesso ao arquivo. Por exemplo, se os 90 dias mais recentes de dados forem necessários e o aplicativo estiver acessando esse período de 90 dias de dados, os dados permanecerão na categoria de performance. Ao definir `tiering-minimum-cooling-days` o período como 2, você obtém a disposição imediata de categorias após os dados ficarem menos ativos.

```
`auto`A política é necessária para impulsionar a disposição em camadas desses blocos, porque somente a `auto` política afeta os blocos que estão no sistema de arquivos ativo.
```



Qualquer tipo de acesso aos dados repõe os dados do mapa de calor. A verificação de vírus, a indexação e até mesmo a atividade de backup que lê os arquivos de origem impedem a categorização porque o limite necessário `tiering-minimum-cooling-days` nunca é atingido.

Disposição em camadas de arquivos parcial

Como o FabricPool funciona no nível de bloco, os arquivos sujeitos a alteração podem ser parcialmente dispostos em camadas no storage de objetos, permanecendo também parcialmente na camada de performance.

Isso é comum com bancos de dados. Bancos de dados que contêm blocos inativos também são candidatos à disposição em camadas do FabricPool. Por exemplo, um banco de dados de gerenciamento da cadeia de suprimentos pode conter informações históricas que devem estar disponíveis se necessário, mas não são acessadas durante operações normais. O FabricPool pode ser usado para realocar seletivamente os blocos inativos.

Por exemplo, os arquivos de dados executados em um volume FabricPool com `tiering-minimum-cooling-days` um período de 90 dias mantêm todos os blocos acessados nos 90 dias anteriores no nível de performance. No entanto, qualquer coisa que não seja acessada por 90 dias é realocada para o nível de capacidade. Em outros casos, a atividade normal da aplicação preserva os blocos corretos no nível correto. Por exemplo, se um banco de dados é normalmente usado para processar os 60 dias anteriores de dados regularmente, um período muito menor `tiering-minimum-cooling-days` pode ser definido porque a atividade natural do aplicativo garante que os blocos não sejam transferidos prematuramente.

A `auto` política deve ser usada com cuidado com bancos de dados. Muitos bancos de dados têm atividades periódicas, como processo de fim de trimestre ou operações de reindexação. Se o período dessas operações for maior do que os `tiering-minimum-cooling-days` problemas de desempenho podem ocorrer. Por exemplo, se o processamento no final do trimestre exigir 1TB TB de dados que, de outra forma, foram intocados, esses dados podem agora estar presentes na camada de capacidade. As leituras do nível de capacidade geralmente são extremamente rápidas e podem não causar problemas de desempenho, mas os resultados exatos dependerão da configuração do armazenamento de objetos.

Políticas

A `tiering-minimum-cooling-days` política deve ser definida suficientemente alta para reter arquivos que possam ser necessários no nível de desempenho. Por exemplo, um banco de dados em que os 60 dias mais recentes de dados possam ser necessários com desempenho ideal garantiria a definição `tiering-minimum-cooling-days` do período para 60 dias. Resultados semelhantes também podem ser alcançados com base nos padrões de acesso dos arquivos. Por exemplo, se os 90 dias mais recentes de dados forem necessários e o aplicativo estiver acessando esse período de 90 dias de dados, os dados permanecerão no nível de desempenho. Definir o `tiering-minimum-cooling-days` período para 2 dias classificaria os dados imediatamente após os dados ficarem menos ativos.

```
`auto`A política é necessária para impulsionar a disposição em camadas desses blocos, porque somente a `auto` política afeta os blocos que estão no sistema de arquivos ativo.
```



Qualquer tipo de acesso aos dados repõe os dados do mapa de calor. Portanto, varreduras completas de tabela de banco de dados e até mesmo atividades de backup que leem os arquivos de origem impedem a categorização porque o limite necessário `tiering-minimum-cooling-days` nunca é atingido.

Disposição em camadas do log de arquivamento

Talvez o uso mais importante para o FabricPool seja melhorar a eficiência de dados inativos conhecidos, como logs de transações de banco de dados.

A maioria dos bancos de dados relacionais opera no modo de arquivamento de log de transações para fornecer recuperação pontual. As alterações nos bancos de dados são confirmadas registrando as alterações nos logs de transação e o log de transação é mantido sem ser substituído. O resultado pode ser um requisito para manter um enorme volume de logs de transações arquivados. Exemplos semelhantes existem com muitos outros fluxos de trabalho de aplicações que geram dados que precisam ser retidos, mas é altamente improvável que nunca sejam acessados.

A FabricPool soluciona esses problemas ao fornecer uma única solução com disposição em camadas integrada. Os arquivos são armazenados e permanecem acessíveis em seu local habitual, mas praticamente não ocupam espaço na matriz primária.

Políticas

O uso de uma `tiering-minimum-cooling-days` política de poucos dias resulta na retenção de blocos nos arquivos criados recentemente (que são os arquivos mais prováveis de serem necessários no curto prazo) na categoria de performance. Os blocos de dados de arquivos mais antigos são movidos para o nível de capacidade.

O `auto` aplica a disposição em camadas quando o limite de resfriamento tiver sido atingido, independentemente de os logs terem sido excluídos ou continuarem a existir no sistema de arquivos primário. Armazenar todos os logs potencialmente necessários em um único local no sistema de arquivos ativo também simplifica o gerenciamento. Não há razão para pesquisar instantâneos para localizar um arquivo que precisa ser restaurado.

Alguns aplicativos, como o Microsoft SQL Server, truncam arquivos de log de transações durante operações de backup para que os logs não estejam mais no sistema de arquivos ativo. A capacidade pode ser salva

usando a `snapshot-only` política de disposição em camadas, mas a `auto` política não é útil para dados de log, porque raramente devem existir dados de log refrigerados no sistema de arquivos ativo.

Disposição do Snapshot em camadas

A versão inicial do FabricPool visou o caso de uso de backup. O único tipo de blocos que podiam ser dispostos em camadas eram blocos que não estavam mais associados aos dados no sistema de arquivos ativo. Portanto, apenas os blocos de dados do snapshot podem ser movidos para a camada de capacidade. Essa continua sendo uma das opções de disposição em categorias mais seguras quando você precisa garantir que a performance nunca seja afetada.

Políticas - instantâneos locais

Existem duas opções para a disposição em camadas de blocos snapshot inativos na camada de capacidade. Primeiro, a `snapshot-only` política segmenta apenas os blocos de snapshot. Embora `auto` a política inclua os `snapshot-only` blocos, ela também dispõe blocos do sistema de arquivos ativo. Isso pode não ser desejável.

O `tiering-minimum-cooling-days` valor deve ser definido para um período de tempo que torne os dados que possam ser necessários durante uma restauração disponíveis no nível de desempenho. Por exemplo, a maioria dos cenários de restauração de um banco de dados de produção crítico inclui um ponto de restauração em algum momento nos últimos dias. Definir um `tiering-minimum-cooling-days` valor de 3 garantirá que qualquer restauração do arquivo resulte em um arquivo que imediatamente forneça o máximo de desempenho. Todos os blocos nos arquivos ativos ainda estão presentes no storage rápido sem a necessidade de recuperá-los da camada de capacidade.

Políticas - instantâneos replicados

Um snapshot replicado com o SnapMirror ou o SnapVault que é usado somente para recuperação geralmente deve usar a política FabricPool `all`. Com essa política, os metadados são replicados, mas todos os blocos de dados são imediatamente enviados para a categoria de capacidade, o que proporciona o máximo de performance. A maioria dos processos de recuperação envolve e/S sequenciais, que é inerentemente eficiente. O tempo de recuperação do destino do armazenamento de objetos deve ser avaliado, mas, em uma arquitetura bem projetada, esse processo de recuperação não precisa ser significativamente mais lento do que a recuperação de dados locais.

Se os dados replicados também se destinarem a ser usados para clonagem, a `auto` política é mais apropriada, com um `tiering-minimum-cooling-days` valor que engloba dados que se espera que sejam usados regularmente em um ambiente de clonagem. Por exemplo, o conjunto de trabalho ativo de um banco de dados pode incluir dados lidos ou gravados nos três dias anteriores, mas também pode incluir outros 6 meses de dados históricos. Em caso afirmativo, a `auto` política no destino do SnapMirror torna o conjunto de trabalho disponível no nível de desempenho.

Disposição em camadas do backup

Os backups tradicionais de aplicativos incluem produtos como o Oracle Recovery Manager, que criam backups baseados em arquivos fora do local do banco de dados original.

`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.

A `auto` política é a política mais adequada para dados de backup. Isso garante a disposição de camadas de prompt quando o limite de resfriamento tiver sido atingido, independentemente de os arquivos terem sido excluídos ou continuarem existindo no sistema de arquivos primário. Armazenar todos os arquivos potencialmente necessários em um único local no sistema de arquivos ativo também simplifica o gerenciamento. Não há razão para pesquisar instantâneos para localizar um arquivo que precisa ser restaurado.

A `snapshot-only` política poderia ser feita para funcionar, mas essa política só se aplica a blocos que não estão mais no sistema de arquivos ativo. Portanto, os arquivos em um compartilhamento NFS ou SMB devem ser excluídos primeiro antes que os dados possam ser categorizados.

Esta política seria ainda menos eficiente com uma configuração LUN porque a exclusão de um arquivo de um LUN só remove as referências de arquivo dos metadados do sistema de arquivos. Os blocos reais nos LUNs permanecem no lugar até que sejam substituídos. Essa situação pode criar um longo atraso entre o tempo em que um arquivo é excluído e o tempo em que os blocos são substituídos e se tornam candidatos à disposição em camadas. Há algum benefício ao migrar `snapshot-only` os blocos para a categoria de capacidade, mas, no geral, o gerenciamento de dados de backup da FabricPool funciona melhor com essa `auto` política.



Essa abordagem ajuda os usuários a gerenciar o espaço necessário para backups com mais eficiência, mas o próprio FabricPool não é uma tecnologia de backup. A disposição em camadas dos arquivos de backup no armazenamento de objetos simplifica o gerenciamento porque os arquivos ainda estão visíveis no sistema de storage original, mas os blocos de dados no destino do armazenamento de objetos dependem do sistema de storage original. Se o volume de origem for perdido, os dados do armazenamento de objetos não serão mais utilizáveis.

Interrupções de acesso ao armazenamento de objetos

A disposição em categorias de um conjunto de dados com o FabricPool resulta em uma dependência entre o storage array primário e a categoria de armazenamento de objetos. Há muitas opções de armazenamento de objetos que oferecem níveis variados de disponibilidade. É importante entender o impacto de uma possível perda de conectividade entre o storage array primário e a camada de storage de objetos.

Se uma e/S emitida para o ONTAP exigir dados da camada de capacidade e o ONTAP não puder alcançar o nível de capacidade para recuperar blocos, a e/S eventualmente expirará. O efeito deste tempo limite depende do protocolo utilizado. Em um ambiente NFS, o ONTAP responde com uma resposta EJUKEBOX ou EDELAY, dependendo do protocolo. Alguns sistemas operacionais mais antigos podem interpretar isso como um erro, mas os sistemas operacionais atuais e os níveis de patch atuais do cliente Oracle Direct NFS tratam isso como um erro recuperável e continuam aguardando a conclusão da e/S.

Um tempo limite mais curto se aplica a ambientes SAN. Se um bloco no ambiente de armazenamento de objetos for necessário e permanecer inacessível por dois minutos, um erro de leitura será retornado ao host.

O volume ONTAP e os LUNs permanecem online, mas o sistema operacional do host pode sinalizar o sistema de arquivos como estando em um estado de erro.

A política de problemas de conectividade de storage de objetos `snapshot-only` é menos preocupante porque apenas os dados de backup são dispostos em camadas. Problemas de comunicação atrasariam a recuperação de dados, mas de outra forma não afetariam os dados que estão sendo usados ativamente. As políticas `auto` e `all` permitem a disposição em camadas de dados inativos do LUN ativo, o que significa que um erro durante a recuperação de dados do armazenamento de objetos pode afetar a disponibilidade do banco de dados. Uma implantação de SAN com essas políticas deve ser usada somente com conexões de rede e storage de objetos de classe empresarial projetadas para alta disponibilidade. NetApp StorageGRID é a opção superior.

Proteção de dados Oracle

Proteção de dados com o ONTAP

O NetApp sabe que os dados mais críticos são encontrados em bancos de dados.

Uma empresa não pode operar sem acesso a seus dados e, às vezes, os dados definem o negócio. No entanto, a proteção de dados é mais do que apenas garantir um backup utilizável. No entanto, é preciso realizar os backups de maneira rápida e confiável, além de armazená-los em segurança.

O outro lado da proteção de dados é a recuperação de dados. Quando os dados estão inacessíveis, a empresa é afetada e pode estar inoperante até que os dados sejam restaurados. Este processo deve ser rápido e confiável. Finalmente, a maioria dos bancos de dados deve ser protegida contra desastres, o que significa manter uma réplica do banco de dados. A réplica deve estar suficientemente atualizada. Também deve ser rápido e simples tornar a réplica um banco de dados totalmente operacional.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-4591: Proteção de dados Oracle: Backup, recuperação e replicação*.

Planejamento

A arquitetura de proteção de dados empresariais certa depende dos requisitos de negócios relacionados à retenção de dados, capacidade de recuperação e tolerância de interrupções durante vários eventos.

Por exemplo, considere o número de aplicações, bancos de dados e conjuntos de dados importantes no escopo. Criar uma estratégia de backup para um único conjunto de dados que garanta a conformidade com SLAs típicos é bastante simples, pois não há muitos objetos para gerenciar. À medida que o número de conjuntos de dados aumenta, o monitoramento se torna mais complicado e os administradores podem ser forçados a gastar um tempo cada vez maior lidando com falhas de backup. À medida que um ambiente atinge a nuvem e o provedor de serviços dimensiona, é necessária uma abordagem totalmente diferente.

O tamanho do conjunto de dados também afeta a estratégia. Por exemplo, existem muitas opções para backup e recuperação com um banco de dados 100GB porque o conjunto de dados é tão pequeno. A simples cópia de dados de Mídia de backup com ferramentas tradicionais normalmente fornece um rto suficiente para recuperação. Um banco de dados 100TB normalmente precisa de uma estratégia completamente diferente, a menos que o rto permita uma interrupção de vários dias, caso em que um procedimento tradicional de backup e recuperação baseado em cópia pode ser aceitável.

Finalmente, existem fatores fora do próprio processo de backup e recuperação. Por exemplo, existem bancos de dados que suportam atividades críticas de produção, tornando a recuperação um evento raro que só é realizado por DBAs qualificados? Alternativamente, os bancos de dados fazem parte de um grande ambiente

de desenvolvimento em que a recuperação é uma ocorrência frequente e gerenciada por uma EQUIPE DE TI generalista?

Planejamento de rto, RPO e SLA

O ONTAP permite que você personalize facilmente uma estratégia de proteção de dados de banco de dados Oracle de acordo com suas necessidades de negócios.

Esses requisitos incluem fatores como a velocidade de recuperação, a perda máxima de dados permitida e as necessidades de retenção de backup. O plano de proteção de dados também deve levar em consideração vários requisitos regulatórios para retenção e restauração de dados. Finalmente, diferentes cenários de recuperação de dados devem ser considerados, desde a recuperação típica e previsível resultante de erros de usuário ou aplicativo até cenários de recuperação de desastres que incluem a perda completa de um site.

Pequenas alterações nas políticas de proteção e recuperação de dados podem ter um efeito significativo na arquitetura geral de storage, backup e recuperação. É fundamental definir e documentar padrões antes de iniciar o trabalho de design para evitar complicar uma arquitetura de proteção de dados. Recursos desnecessários ou níveis de proteção levam a custos desnecessários e sobrecarga de gerenciamento, e um requisito inicialmente negligenciado pode levar um projeto na direção errada ou exigir alterações de design de última hora.

Objetivo de tempo de recuperação

O objetivo de tempo de recuperação (rto) define o tempo máximo permitido para a recuperação de um serviço. Por exemplo, um banco de dados de recursos humanos pode ter um rto de 24 horas porque, embora seja muito inconveniente perder o acesso a esses dados durante o dia de trabalho, a empresa ainda pode operar. Em contraste, um banco de dados que suporta o Registro geral de um banco teria um rto medido em minutos ou mesmo segundos. Um rto de zero não é possível, porque deve haver uma maneira de diferenciar entre uma interrupção de serviço real e um evento de rotina, como um pacote de rede perdido. No entanto, um rto quase zero é um requisito típico.

Objetivo do ponto de restauração

O objetivo do ponto de restauração (RPO) define a perda máxima de dados tolerável. Em muitos casos, o RPO é determinado exclusivamente pela frequência de snapshots ou atualizações do SnapMirror.

Em alguns casos, o RPO pode ser tornado mais agressivo, protegendo determinados dados de forma seletiva com mais frequência. Em um contexto de banco de dados, o RPO geralmente é uma questão de quanto dados de log podem ser perdidos em uma situação específica. Em um cenário típico de recuperação em que um banco de dados é danificado devido a um bug de produto ou erro de usuário, o RPO deve ser zero, o que significa que não deve haver perda de dados. O procedimento de recuperação envolve restaurar uma cópia anterior dos arquivos de banco de dados e, em seguida, rereproduzir os arquivos de log para trazer o estado do banco de dados até o ponto desejado no tempo. Os arquivos de registro necessários para esta operação já devem estar no local original.

Em cenários incomuns, os dados de log podem ser perdidos. Por exemplo, um acidental ou malicioso `rm -rf *` de arquivos de banco de dados pode resultar na exclusão de todos os dados. A única opção seria restaurar do backup, incluindo arquivos de log, e alguns dados seriam inevitavelmente perdidos. A única opção para melhorar o RPO em um ambiente de backup tradicional seria executar backups repetidos dos dados de log. Isso tem limitações, no entanto, devido à constante movimentação de dados e à dificuldade de manter um sistema de backup como um serviço em constante execução. Um dos benefícios dos sistemas avançados de storage é a capacidade de proteger dados de danos acidentais ou mal-intencionados nos arquivos, além de fornecer um RPO melhor sem movimentação de dados.

Recuperação de desastres

A recuperação de desastres inclui a arquitetura, as políticas e os procedimentos DE TI necessários para recuperar um serviço em caso de desastre físico. Isso pode incluir inundações, incêndios ou pessoas agindo com intenção maliciosa ou negligente.

A recuperação de desastres é mais do que apenas um conjunto de procedimentos de recuperação. É o processo completo de identificar os vários riscos, definir os requisitos de recuperação de dados e continuidade do serviço e fornecer a arquitetura certa com os procedimentos associados.

Ao estabelecer os requisitos de proteção de dados, é essencial diferenciar entre os requisitos típicos de RPO e rto e os requisitos de RPO e rto necessários para a recuperação de desastres. Alguns ambientes de aplicações exigem um RPO de zero e um rto quase zero para situações de perda de dados que vão desde um erro de usuário relativamente normal até um incêndio que destrói um data center. No entanto, existem consequências administrativas e de custos para estes elevados níveis de proteção.

Em geral, os requisitos de recuperação de dados que não sejam de desastres devem ser rigorosos por dois motivos. Primeiro, erros de aplicativos e erros de usuário que danificam dados são previsíveis ao ponto de serem quase inevitáveis. Em segundo lugar, não é difícil projetar uma estratégia de backup que possa fornecer um RPO de zero e um rto baixo, contanto que o sistema de storage não seja destruído. Não há motivo para não solucionar um risco significativo que seja facilmente remediado, e é por isso que os destinos RPO e rto para recuperação local devem ser agressivos.

Os requisitos de rto e RPO para recuperação de desastres variam mais amplamente com base na probabilidade de um desastre e nas consequências da perda ou interrupção de dados associada a uma empresa. Os requisitos de RPO e rto devem ser baseados nas necessidades reais dos negócios e não em princípios gerais. Eles devem levar em conta vários cenários de desastre lógico e físico.

Desastres lógicos

Os desastres lógicos incluem corrupção de dados causada por usuários, erros de aplicativos ou sistemas operacionais e falhas de software. Os desastres lógicos também podem incluir ataques maliciosos de terceiros com vírus ou worms ou explorando vulnerabilidades de aplicativos. Nesses casos, a infraestrutura física não está danificada, mas os dados subjacentes não são mais válidos.

Um tipo cada vez mais comum de desastre lógico é conhecido como ransomware, no qual um vetor de ataque é usado para criptografar dados. A criptografia não danifica os dados, mas torna-os indisponíveis até que o pagamento seja feito a terceiros. Um número cada vez maior de empresas está sendo alvo específico de hacks de ransomware. Para essa ameaça, o NetApp oferece snapshots à prova de violações, onde nem mesmo o administrador de storage pode alterar os dados protegidos antes da data de expiração configurada.

Desastres físicos

Os desastres físicos incluem a falha de componentes de uma infraestrutura que excede seus recursos de redundância e resultam em perda de dados ou perda estendida de serviço. Por exemplo, a proteção RAID fornece redundância de unidade de disco e o uso de HBAs fornece redundância de porta FC e cabo FC. Falhas de hardware de tais componentes são previsíveis e não afetam a disponibilidade.

Em um ambiente corporativo, geralmente é possível proteger a infraestrutura de um local inteiro com componentes redundantes até o ponto em que o único cenário de desastre físico previsível é a perda completa do local. O Planejamento de recuperação de desastre depende da replicação local a local.

Proteção de dados síncrona e assíncrona

Em um mundo ideal, todos os dados seriam replicados de forma síncrona em locais geograficamente

dispersos. Essa replicação nem sempre é viável ou até possível por várias razões:

- A replicação síncrona aumenta inevitavelmente a latência de gravação porque todas as alterações devem ser replicadas em ambos os locais antes que a aplicação/banco de dados possa prosseguir com o processamento. O efeito de desempenho resultante às vezes é inaceitável, descartando o uso do espelhamento síncrono.
- O aumento da adoção do storage SSD de 100% significa que a latência de gravação adicional provavelmente será notada porque as expectativas de desempenho incluem centenas de milhares de IOPS e latência inferior a milissegundos. Aproveitar todos os benefícios do uso de SSDs de 100% pode exigir uma nova visita à estratégia de recuperação de desastres.
- Os conjuntos de dados continuam a crescer em termos de bytes, criando desafios com a garantia de largura de banda suficiente para sustentar a replicação síncrona.
- Os conjuntos de dados também crescem em termos de complexidade, criando desafios com o gerenciamento da replicação síncrona de grande escala.
- Estratégias baseadas na nuvem frequentemente envolvem maiores distâncias de replicação e latência, o que impede o uso do espelhamento síncrono.

A NetApp oferece soluções que incluem replicação síncrona para as demandas mais exigentes de recuperação de dados e soluções assíncronas que proporcionam melhor desempenho e flexibilidade. Além disso, a tecnologia NetApp se integra perfeitamente a muitas soluções de replicação de terceiros, como o Oracle DataGuard

Tempo de retenção

O último aspecto de uma estratégia de proteção de dados é o tempo de retenção de dados, que pode variar muito.

- Um requisito típico é de 14 dias de backups noturnos no local principal e 90 dias de backups armazenados em um local secundário.
- Muitos clientes criam arquivos trimestrais autônomos armazenados em Mídias diferentes.
- Um banco de dados constantemente atualizado pode não ter necessidade de dados históricos, e os backups precisam ser mantidos apenas por alguns dias.
- Os requisitos regulatórios podem exigir recuperação até o ponto de qualquer transação arbitrária em uma janela de 365 dias.

Disponibilidade do banco de dados

O ONTAP foi projetado para oferecer a máxima disponibilidade de banco de dados Oracle. Uma descrição completa dos recursos de alta disponibilidade do ONTAP está além do escopo deste documento. No entanto, assim como na proteção de dados, uma compreensão básica dessa funcionalidade é importante ao projetar uma infraestrutura de banco de dados.

Pares HA

A unidade básica de alta disponibilidade é o par de HA. Cada par contém links redundantes para suportar a replicação de dados para o NVRAM. NVRAM não é um cache de gravação. A RAM dentro do controlador serve como cache de gravação. O objetivo do NVRAM é registrar temporariamente os dados como uma salvaguarda contra falhas inesperadas do sistema. A este respeito, é semelhante a um log refazer banco de dados.

Tanto o NVRAM quanto o log refazer de banco de dados são usados para armazenar dados rapidamente, permitindo que as alterações aos dados sejam confirmadas o mais rápido possível. A atualização para os dados persistentes em unidades (ou datafiles) não ocorre até mais tarde durante um processo chamado ponto de verificação em plataformas ONTAP e na maioria dos bancos de dados. Nem os dados do NVRAM nem os logs de refazer do banco de dados são lidos durante operações normais.

Se um controlador falhar abruptamente, é provável que haja alterações pendentes armazenadas no NVRAM que ainda não foram gravadas nas unidades. O controlador do parceiro detecta a falha, controla as unidades e aplica as alterações necessárias que foram armazenadas no NVRAM.

Takeover e giveback

Takeover e giveback refere-se ao processo de transferência de responsabilidade por recursos de storage entre nós em um par de HA. Há dois aspectos para a aquisição e a giveback:

- Gestão da conectividade de rede que permite o acesso às unidades
- Gestão das próprias unidades

As interfaces de rede que suportam o tráfego CIFS e NFS são configuradas com um local de origem e failover. Uma aquisição inclui mover as interfaces de rede para sua casa temporária em uma interface física localizada na(s) mesma(s) sub-rede(s) do local original. A giveback inclui mover as interfaces de rede de volta para seus locais originais. O comportamento exato pode ser ajustado conforme necessário.

As interfaces de rede que suportam protocolos de bloco SAN, como iSCSI e FC, não são relocadas durante a aquisição e a giveback. Em vez disso, os LUNs devem ser provisionados com caminhos que incluam um par de HA completo, o que resulta em um caminho primário e um caminho secundário.



Caminhos adicionais para controladores adicionais também podem ser configurados para dar suporte à realocação de dados entre nós em um cluster maior, mas isso não faz parte do processo de HA.

O segundo aspecto da aquisição e da giveback é a transferência da propriedade do disco. O processo exato depende de vários fatores, incluindo o motivo da aquisição/giveback e as opções de linha de comando emitidas. O objetivo é realizar a operação da forma mais eficiente possível. Embora o processo geral possa parecer exigir vários minutos, o momento real em que a propriedade da unidade é transferida de nó para nó geralmente pode ser medido em segundos.

Tempo de takeover

A e/S do host tem uma pequena pausa na e/S durante as operações de takeover e giveback, mas não deve haver interrupção do aplicativo em um ambiente configurado corretamente. O processo de transição real no qual a e/S é atrasada geralmente é medido em segundos, mas o host pode exigir tempo adicional para reconhecer a alteração nos caminhos de dados e reenviar operações de e/S.

A natureza da interrupção depende do protocolo:

- Uma interface de rede que suporta tráfego NFS e CIFS emite uma solicitação ARP (Address Resolution Protocol) para a rede após a transição para um novo local físico. Isso faz com que os switches de rede atualizem suas tabelas de endereços de controle de acesso de Mídia (MAC) e retomem a e/S de processamento. A interrupção no caso de aquisição planejada e a giveback geralmente é medida em segundos e, em muitos casos, não é detectável. Algumas redes podem ser mais lentas para reconhecer completamente a mudança no caminho da rede, e alguns sistemas operacionais podem colocar em fila muita e/S em um tempo muito curto que deve ser tentado novamente. Isso pode prolongar o tempo necessário para retomar a I/O.

- Uma interface de rede que suporte protocolos SAN não faz a transição para um novo local. Um sistema operacional do host deve alterar o caminho ou os caminhos em uso. A pausa em I/O observada pelo host depende de vários fatores. Do ponto de vista do sistema de armazenamento, o período em que a e/S não pode ser atendida é de apenas alguns segundos. No entanto, diferentes sistemas operacionais de host podem exigir tempo adicional para permitir que uma e/S termine o tempo limite antes de tentar novamente. Os sistemas operacionais mais novos são mais capazes de reconhecer uma mudança de caminho muito mais rapidamente, mas os sistemas operacionais mais antigos geralmente exigem até 30 segundos para reconhecer uma mudança.

Os tempos de aquisição esperados durante os quais o sistema de storage não pode fornecer dados a um ambiente de aplicativo são mostrados na tabela abaixo. Não deve haver erros em qualquer ambiente de aplicativo, o controle deve aparecer como uma pequena pausa no processamento de e/S.

	NFS	AFF	ASA
Takeover planejado	15 seg	6-10 seg	2-3 seg
Takeover não planejado	30 seg	6-10 seg	2-3 seg

Somas de verificação e integridade de dados

O ONTAP e seus protocolos compatíveis incluem vários recursos que protegem a integridade do banco de dados Oracle, incluindo dados em repouso e dados transmitidos pela rede de rede.

A proteção de dados lógicos no ONTAP consiste em três requisitos principais:

- Os dados devem estar protegidos contra corrupção de dados.
- Os dados devem estar protegidos contra falha da unidade.
- As alterações nos dados devem ser protegidas contra perda.

Essas três necessidades são discutidas nas seções a seguir.

Corrupção de rede: Somas de verificação

O nível mais básico de proteção de dados é a soma de verificação, que é um código especial de detecção de erros armazenado junto com os dados. A corrupção de dados durante a transmissão da rede é detetada com o uso de uma soma de verificação e, em alguns casos, várias somas de verificação.

Por exemplo, um quadro FC inclui uma forma de checksum chamada de verificação de redundância cíclica (CRC) para garantir que a carga útil não esteja corrompida em trânsito. O transmissor envia os dados e o CRC dos dados. O receptor de um quadro FC recalcula o CRC dos dados recebidos para garantir que ele corresponda ao CRC transmitido. Se o CRC recém-calculado não corresponder ao CRC anexado ao quadro, os dados estarão corrompidos e o quadro FC será descartado ou rejeitado. Uma operação de e/S iSCSI inclui somas de verificação nas camadas TCP/IP e Ethernet e, para proteção adicional, também pode incluir proteção CRC opcional na camada SCSI. Qualquer corrupção de bits no fio é detetada pela camada TCP ou camada IP, o que resulta na retransmissão do pacote. Assim como no FC, erros no CRC SCSI resultam em uma rejeição ou rejeição da operação.

Drive corruption: Somas de verificação

As somas de verificação também são usadas para verificar a integridade dos dados armazenados nas unidades. Os blocos de dados gravados nas unidades são armazenados com uma função de checksum que

produz um número imprevisível vinculado aos dados originais. Quando os dados são lidos a partir da unidade, a soma de verificação é recalculada e comparada com a soma de verificação armazenada. Se não corresponder, os dados ficaram corrompidos e devem ser recuperados pela camada RAID.

Corrupção de dados: Gravações perdidas

Um dos tipos mais difíceis de detetar é uma gravação perdida ou extraviada. Quando uma escrita é reconhecida, ela deve ser escrita para a Mídia no local correto. A corrupção de dados no local é relativamente fácil de detetar usando uma soma de verificação simples armazenada com os dados. No entanto, se a gravação é simplesmente perdida, então a versão anterior dos dados ainda pode existir e a soma de verificação estaria correta. Se a gravação for colocada no local físico errado, a soma de verificação associada seria mais uma vez válida para os dados armazenados, mesmo que a gravação tenha destruído outros dados.

A solução para este desafio é a seguinte:

- Uma operação de gravação deve incluir metadados que indicam o local onde a gravação deve ser encontrada.
- Uma operação de gravação deve incluir algum tipo de identificador de versão.

Quando o ONTAP grava um bloco, ele inclui dados sobre a localização do bloco. Se uma leitura subsequente identificar um bloco, mas os metadados indicarem que ele pertence ao local 123 quando foi encontrado no local 456, então a gravação foi extraviada.

Detetar uma escrita totalmente perdida é mais difícil. A explicação é muito complicada, mas essencialmente o ONTAP está armazenando metadados de uma forma que uma operação de gravação resulta em atualizações para dois locais diferentes nas unidades. Se uma gravação for perdida, uma leitura posterior dos dados e metadados associados mostrará duas identidades de versão diferentes. Isso indica que a gravação não foi concluída pela unidade.

A corrupção de gravação perdida e extraviada é extremamente rara, mas, à medida que as unidades continuam a crescer e os conjuntos de dados aumentam a escala de exabytes, o risco aumenta. A detecção de gravações perdidas deve ser incluída em qualquer sistema de storage que suporte workloads de banco de dados.

Falhas de unidade: RAID, RAID DP e RAID-teC

Se um bloco de dados em uma unidade for descoberto como corrompido ou toda a unidade falhar e estiver totalmente indisponível, os dados devem ser reconstituídos. Isso é feito no ONTAP usando unidades de paridade. Os dados são distribuídos em várias unidades de dados e, em seguida, os dados de paridade são gerados. Este é armazenado separadamente dos dados originais.

O ONTAP usou originalmente o RAID 4, que usa uma única unidade de paridade para cada grupo de unidades de dados. O resultado foi que qualquer unidade do grupo poderia falhar sem resultar em perda de dados. Se a unidade de paridade falhar, nenhum dado será danificado e uma nova unidade de paridade poderá ser construída. Se uma única unidade de dados falhar, as unidades restantes poderão ser usadas com a unidade de paridade para regenerar os dados em falta.

Quando as unidades eram pequenas, a chance estatística de duas unidades falharem simultaneamente foi insignificante. À medida que as capacidades da unidade crescem, também tem o tempo necessário para reconstruir os dados após uma falha de unidade. Isso aumentou a janela na qual uma segunda falha da unidade resultaria em perda de dados. Além disso, o processo de reconstrução cria muitas e/S adicionais nas unidades sobreviventes. À medida que as unidades envelhecem, o risco de carga adicional que leva a uma segunda falha da unidade também aumenta. Finalmente, mesmo que o risco de perda de dados não tenha aumentado com o uso continuado do RAID 4, as consequências da perda de dados se tornariam mais graves. Quanto mais dados forem perdidos no caso de uma falha do grupo RAID, mais tempo levaria para recuperar

os dados, estendendo a interrupção dos negócios.

Esses problemas levaram a NetApp a desenvolver a tecnologia NetApp RAID DP, uma variante do RAID 6. Essa solução inclui duas unidades de paridade, o que significa que todas as duas unidades em um grupo RAID podem falhar sem criar perda de dados. As unidades continuaram a crescer em tamanho, o que levou a NetApp a desenvolver a tecnologia NetApp RAID-teC, que introduz uma terceira unidade de paridade.

Algumas práticas recomendadas de banco de dados históricos recomendam o uso do RAID-10, também conhecido como espelhamento distribuído. Isso oferece menos proteção de dados do que até mesmo o RAID DP porque há vários cenários de falha de dois discos, enquanto que no RAID DP não há nenhum.

Há também algumas práticas recomendadas de banco de dados históricos que indicam que RAID-10 é preferível às opções RAID-4/5/6 devido a problemas de desempenho. Essas recomendações às vezes se referem a uma penalidade de RAID. Embora essas recomendações geralmente estejam corretas, elas não são aplicáveis às implementações de RAID no ONTAP. O problema de desempenho está relacionado com a regeneração de paridade. Com as implementações tradicionais de RAID, o processamento das gravações aleatórias de rotina executadas por um banco de dados requer várias leituras de disco para regenerar os dados de paridade e concluir a gravação. A penalidade é definida como o IOPS de leitura adicional necessário para executar operações de gravação.

O ONTAP não incorre em uma penalidade de RAID porque as gravações são encenadas na memória em que a paridade é gerada e, em seguida, gravadas no disco como um único stripe RAID. Não são necessárias leituras para concluir a operação de gravação.

Em resumo, quando comparado ao RAID 10, o RAID DP e o RAID-teC oferecem muito mais capacidade utilizável, melhor proteção contra falha de unidade e nenhum sacrifício de performance.

Proteção contra falhas de hardware: NVRAM

Qualquer storage array que atenda a um workload de banco de dados precisa atender às operações de gravação o mais rápido possível. Além disso, uma operação de gravação deve ser protegida contra perda de um evento inesperado, como uma falha de energia. Isso significa que qualquer operação de gravação deve ser armazenada com segurança em pelo menos dois locais.

Os sistemas AFF e FAS contam com a NVRAM para atender a esses requisitos. O processo de escrita funciona da seguinte forma:

1. Os dados de gravação de entrada são armazenados na RAM.
2. As alterações que devem ser feitas nos dados no disco são registradas no NVRAM no nó local e no nó parceiro. O NVRAM não é um cache de gravação; em vez disso, é um diário semelhante a um log de refazer de banco de dados. Em condições normais, não é lido. Ele é usado apenas para recuperação, como após uma falha de energia durante o processamento de e/S.
3. A gravação é então reconhecida para o host.

O processo de gravação nesta fase é concluído do ponto de vista da aplicação, e os dados são protegidos contra perda porque são armazenados em dois locais diferentes. Eventualmente, as alterações são gravadas no disco, mas esse processo está fora da banda do ponto de vista do aplicativo, porque ocorre depois que a gravação é reconhecida e, portanto, não afeta a latência. Este processo é mais uma vez semelhante ao log de banco de dados. Uma alteração ao banco de dados é registrada nos logs de refazer o mais rápido possível, e a alteração é então reconhecida como comprometida. As atualizações para os datafiles ocorrem muito mais tarde e não afetam diretamente a velocidade de processamento.

No caso de uma falha do controlador, o controlador do parceiro assume a propriedade dos discos necessários e replica os dados registados no NVRAM para recuperar quaisquer operações de e/S que estivessem em

trânsito quando a falha ocorreu.

Proteção contra falhas de hardware: NVFAIL

Como discutido anteriormente, uma gravação não é reconhecida até que ela tenha sido registrada no NVRAM local e no NVRAM em pelo menos um outro controlador. Essa abordagem garante que uma falha de hardware ou falha de energia não resulte na perda de e/S em trânsito. Se o NVRAM local falhar ou a conectividade com o parceiro de HA falhar, esses dados em trânsito não serão mais espelhados.

Se o NVRAM local relatar um erro, o nó será encerrado. Esse desligamento resulta em failover para uma controladora de parceiro de HA. Nenhum dado é perdido porque o controlador que sofre a falha não reconheceu a operação de gravação.

O ONTAP não permite um failover quando os dados estão fora de sincronia, a menos que o failover seja forçado. Forçar uma alteração de condições desta forma reconhece que os dados podem ser deixados para trás no controlador original e que a perda de dados é aceitável.

Os bancos de dados são especialmente vulneráveis à corrupção se um failover for forçado porque os bancos de dados mantêm grandes caches internos de dados no disco. Se ocorrer um failover forçado, as alterações anteriormente confirmadas serão efetivamente descartadas. O conteúdo da matriz de armazenamento salta efetivamente para trás no tempo, e o estado do cache do banco de dados não reflete mais o estado dos dados no disco.

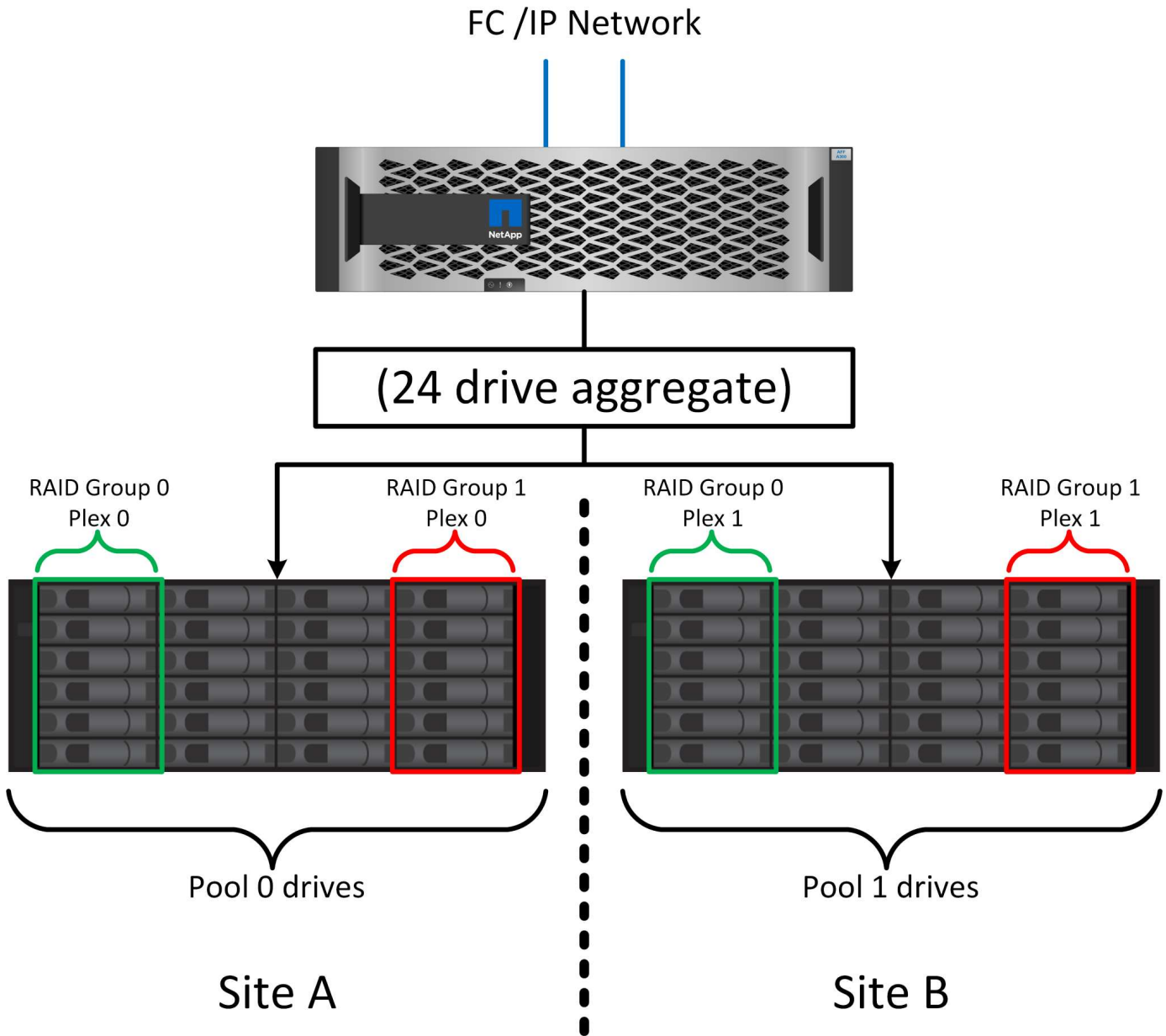
Para proteger os dados contra essa situação, o ONTAP permite que os volumes sejam configurados para proteção especial contra falhas do NVRAM. Quando acionado, esse mecanismo de proteção resulta em um volume entrando em um estado chamado NVFAIL. Esse estado resulta em erros de e/S que causam o desligamento de um aplicativo para que eles não usem dados obsoletos. Os dados não devem ser perdidos porque qualquer gravação reconhecida deve estar presente no storage array.

As próximas etapas usuais são para que um administrador desligue totalmente os hosts antes de colocar manualmente os LUNs e volumes novamente on-line. Embora essas etapas possam envolver algum trabalho, essa abordagem é a maneira mais segura de garantir a integridade dos dados. Nem todos os dados exigem essa proteção, e é por isso que o comportamento do NVFAIL pode ser configurado volume a volume.

Proteção contra falhas no local e no compartimento: SyncMirror e plexos

O SyncMirror é uma tecnologia de espelhamento que aprimora, mas não substitui, o RAID DP ou o RAID-teC. Ele espelha o conteúdo de dois grupos RAID independentes. A configuração lógica é a seguinte:

- As unidades são configuradas em dois pools com base no local. Um pool é composto por todas as unidades no local A, e o segundo pool é composto por todas as unidades no local B..
- Um pool comum de armazenamento, conhecido como agregado, é criado com base em conjuntos espelhados de grupos RAID. Um número igual de unidades é extraído de cada local. Por exemplo, um agregado SyncMirror de 20 unidades seria composto por 10 unidades do local A e 10 unidades do local B..
- Cada conjunto de unidades em um determinado local é configurado automaticamente como um ou mais grupos RAID-DP ou RAID-teC totalmente redundantes, independentemente do uso do espelhamento. Isso fornece proteção contínua de dados, mesmo após a perda de um site.



A figura acima ilustra um exemplo de configuração do SyncMirror. Um agregado de 24 unidades foi criado na controladora com 12 unidades de um compartimento alocado no local A e 12 unidades de um compartimento alocado no local B. as unidades foram agrupadas em dois grupos RAID espelhados. RAID Group 0 inclui um Plex de 6 unidades no local Um espelhado para um Plex de 6 unidades no local B. da mesma forma, RAID Group 1 inclui um Plex de 6 unidades no local Um espelhado para um Plex de 6 unidades no local B.

O SyncMirror normalmente é usado para fornecer espelhamento remoto com sistemas MetroCluster, com uma cópia dos dados em cada local. Ocasionalmente, ele tem sido usado para fornecer um nível extra de redundância em um único sistema. Em particular, ele fornece redundância em nível de prateleira. Um compartimento de unidades já contém fontes de alimentação duplas e controladores e é, no geral, pouco mais do que chapas metálicas, mas em alguns casos, a proteção extra pode ser garantida. Por exemplo, um cliente da NetApp implantou o SyncMirror para uma plataforma móvel de análise em tempo real usada durante testes automotivos. O sistema foi separado em dois racks físicos fornecidos por alimentações de energia independentes de sistemas UPS independentes.

Somas de verificação

O tópico de checksums é de particular interesse para DBAs que estão acostumados a usar backups de streaming Oracle RMAN migram para backups baseados em snapshot. Um recurso do RMAN é que ele executa verificações de integridade durante operações de backup. Embora esse recurso tenha algum valor, seu principal benefício é para um banco de dados que não é usado em um storage array moderno. Quando as unidades físicas são usadas para um banco de dados Oracle, é quase certo que a corrupção eventualmente ocorre à medida que as unidades envelhecem, um problema que é resolvido por somas de verificação baseadas em array em arrays de armazenamento reais.

Com um storage array real, a integridade de dados é protegida pelo uso de somas de verificação em vários níveis. Se os dados estiverem corrompidos em uma rede baseada em IP, a camada TCP (Transmission Control Protocol) rejeita os dados do pacote e solicita a retransmissão. O protocolo FC inclui somas de verificação, assim como os dados SCSI encapsulados. Depois que ele está no array, o ONTAP tem proteção RAID e checksum. A corrupção pode ocorrer, mas, como na maioria dos arrays corporativos, ela é detetada e corrigida. Normalmente, uma unidade inteira falha, solicitando uma reconstrução RAID e a integridade do banco de dados não é afetada. Ainda é possível que bytes individuais em uma unidade sejam danificados por radiação cósmica ou células flash com falha. Se isso acontecer, a verificação de paridade falharia, a unidade falharia e uma reconstrução RAID começaria. Mais uma vez, a integridade dos dados não é afetada. A linha final de defesa é o uso de somas de verificação. Se, por exemplo, um erro de firmware catastrófico em uma unidade corrompeu dados de uma forma que de alguma forma não fosse detetado por uma verificação de paridade RAID, a soma de verificação não corresponderia e o ONTAP impediria a transferência de um bloco corrompido antes que o banco de dados Oracle pudesse recebê-lo.

A arquitetura Oracle datafile e refazer log também foi projetada para fornecer o mais alto nível possível de integridade de dados, mesmo em circunstâncias extremas. No nível mais básico, os blocos Oracle incluem checksum e verificações lógicas básicas com quase todas as I/O. Se o Oracle não travou ou uma tablespace off-line, então os dados estão intactos. O grau de verificação da integridade dos dados é ajustável, e o Oracle também pode ser configurado para confirmar gravações. Como resultado, quase todos os cenários de falha e falha podem ser recuperados e, no caso extremamente raro de uma situação irrecuperável, a corrupção é imediatamente detetada.

A maioria dos clientes NetApp que usam bancos de dados Oracle descontinuam o uso de RMAN e outros produtos de backup após a migração para backups baseados em snapshot. Ainda existem opções nas quais o RMAN pode ser usado para executar a recuperação em nível de bloco com o SnapCenter. No entanto, no dia a dia, o RMAN, o NetBackup e outros produtos só são usados ocasionalmente para criar cópias de arquivamento mensais ou trimestrais.

Alguns clientes optam por executar `dbv` periodicamente para realizar verificações de integridade em seus bancos de dados existentes. NetApp desencoraja essa prática porque cria carga de e/S desnecessária. Como discutido acima, se o banco de dados não estava enfrentando problemas anteriormente, a chance de `dbv` detetar um problema é perto de zero, e este utilitário cria uma carga de e/S sequencial muito alta na rede e no sistema de armazenamento. A menos que haja razão para acreditar que existe corrupção, como a exposição a um bug conhecido da Oracle, não há razão para ser executado `dbv`.

Noções básicas de backup e recuperação

Backups baseados em snapshot

A base da proteção de dados de banco de dados Oracle no ONTAP é a tecnologia NetApp Snapshot.

Os valores-chave são os seguintes:

- **Simplicidade.** Um snapshot é uma cópia somente leitura do conteúdo de um contentor de dados em um determinado momento.
- **Eficiência.** Os instantâneos não requerem espaço no momento da criação. O espaço só é consumido quando os dados são alterados.
- **Capacidade de gerenciamento.** Uma estratégia de backup baseada em snapshots é fácil de configurar e gerenciar, pois os snapshots são uma parte nativa do sistema operacional de storage. Se o sistema de armazenamento estiver ligado, ele estará pronto para criar backups.
- **Escalabilidade.** É possível preservar até 1024 backups de um único contêiner de arquivos e LUNs. Para conjuntos de dados complexos, vários contêineres de dados podem ser protegidos por um único conjunto consistente de snapshots.
- O desempenho não é afetado, independentemente de um volume conter 1024 snapshots ou nenhum.

Embora muitos fornecedores de storage ofereçam tecnologia Snapshot, a tecnologia Snapshot no ONTAP é única e oferece benefícios significativos para ambientes de aplicações e bancos de dados empresariais:

- As cópias snapshot fazem parte do layout de arquivo em qualquer lugar (WAFL) subjacente. Eles não são uma tecnologia adicional ou externa. Isso simplifica o gerenciamento porque o sistema de storage é o sistema de backup.
- As cópias snapshot não afetam a performance, exceto em alguns casos de borda, como quando muitos dados são armazenados em snapshots que o sistema de storage subjacente preenche.
- O termo "grupo de consistência" é frequentemente usado para se referir a um agrupamento de objetos de armazenamento que são gerenciados como uma coleta consistente de dados. Um snapshot de um determinado volume ONTAP constitui um backup de grupo de consistência.

Os snapshots do ONTAP também são mais dimensionados do que a tecnologia da concorrência. Os clientes podem armazenar snapshots 5, 50 ou 500 sem afetar a performance. O número máximo de instantâneos atualmente permitido em um volume é 1024. Se a retenção adicional de snapshot for necessária, há opções para colocar os snapshots em cascata em volumes adicionais.

Como resultado, proteger um conjunto de dados hospedado no ONTAP é simples e altamente dimensionável. Os backups não exigem movimentação de dados, portanto, uma estratégia de backup pode ser adaptada às necessidades da empresa, em vez das limitações de taxas de transferência de rede, grande número de unidades de fita ou áreas de preparo de disco.

Um instantâneo é um backup?

Uma pergunta comumente feita sobre o uso de snapshots como estratégia de proteção de dados é o fato de que os dados "reais" e os dados instantâneos estão localizados nas mesmas unidades. A perda dessas unidades resultaria na perda dos dados primários e do backup.

Este é um problema válido. Os snapshots locais são usados para necessidades diárias de backup e recuperação e, nesse aspecto, o snapshot é um backup. Cerca de 99% de todos os cenários de recuperação em ambientes NetApp dependem de snapshots para atender até aos requisitos mais agressivos de rto.

No entanto, os snapshots locais nunca devem ser a única estratégia de backup. É por isso que a NetApp oferece tecnologia como replicação SnapMirror e SnapVault para replicar snapshots para um conjunto independente de unidades com rapidez e eficiência. Em uma solução de arquitetura adequada com snapshots e replicação de snapshot, o uso da fita pode ser minimizado para talvez um arquivamento trimestral ou eliminado totalmente.

Backups baseados em snapshot

Há muitas opções para o uso de cópias ONTAP Snapshot para proteger seus dados. Além disso, os snapshots são a base de muitos outros recursos da ONTAP, incluindo replicação, recuperação de desastres e clonagem. Uma descrição completa da tecnologia de instantâneos está além do escopo deste documento, mas as seções a seguir fornecem uma visão geral.

Existem duas abordagens principais para criar um instantâneo de um conjunto de dados:

- Backups consistentes com falhas
- Backups consistentes com aplicativos

Um backup consistente com falhas de um conjunto de dados refere-se à captura de toda a estrutura do conjunto de dados em um único ponto no tempo. Se o conjunto de dados for armazenado em um único volume, o processo será simples. É possível criar um Snapshot a qualquer momento. Se um conjunto de dados abranger volumes, é necessário criar um instantâneo de grupo de consistência (CG). Existem várias opções para criar snapshots CG, incluindo o software NetApp SnapCenter, recursos nativos do grupo de consistência do ONTAP e scripts mantidos pelo usuário.

Backups consistentes com falhas são usados principalmente quando a recuperação do ponto de backup é suficiente. Quando uma recuperação mais granular é necessária, backups consistentes com aplicações geralmente são necessários.

A palavra "consistente" em "consistente com aplicativos" é muitas vezes um erro. Por exemplo, colocar um banco de dados Oracle no modo de backup é referido como um backup consistente com aplicativos, mas os dados não são consistentes ou silenciosos de forma alguma. Os dados continuam a mudar durante todo o backup. Em contraste, a maioria dos backups MySQL e Microsoft SQL Server realmente silenciam os dados antes de executar o backup. A VMware pode ou não tornar certos arquivos consistentes.

Grupos de consistência

O termo "grupo de consistência" refere-se à capacidade de um storage array gerenciar vários recursos de armazenamento como uma única imagem. Por exemplo, um banco de dados pode consistir em 10 LUNs. O array deve ser capaz de fazer backup, restaurar e replicar esses 10 LUNs de maneira consistente. A restauração não é possível se as imagens dos LUNs não forem consistentes no ponto de backup. A replicação desses 10 LUNs requer que todas as réplicas sejam perfeitamente sincronizadas umas com as outras.

O termo "grupo de consistência" não é frequentemente usado quando se discute ONTAP porque consistência sempre foi uma função básica da arquitetura de volume e agregado dentro do ONTAP. Muitos outros storage arrays gerenciam LUNs ou sistemas de arquivos como unidades individuais. Eles poderiam, então, ser opcionalmente configurados como um "grupo de consistência" para fins de proteção de dados, mas esta é uma etapa extra na configuração.

O ONTAP sempre foi capaz de capturar imagens de dados locais e replicados consistentes. Embora os vários volumes em um sistema ONTAP não sejam geralmente formalmente descritos como um grupo de consistência, é isso que eles são. Um snapshot desse volume é uma imagem de grupo de consistência, a restauração para esse snapshot é uma restauração de grupo de consistência, e o SnapMirror e o SnapVault oferecem replicação de grupo de consistência.

Instantâneos do grupo de consistência

Os snapshots de grupos de consistência (snapshots cg) são uma extensão da tecnologia básica do ONTAP Snapshot. Uma operação de snapshot padrão cria uma imagem consistente de todos os dados em um único volume, mas às vezes é necessário criar um conjunto consistente de snapshots em vários volumes e até

mesmo em vários sistemas de storage. O resultado é um conjunto de instantâneos que podem ser usados da mesma forma que um instantâneo de apenas um volume individual. Eles podem ser usados para recuperação de dados local, replicados para fins de recuperação de desastres ou clonados como uma única unidade consistente.

O maior uso conhecido de snapshots cg é para um ambiente de banco de dados de aproximadamente 1PB TB de tamanho abrangendo 12 controladoras. Os snapshots cg criados neste sistema foram usados para backup, recuperação e clonagem.

Na maioria das vezes, quando um conjunto de dados abrange volumes e ordem de gravação deve ser preservado, um cg-snapshot é usado automaticamente pelo software de gerenciamento escolhido. Nesses casos, não há necessidade de entender os detalhes técnicos dos instantâneos cg. No entanto, há situações em que requisitos complicados de proteção de dados exigem controle detalhado sobre o processo de replicação e proteção de dados. Fluxos de trabalho de automação ou o uso de scripts personalizados para chamar APIs cg-snapshot são algumas das opções. Compreender a melhor opção e o papel do cg-snapshot requer uma explicação mais detalhada da tecnologia.

A criação de um conjunto de instantâneos cg é um processo de duas etapas:

1. Estabeleça cercas de gravação em todos os volumes de destino.
2. Crie instantâneos desses volumes enquanto estiver no estado cercado.

A esgrima de escrita é estabelecida em série. Isso significa que, à medida que o processo de esgrima é configurado em vários volumes, a e/S de gravação é congelada no primeiro volume da sequência, uma vez que continua a ser comprometida com volumes que aparecem mais tarde. Isso pode inicialmente parecer violar o requisito para que a ordem de gravação seja preservada, mas isso só se aplica a e/S que é emitida assincronamente no host e não depende de outras gravações.

Por exemplo, um banco de dados pode emitir muitas atualizações assíncronas de arquivos de dados e permitir que o sistema operacional reordene a e/S e as complete de acordo com sua própria configuração de agendador. A ordem deste tipo de e/S não pode ser garantida porque a aplicação e o sistema operativo já lançaram a exigência de preservar a ordem de escrita.

Como um exemplo de contador, a maioria das atividades de Registro de banco de dados é síncrona. O banco de dados não prossegue com outras gravações de log até que a e/S seja reconhecida, e a ordem dessas gravações deve ser preservada. Se uma e/S de log chegar em um volume cercado, ela não será reconhecida e a aplicação será bloqueada em outras gravações. Da mesma forma, a e/S de metadados do sistema de arquivos geralmente é síncrona. Por exemplo, uma operação de exclusão de arquivos não deve ser perdida. Se um sistema operacional com um sistema de arquivos xfs excluísse um arquivo e a e/S que atualizasse os metadados do sistema de arquivos xfs para remover a referência a esse arquivo aterrado em um volume cercado, a atividade do sistema de arquivos pausará. Isso garante a integridade do sistema de arquivos durante operações cg-snapshot.

Depois que o grima de gravação é configurado nos volumes de destino, eles estão prontos para a criação de snapshot. Os instantâneos não precisam ser criados exatamente ao mesmo tempo porque o estado dos volumes é congelado de um ponto de vista de gravação dependente. Para se proteger contra uma falha na aplicação criando os instantâneos cg, a esgrima de gravação inicial inclui um tempo limite configurável no qual o ONTAP libera automaticamente a esgrima e retoma o processamento de gravação após um número definido de segundos. Se todos os instantâneos forem criados antes do período de tempo limite expirar, o conjunto de instantâneos resultante será um grupo de consistência válido.

Ordem de escrita dependente

Do ponto de vista técnico, a chave para um grupo de consistência é preservar a ordem de gravação e, especificamente, a ordem de gravação dependente. Por exemplo, um banco de dados gravando em 10 LUNs

grava simultaneamente em todos eles. Muitas gravações são emitidas assincronamente, o que significa que a ordem em que são concluídas não é importante e a ordem real que são concluídas varia de acordo com o sistema operacional e o comportamento da rede.

Algumas operações de gravação devem estar presentes no disco antes que o banco de dados possa continuar com gravações adicionais. Essas operações críticas de gravação são chamadas de gravações dependentes. A e/S de gravação subsequente depende da presença dessas gravações no disco. Qualquer snapshot, recuperação ou replicação desses 10 LUNs deve garantir que a ordem de gravação dependente seja garantida. As atualizações do sistema de arquivos são outro exemplo de gravações dependentes da ordem de gravação. A ordem em que as alterações do sistema de arquivos são feitas deve ser preservada ou todo o sistema de arquivos pode ficar corrompido.

Estratégias

Há duas abordagens principais para backups baseados em snapshot:

- Backups consistentes com falhas
- Backups ativos protegidos por snapshot

Um backup consistente com falhas de um banco de dados refere-se à captura de toda a estrutura do banco de dados, incluindo datafiles, logs de refazer e arquivos de controle, em um único ponto no tempo. Se o banco de dados for armazenado em um único volume, o processo será simples; uma captura Instantânea pode ser criada a qualquer momento. Se um banco de dados abranger volumes, um snapshot de grupo de consistência (CG) deve ser criado. Existem várias opções para criar snapshots CG, incluindo o software NetApp SnapCenter, recursos nativos do grupo de consistência do ONTAP e scripts mantidos pelo usuário.

Os backups Snapshot consistentes com falhas são usados principalmente quando a recuperação do ponto de backup é suficiente. Registros de arquivo podem ser aplicados em algumas circunstâncias, mas quando uma recuperação pontual mais granular é necessária, um backup on-line é preferível.

O procedimento básico para um backup on-line baseado em snapshot é o seguinte:

1. Coloque a base de dados `backup` no modo.
2. Crie um instantâneo de todos os volumes que hospedam datafiles.
3. Sair `backup` do modo.
4. Execute o comando `alter system archive log current` para forçar o arquivamento de logs.
5. Crie instantâneos de todos os volumes que hospedam os logs do arquivo.

Este procedimento produz um conjunto de instantâneos contendo datafiles no modo de backup e os logs críticos de arquivo gerados no modo de backup. Estes são os dois requisitos para recuperar um banco de dados. Arquivos como arquivos de controle também devem ser protegidos por conveniência, mas o único requisito absoluto é a proteção para arquivos de dados e logs de arquivo.

Embora clientes diferentes possam ter estratégias muito diferentes, quase todas essas estratégias são baseadas nos mesmos princípios descritos abaixo.

Recuperação baseada em Snapshot

Ao projetar layouts de volume para bancos de dados Oracle, a primeira decisão é se usar a tecnologia NetApp SnapRestore baseada em volume (VBSR).

O SnapRestore baseado em volume permite que um volume seja revertido quase instantaneamente para um ponto anterior no tempo. Como todos os dados no volume são revertidos, o VBSR pode não ser apropriado

para todos os casos de uso. Por exemplo, se um banco de dados inteiro, incluindo datafiles, refazer logs e Registros de arquivamento, for armazenado em um único volume e esse volume for restaurado com VBSR, os dados serão perdidos porque o Registro de arquivo mais recente e os dados de refazer são descartados.

VBSR não é necessário para restaurar. Muitos bancos de dados podem ser restaurados usando o SnapRestore de arquivo único (SFSR) baseado em arquivo ou simplesmente copiando arquivos do snapshot de volta para o sistema de arquivos ativo.

O VBSR é preferido quando um banco de dados é muito grande ou quando ele deve ser recuperado o mais rápido possível, e o uso do VBSR requer isolamento dos arquivos de dados. Em um ambiente NFS, os arquivos de dados de um determinado banco de dados devem ser armazenados em volumes dedicados que não estejam contaminados por qualquer outro tipo de arquivo. Em um ambiente SAN, os arquivos de dados devem ser armazenados em LUNs dedicados em volumes dedicados. Se um gerenciador de volumes for usado (incluindo Oracle Automatic Storage Management [ASM]), o grupo de discos também deve ser dedicado a arquivos de dados.

Isolar datafiles desta maneira permite que eles sejam revertidos para um estado anterior sem danificar outros sistemas de arquivos.

Reserva do Snapshot

Para cada volume com dados Oracle em um ambiente SAN, o `percent-snapshot-space` deve ser definido como zero porque reservar espaço para um snapshot em um ambiente LUN não é útil. Se a reserva fracionária estiver definida como 100, um instantâneo de um volume com LUNs requer espaço livre suficiente no volume, excluindo a reserva instantânea, para absorver 100% de rotatividade de todos os dados. Se a reserva fracionária for definida para um valor mais baixo, uma quantidade correspondente menor de espaço livre será necessária, mas sempre exclui a reserva instantânea. Isso significa que o espaço de reserva do snapshot em um ambiente LUN é desperdiçado.

Em um ambiente NFS, há duas opções:

- Defina o `percent-snapshot-space` com base no consumo de espaço esperado do instantâneo.
- Defina o `percent-snapshot-space` como zero e gerencie o consumo de espaço ativo e instantâneo coletivamente.

Com a primeira opção, `percent-snapshot-space` é definido para um valor diferente de zero, normalmente em torno de 20%. Este espaço é então escondido do usuário. Esse valor não cria, no entanto, um limite de utilização. Se um banco de dados com uma reserva de 20% sofrer 30% de rotatividade, o espaço instantâneo pode crescer além dos limites da reserva de 20% e ocupar espaço não reservado.

O principal benefício de definir uma reserva para um valor como 20% é verificar se algum espaço está sempre disponível para instantâneos. Por exemplo, um volume 1TB com uma reserva de 20% permitiria apenas que um administrador de banco de dados (DBA) armazenasse 800GB TB de dados. Essa configuração garante pelo menos 200GBMB de espaço para consumo de snapshot.

```
`percent-snapshot-space`Quando está definido como zero, todo o espaço no volume está disponível para o usuário final, o que proporciona melhor visibilidade. O DBA deve entender que, se ele ou ela vir um volume de 1TB TB que aproveita snapshots, esse 1TB TB de espaço será compartilhado entre dados ativos e a rotatividade do Snapshot.
```

Não há preferência clara entre a opção um e a opção dois entre os usuários finais.

ONTAP e snapshots de terceiros

O Oracle Doc ID 604683,1 explica os requisitos para suporte a instantâneos de terceiros e as várias opções disponíveis para operações de backup e restauração.

O fornecedor terceirizado deve garantir que os snapshots da empresa estejam em conformidade com os seguintes requisitos:

- Os snapshots devem ser integrados às operações de restauração e recuperação recomendadas pela Oracle.
- Os snapshots devem ser consistentes com falhas de banco de dados no ponto do snapshot.
- A ordenação de gravação é preservada para cada arquivo dentro de um snapshot.

Os produtos de gerenciamento ONTAP e NetApp da Oracle atendem a esses requisitos.

SnapRestore

A restauração rápida de dados no ONTAP a partir de um snapshot é fornecida pela tecnologia NetApp SnapRestore.

Quando um conjunto de dados essencial não está disponível, as operações de negócios essenciais estão inativas. As fitas podem quebrar, e até mesmo as restaurações de backups baseados em disco podem ser lentas para serem transferidas pela rede. O SnapRestore evita esses problemas fornecendo restauração quase instantânea de conjuntos de dados. Mesmo os bancos de dados em escala de petabyte podem ser completamente restaurados com apenas alguns minutos de esforço.

Existem duas formas de SnapRestore - baseado em arquivo/LUN e baseado em volume.

- Arquivos individuais ou LUNs podem ser restaurados em segundos, seja um arquivo 2TB LUN ou 4KB.
- O volume de arquivos ou LUNs pode ser restaurado em segundos, seja 10GB ou 100TB TB de dados.

Um "contentor de arquivos ou LUNs" normalmente se referiria a um FlexVol volume. Por exemplo, você pode ter 10 LUNs que compõem um grupo de discos LVM em um único volume ou um volume pode armazenar os diretórios base NFS de usuários do 1000. Em vez de executar uma operação de restauração para cada arquivo individual ou LUN, você pode restaurar todo o volume como uma única operação. Esse processo também funciona com contêineres com escalabilidade horizontal que incluem vários volumes, como um FlexGroup ou um Grupo de consistência do ONTAP.

O motivo pelo qual o SnapRestore funciona tão rápido e eficientemente é devido à natureza de um snapshot, que é essencialmente uma visualização paralela somente leitura do conteúdo de um volume em um determinado momento. Os blocos ativos são os blocos reais que podem ser alterados, enquanto o snapshot é uma visualização somente leitura no estado dos blocos que constituem os arquivos e LUNs no momento em que o snapshot foi criado.

O ONTAP só permite acesso somente leitura a dados instantâneos, mas os dados podem ser reativados com o SnapRestore. O instantâneo é reativado como uma visualização de leitura e gravação dos dados, retornando os dados ao seu estado anterior. O SnapRestore pode operar no volume ou no nível do arquivo. A tecnologia é essencialmente a mesma com algumas pequenas diferenças de comportamento.

Volume SnapRestore

O SnapRestore baseado em volume retorna todo o volume de dados para um estado anterior. Essa operação não requer movimentação de dados, o que significa que o processo de restauração é essencialmente instantâneo, embora a operação de API ou CLI possa levar alguns segundos para ser processada. Restaurar

1GB TB de dados não é mais complicado ou demorado do que restaurar 1PB TB de dados. Essa funcionalidade é a principal razão pela qual muitos clientes empresariais migram para os sistemas de storage da ONTAP. Ele fornece um rto medido em segundos, até mesmo para os maiores conjuntos de dados.

Uma desvantagem para o SnapRestore baseado em volume é causada pelo fato de que as alterações dentro de um volume são cumulativas ao longo do tempo. Portanto, cada snapshot e os dados de arquivo ativos dependem das alterações que levam a esse ponto. Reverter um volume para um estado anterior significa descartar todas as alterações subsequentes que foram feitas aos dados. O que é menos óbvio, no entanto, é que isso inclui instantâneos criados posteriormente. Isso nem sempre é desejável.

Por exemplo, um SLA de retenção de dados pode especificar 30 dias de backups noturnos. Restaurar um conjunto de dados para um instantâneo criado há cinco dias com o volume SnapRestore descartaria todos os snapshots criados nos cinco dias anteriores, violando o SLA.

Existem várias opções disponíveis para resolver esta limitação:

1. Os dados podem ser copiados de um snapshot anterior, em vez de executar um SnapRestore de todo o volume. Esse método funciona melhor com conjuntos de dados menores.
2. Um snapshot pode ser clonado em vez de restaurado. A limitação a essa abordagem é que o snapshot de origem é uma dependência do clone. Portanto, ele não pode ser excluído a menos que o clone também seja excluído ou seja dividido em um volume independente.
3. Uso de SnapRestore baseado em arquivos.

File SnapRestore (ficheiro)

O SnapRestore baseado em arquivo é um processo de restauração mais granular baseado em snapshot. Em vez de reverter o estado de um volume inteiro, o estado de um arquivo individual ou LUN é revertido. Não é necessário eliminar instantâneos, nem esta operação cria qualquer dependência de um instantâneo anterior. O ficheiro ou LUN fica imediatamente disponível no volume ativo.

Nenhuma movimentação de dados é necessária durante uma restauração do SnapRestore de um arquivo ou LUN. No entanto, algumas atualizações internas de metadados são necessárias para refletir o fato de que os blocos subjacentes em um arquivo ou LUN agora existem em um snapshot e no volume ativo. Não deve haver efeito no desempenho, mas esse processo bloqueia a criação de snapshots até que ele esteja concluído. A taxa de processamento é de aproximadamente 5Gbps (18TBMB/hora) com base no tamanho total dos arquivos restaurados.

Backups online

Dois conjuntos de dados são necessários para proteger e recuperar um banco de dados Oracle no modo de backup. Note que esta não é a única opção de backup Oracle, mas é a mais comum.

- Um instantâneo dos arquivos de dados no modo de backup
- Os logs de arquivo criados enquanto os datafiles estavam no modo de backup

Se a recuperação completa, incluindo todas as transações confirmadas, é necessário um terceiro item:

- Um conjunto de registos de refazer atuais

Existem várias maneiras de impulsionar a recuperação de um backup on-line. Muitos clientes restauram snapshots usando a CLI do ONTAP e, em seguida, usando o Oracle RMAN ou sqlplus para concluir a recuperação. Isso é especialmente comum com grandes ambientes de produção em que a probabilidade e a

frequência de restaurações de banco de dados são extremamente baixas e qualquer procedimento de restauração é tratado por um DBA qualificado. Para automação completa, soluções como o NetApp SnapCenter incluem um plug-in Oracle com interfaces gráficas e de linha de comando.

Alguns clientes de grande escala adotaram uma abordagem mais simples, configurando scripts básicos nos hosts para colocar os bancos de dados no modo de backup em um momento específico, em preparação para um snapshot agendado. Por exemplo, programe o comando `alter database begin backup` às 23:58, `alter database end backup` às 00:02 e, em seguida, programe instantâneos diretamente no sistema de armazenamento à meia-noite. O resultado é uma estratégia de backup simples e altamente dimensionável que não requer software ou licenças externos.

Layout de dados

O layout mais simples é isolar datafiles em um ou mais volumes dedicados. Eles devem ser não contaminados por qualquer outro tipo de arquivo. Isso é para garantir que os volumes de arquivo de dados possam ser restaurados rapidamente através de uma operação SnapRestore sem destruir um log refazer importante, controlfile ou log de arquivo.

A SAN tem requisitos semelhantes para isolamento de arquivos de dados dentro de volumes dedicados. Com um sistema operacional como o Microsoft Windows, um único volume pode conter vários LUNs de arquivo de dados, cada um com um sistema de arquivos NTFS. Com outros sistemas operacionais, geralmente há um gerenciador de volumes lógico. Por exemplo, com o Oracle ASM, a opção mais simples seria limitar os LUNs de um grupo de discos ASM a um único volume que pode ser feito backup e restaurado como uma unidade. Se forem necessários volumes adicionais por motivos de gerenciamento de performance ou capacidade, a criação de um grupo de discos adicional no novo volume resultará em um gerenciamento mais simples.

Se essas diretrizes forem seguidas, os snapshots poderão ser agendados diretamente no sistema de storage sem a necessidade de realizar um snapshot de grupo de consistência. A razão é que os backups Oracle não exigem que os datafiles sejam copiados ao mesmo tempo. O procedimento de backup on-line foi projetado para permitir que os arquivos de dados continuem sendo atualizados, pois são transmitidos lentamente para a fita ao longo de horas.

Uma complicação surge em situações como o uso de um grupo de discos ASM que é distribuído entre volumes. Nesses casos, um `cg-snapshot` deve ser executado para garantir que os metadados ASM sejam consistentes em todos os volumes constituintes.

Atenção: Verifique se o ASM `spfile` e `passwd` os arquivos não estão no grupo de discos que hospeda os arquivos de dados. Isso interfere na capacidade de restaurar seletivamente datafiles e apenas datafiles.

Procedimento de recuperação local – NFS

Este procedimento pode ser conduzido manualmente ou através de uma aplicação como o SnapCenter. O procedimento básico é o seguinte:

1. Encerre o banco de dados.
2. Recupere o(s) volume(s) de arquivo de dados para o instantâneo imediatamente antes do ponto de restauração desejado.
3. Reproduza registros de arquivo até ao ponto pretendido.
4. Repita os logs atuais de refazer se a recuperação completa for desejada.

Este procedimento pressupõe que os registros de arquivo desejados ainda estão presentes no sistema de ficheiros ativo. Se não estiverem, os logs do arquivo devem ser restaurados ou `rman/sqlplus` podem ser direcionados para os dados no diretório instantâneo.

Além disso, para bancos de dados menores, os arquivos de dados podem ser recuperados por um usuário final diretamente . `snapshot` do diretório sem a ajuda de ferramentas de automação ou administradores de armazenamento para executar um `snapprestore` comando.

Procedimento de recuperação local – SAN

Este procedimento pode ser conduzido manualmente ou através de uma aplicação como o SnapCenter. O procedimento básico é o seguinte:

1. Encerre o banco de dados.
2. Quiesce o(s) grupo(s) de discos que hospedam os arquivos de dados. O procedimento varia consoante o gestor de volume lógico escolhido. Com ASM, o processo requer a desmontagem do grupo de discos. Com o Linux, os sistemas de arquivos devem ser desmontados e os volumes lógicos e grupos de volumes devem ser desativados. O objetivo é parar todas as atualizações no grupo de volume alvo a serem restauradas.
3. Restaure os grupos de discos de arquivo de dados para o instantâneo imediatamente antes do ponto de restauração desejado.
4. Reative os grupos de discos recentemente restaurados.
5. Reproduza registos de arquivo até ao ponto pretendido.
6. Repita todos os logs de refazer se a recuperação completa for desejada.

Este procedimento pressupõe que os registos de arquivo desejados ainda estão presentes no sistema de ficheiros ativo. Se não estiverem, os registos de arquivo devem ser restaurados colocando os LUNs de registo de arquivo offline e executando uma restauração. Este também é um exemplo no qual dividir os logs de arquivo em volumes dedicados é útil. Se os logs de arquivo compartilharem um grupo de volumes com os logs de refazer, os logs de refazer devem ser copiados em outro lugar antes da restauração do conjunto geral de LUNs. Esta etapa impede a perda dessas transações finais registradas.

Backups otimizados para Storage Snapshot

Backup e recuperação baseados em snapshot se tornaram ainda mais simples quando o Oracle 12c foi lançado porque não há necessidade de colocar um banco de dados no modo hot backup. O resultado é a capacidade de agendar backups baseados em snapshot diretamente em um sistema de storage e ainda preservar a capacidade de executar recuperação completa ou pontual.

Embora o procedimento de recuperação de hot backup seja mais familiar aos DBAs, há muito tempo foi possível usar snapshots que não foram criados enquanto o banco de dados estava no modo hot backup. Etapas manuais extras foram necessárias com o Oracle 10gi e 11gi durante a recuperação para tornar o banco de dados consistente. Com o Oracle 12ci, `sqlplus` e `rman` conter a lógica extra para reproduzir logs de arquivo em backups de arquivos de dados que não estavam no modo de backup ativo.

Como discutido anteriormente, a recuperação de um hot backup baseado em snapshot requer dois conjuntos de dados:

- Um instantâneo dos arquivos de dados criados no modo de backup
- Os logs de arquivo gerados enquanto os datafiles estavam no modo hot backup

Durante a recuperação, o banco de dados lê metadados dos arquivos de dados para selecionar os logs de arquivo necessários para recuperação.

A recuperação otimizada para snapshot de storage requer conjuntos de dados ligeiramente diferentes para alcançar os mesmos resultados:

- Um instantâneo dos arquivos de dados, além de um método para identificar a hora em que o snapshot foi criado
- Arquive logs do tempo do ponto de verificação mais recente do arquivo de dados até a hora exata do instantâneo

Durante a recuperação, o banco de dados lê metadados dos arquivos de dados para identificar o Registro de arquivo mais antigo necessário. A recuperação completa ou pontual pode ser realizada. Ao executar uma recuperação pontual, é fundamental saber o tempo do snapshot dos arquivos de dados. O ponto de recuperação especificado deve ser após o tempo de criação dos instantâneos. A NetApp recomenda adicionar pelo menos alguns minutos à hora do instantâneo para contabilizar a variação do relógio.

Para obter detalhes completos, consulte a documentação da Oracle sobre o tópico "recuperação usando Otimização de Snapshot de armazenamento" disponível em várias versões da documentação do Oracle 12c. Além disso, consulte Oracle Document ID Doc ID 604683,1 sobre o suporte a snapshots de terceiros da Oracle.

Layout de dados

O layout mais simples é isolar os arquivos de dados em um ou mais volumes dedicados. Eles devem ser não contaminados por qualquer outro tipo de arquivo. Isso é para garantir que os volumes de arquivo de dados possam ser restaurados rapidamente com uma operação SnapRestore sem destruir um log de refazer importante, controlfile ou log de arquivo.

A SAN tem requisitos semelhantes para isolamento de arquivos de dados dentro de volumes dedicados. Com um sistema operacional como o Microsoft Windows, um único volume pode conter vários LUNs de arquivo de dados, cada um com um sistema de arquivos NTFS. Com outros sistemas operacionais, geralmente há um gerenciador de volume lógico também. Por exemplo, com o Oracle ASM, a opção mais simples seria limitar grupos de discos a um único volume que pode ser feito backup e restaurado como uma unidade. Se forem necessários volumes adicionais por motivos de gerenciamento de performance ou capacidade, criar um grupo de discos adicional no novo volume resulta em gerenciamento mais fácil.

Se essas diretrizes forem seguidas, os snapshots poderão ser agendados diretamente no ONTAP sem a necessidade de realizar um snapshot de grupo de consistência. O motivo é que backups otimizados para snapshot não exigem que sejam feitos backup de dados ao mesmo tempo.

Uma complicação surge em situações como um grupo de discos ASM que é distribuído entre volumes. Nesses casos, um cg-snapshot deve ser executado para garantir que os metadados ASM sejam consistentes em todos os volumes constituintes.

[Nota]Verifique se os arquivos ASM spfile e passwd não estão no grupo de discos que hospeda os arquivos de dados. Isso interfere na capacidade de restaurar seletivamente datafiles e apenas datafiles.

Procedimento de recuperação local – NFS

Este procedimento pode ser conduzido manualmente ou através de uma aplicação como o SnapCenter. O procedimento básico é o seguinte:

1. Encerre o banco de dados.
2. Recupere o(s) volume(s) de arquivo de dados para o instantâneo imediatamente antes do ponto de restauração desejado.
3. Reproduza registros de arquivo até ao ponto pretendido.

Este procedimento pressupõe que os registos de arquivo desejados ainda estão presentes no sistema de ficheiros ativo. Se não estiverem, os registos de arquivo têm de ser restaurados ou `rman sqlplus` podem ser direcionados para os dados no `.snapshot` diretório.

Além disso, para bancos de dados menores, os arquivos de dados podem ser recuperados por um usuário final diretamente `.snapshot` do diretório sem a ajuda de ferramentas de automação ou um administrador de armazenamento para executar um comando SnapRestore.

Procedimento de recuperação local – SAN

Este procedimento pode ser conduzido manualmente ou através de uma aplicação como o SnapCenter. O procedimento básico é o seguinte:

1. Encerre o banco de dados.
2. Quiesce o(s) grupo(s) de discos que hospedam os arquivos de dados. O procedimento varia consoante o gestor de volume lógico escolhido. Com ASM, o processo requer a desmontagem do grupo de discos. Com o Linux, os sistemas de arquivos devem ser desmontados e os volumes lógicos e grupos de volumes são desativados. O objetivo é parar todas as atualizações no grupo de volume alvo a serem restauradas.
3. Restaure os grupos de discos de arquivo de dados para o instantâneo imediatamente antes do ponto de restauração desejado.
4. Reative os grupos de discos recentemente restaurados.
5. Reproduza registos de arquivo até ao ponto pretendido.

Este procedimento pressupõe que os registos de arquivo desejados ainda estão presentes no sistema de ficheiros ativo. Se não estiverem, os registos de arquivo devem ser restaurados colocando os LUNs de registo de arquivo offline e executando uma restauração. Este também é um exemplo no qual dividir os logs de arquivo em volumes dedicados é útil. Se os logs do arquivo compartilharem um grupo de volumes com os logs de refazer, os logs de refazer devem ser copiados em outro lugar antes da restauração do conjunto geral de LUNs para evitar perder as transações registradas finais.

Exemplo de recuperação completa

Suponha que os arquivos de dados foram corrompidos ou destruídos e a recuperação completa é necessária. O procedimento para o fazer é o seguinte:

```

[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>

```

Exemplo de recuperação pontual

Todo o procedimento de recuperação é um único comando: `recover automatic`.

Se a recuperação pontual for necessária, o carimbo de data/hora do(s) instantâneo(s) deve(m) ser conhecido(s) e pode(m) ser identificado(s) da seguinte forma:

```

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

```

A hora de criação do instantâneo é listada como 9th de Março e 10:10:06. Para estar seguro, um minuto é adicionado à hora do instantâneo:

```

[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';

```

A recuperação agora é iniciada. Ele especificou um tempo instantâneo de 10:11:00, um minuto após o tempo gravado para contabilizar a possível variação do relógio e um tempo de recuperação alvo de 10:44. Em seguida, sqlplus solicita os logs de arquivo necessários para alcançar o tempo de recuperação desejado de 10:44.

```
ORA-00279: change 551760 generated at 03/09/2017 05:06:07 needed for
thread 1
ORA-00289: suggestion : /orlogs_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 : /orlogs_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 : /orlogs_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 : /orlogs_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>
```



A recuperação completa de um banco de dados usando snapshots usando o `recover automatic` comando não requer licenciamento específico, mas a recuperação pontual usando `snapshot time` requer a licença Oracle Advanced Compression.

Ferramentas de gerenciamento e automação de banco de dados

O principal valor do ONTAP em um ambiente de banco de dados Oracle vem das principais tecnologias da ONTAP, como cópias Snapshot instantâneas, replicação simples do SnapMirror e criação eficiente de volumes do FlexClone.

Em alguns casos, a configuração simples desses principais recursos diretamente no ONTAP atende aos requisitos, mas as necessidades mais complicadas exigem uma camada de orquestração.

SnapCenter

O SnapCenter é o principal produto de proteção de dados da NetApp. Em um nível muito baixo, ele é

semelhante aos produtos SnapManager em termos de como executa backups de bancos de dados. No entanto, ele foi criado do zero para fornecer um painel único para gerenciamento de proteção de dados em sistemas de storage da NetApp.

O SnapCenter inclui funções básicas, como backups e restaurações baseados em snapshot, replicação SnapMirror e SnapVault e outros recursos necessários para operar em escala para empresas de grande porte. Esses recursos avançados incluem funcionalidade de controle de acesso baseado em funções (RBAC) expandida, APIs RESTful para integração com produtos de orquestração de terceiros, gerenciamento central sem interrupções de plug-ins do SnapCenter em hosts de banco de dados e uma interface de usuário projetada para ambientes de escala de nuvem.

DESCANSO

O ONTAP também contém um conjunto de APIs RESTful. Isso permite que fornecedores terceirizados de 3rd criem proteção de dados e outros aplicativos de gerenciamento com profunda integração com o ONTAP. Além disso, a API RESTful é fácil de consumir por clientes que desejam criar seus próprios fluxos de trabalho e utilitários de automação.

Recuperação de desastres Oracle

Visão geral

Recuperação de desastres refere-se à restauração de serviços de dados após um evento catastrófico, como um incêndio que destrói um sistema de storage ou até mesmo um local inteiro.



Esta documentação substitui os relatórios técnicos publicados anteriormente *TR-4591: Oracle Data Protection* e *TR-4592: Oracle no MetroCluster*.

A recuperação de desastres pode ser realizada por replicação simples de dados usando o SnapMirror, é claro, com muitos clientes atualizando réplicas espelhadas quantas vezes por hora.

Para a maioria dos clientes, a recuperação de desastres requer mais do que apenas a posse de uma cópia remota de dados. Isso exige a capacidade de usar esses dados rapidamente. A NetApp oferece duas tecnologias que atendem a essa necessidade: MetroCluster e SnapMirror active Sync

O MetroCluster se refere ao ONTAP em uma configuração de hardware que inclui armazenamento de baixo nível espelhado e vários recursos adicionais. Soluções integradas, como o MetroCluster, simplificam as infraestruturas de virtualização, aplicações e banco de dados complicadas e com escalabilidade horizontal atuais. Ele substitui vários produtos e estratégias de proteção de dados externos por um storage array simples e central. Ele também fornece backup integrado, recuperação, recuperação de desastres e alta disponibilidade (HA) em um único sistema de storage em cluster.

A sincronização ativa do SnapMirror (SM-as) é baseada no SnapMirror síncrono. Com o MetroCluster, cada controlador ONTAP é responsável por replicar os dados da unidade para um local remoto. Com o SnapMirror active Sync, você tem essencialmente dois sistemas ONTAP diferentes que mantêm cópias independentes dos seus dados LUN, mas cooperam para apresentar uma única instância desse LUN. Do ponto de vista do host, é uma única entidade LUN.

Comparação de SM-as e MCC

O SM-as e o MetroCluster são semelhantes na funcionalidade geral, mas há diferenças importantes na maneira como a replicação RPO-0 foi implementada e como ela é gerenciada. O SnapMirror assíncrono e

síncrono também pode ser usado como parte de um plano de recuperação de desastres, mas não foram desenvolvidos como tecnologias de repliação de HA.

- Uma configuração do MetroCluster é mais como um cluster integrado com nós distribuídos entre locais. O SM-as se comporta como dois clusters independentes que estão cooperando no fornecimento de LUNs replicados de forma síncrona RPO igual a 0.
- Os dados em uma configuração do MetroCluster só podem ser acessados de um determinado site a qualquer momento. Uma segunda cópia dos dados está presente no site oposto, mas os dados são passivos. Ele não pode ser acessado sem um failover do sistema de storage.
- O espelhamento de desempenho MetroCluster e SM-as ocorre em níveis diferentes. O espelhamento MetroCluster é executado na camada RAID. Os dados de baixo nível são armazenados em um formato espelhado usando SyncMirror. O uso do espelhamento é praticamente invisível nas camadas de LUN, volume e protocolo.
- Em contraste, o espelhamento SM-as ocorre na camada de protocolo. Os dois clusters são, no geral, clusters independentes. Depois que as duas cópias dos dados estiverem sincronizadas, os dois clusters só precisam espelhar gravações. Quando ocorre uma gravação em um cluster, ela é replicada para o outro cluster. A gravação só é reconhecida para o host quando a gravação for concluída em ambos os sites. Além desse comportamento de divisão de protocolo, os dois clusters são, de outra forma, clusters ONTAP normais.
- A função principal do MetroCluster é a replicação em grande escala. É possível replicar um array inteiro com RPO igual a 0 e rto quase zero. Isso simplifica o processo de failover porque há apenas uma "coisa" a fazer failover e é dimensionado extremamente bem em termos de capacidade e IOPS.
- Um dos principais casos de uso para SM-as é a replicação granular. Às vezes, você não quer replicar todos os dados como uma única unidade ou precisa ser capaz de falhar seletivamente em determinados workloads.
- Outro importante caso de uso para SM-as é para operações ativas-ativas, em que você deseja que cópias totalmente utilizáveis de dados estejam disponíveis em dois clusters diferentes localizados em dois locais diferentes com características de desempenho idênticas e, se desejado, não é necessário estender a SAN entre locais. Você pode ter suas aplicações já em execução em ambos os locais, o que reduz o rto geral durante operações de failover.

MetroCluster

Recuperação de desastres com o MetroCluster

O MetroCluster é um recurso ONTAP que pode proteger seus bancos de dados Oracle com espelhamento síncrono de 0 RPO entre locais e escala para oferecer suporte a centenas de bancos de dados em um único sistema MetroCluster.

Também é simples de usar. O uso do MetroCluster não necessariamente adiciona ou altera quaisquer melhores pistas para operar aplicativos e bancos de dados empresariais.

As práticas recomendadas usuais ainda se aplicam. E, se suas necessidades exigirem apenas proteção de dados RPO de 0, essa necessidade será atendida com o MetroCluster. No entanto, a maioria dos clientes usa o MetroCluster não apenas para proteção de dados RPO igual a 0, mas também para aprimorar o rto durante cenários de desastre, bem como para fornecer failover transparente como parte das atividades de manutenção do local.

Arquitetura física

Entender como os bancos de dados Oracle operam em um ambiente MetroCluster requer alguma explicação do design físico de um sistema MetroCluster.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-4592: Oracle no MetroCluster*.

O MetroCluster pode ser usado em 3 configurações diferentes

- Pares HA com conectividade IP
- Pares DE HA com conectividade FC
- Controladora única com conectividade FC



O termo 'conectividade' refere-se à conexão de cluster usada para replicação entre sites. Não se refere aos protocolos de host. Todos os protocolos do lado do host são suportados como de costume em uma configuração MetroCluster, independentemente do tipo de conexão usada para comunicação entre clusters.

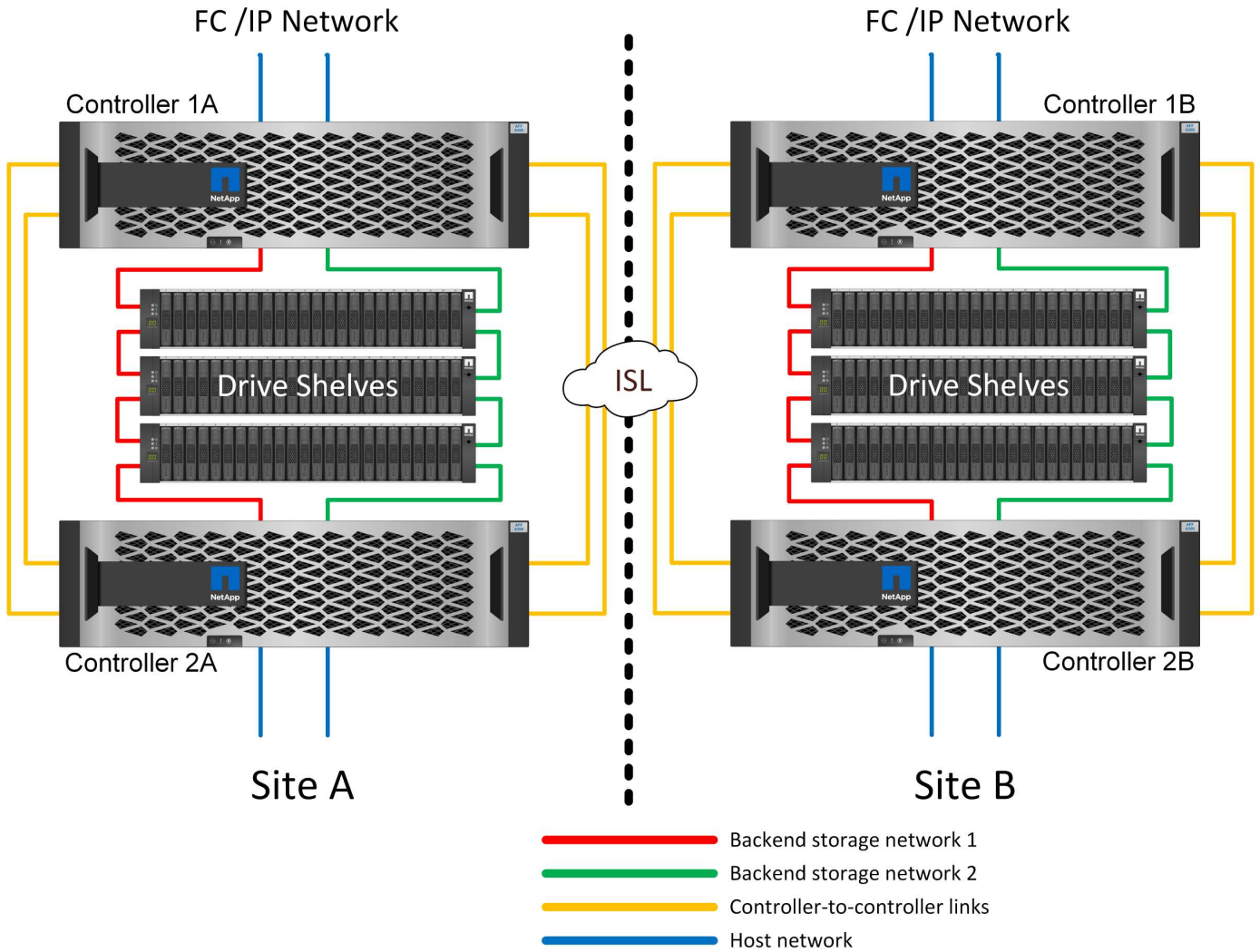
IP MetroCluster

A configuração MetroCluster IP de par de HA usa dois ou quatro nós por local. Essa opção de configuração aumenta a complexidade e os custos em relação à opção de dois nós, mas oferece um benefício importante: A redundância intrasite. Uma simples falha do controlador não requer acesso aos dados na WAN. O acesso aos dados permanece local por meio do controlador local alternativo.

A maioria dos clientes está escolhendo a conectividade IP porque os requisitos de infraestrutura são mais simples. No passado, a conectividade entre locais de alta velocidade era geralmente mais fácil de provisionar usando switches FC e fibra escura, mas hoje em dia os circuitos IP de baixa latência e alta velocidade estão mais prontamente disponíveis.

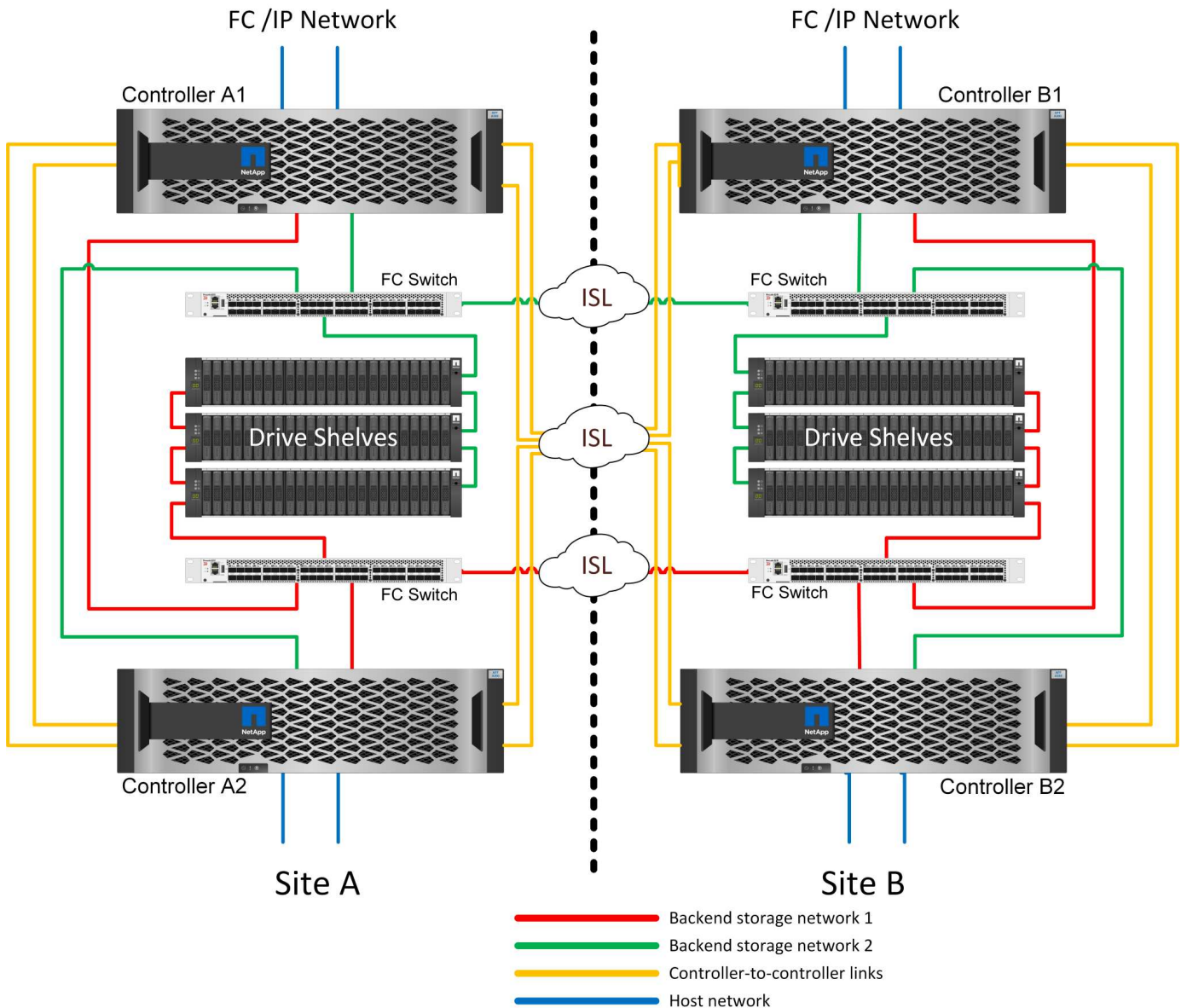
A arquitetura também é mais simples porque as únicas conexões entre locais são para os controladores. No Metroclusters FC SAN conectados, um controlador grava diretamente nas unidades no local oposto e, portanto, requer conexões, switches e bridges SAN adicionais. Em contraste, um controlador em uma configuração IP grava nas unidades opostas através do controlador.

Para obter informações adicionais, consulte a documentação oficial do ONTAP e "[Arquitetura e design da solução IP da MetroCluster](#)".



MetroCluster conectados a FC de par HA SAN

A configuração de MetroCluster FC de par de HA usa dois ou quatro nós por local. Essa opção de configuração aumenta a complexidade e os custos em relação à opção de dois nós, mas oferece um benefício importante: A redundância intrasite. Uma simples falha do controlador não requer acesso aos dados na WAN. O acesso aos dados permanece local por meio do controlador local alternativo.

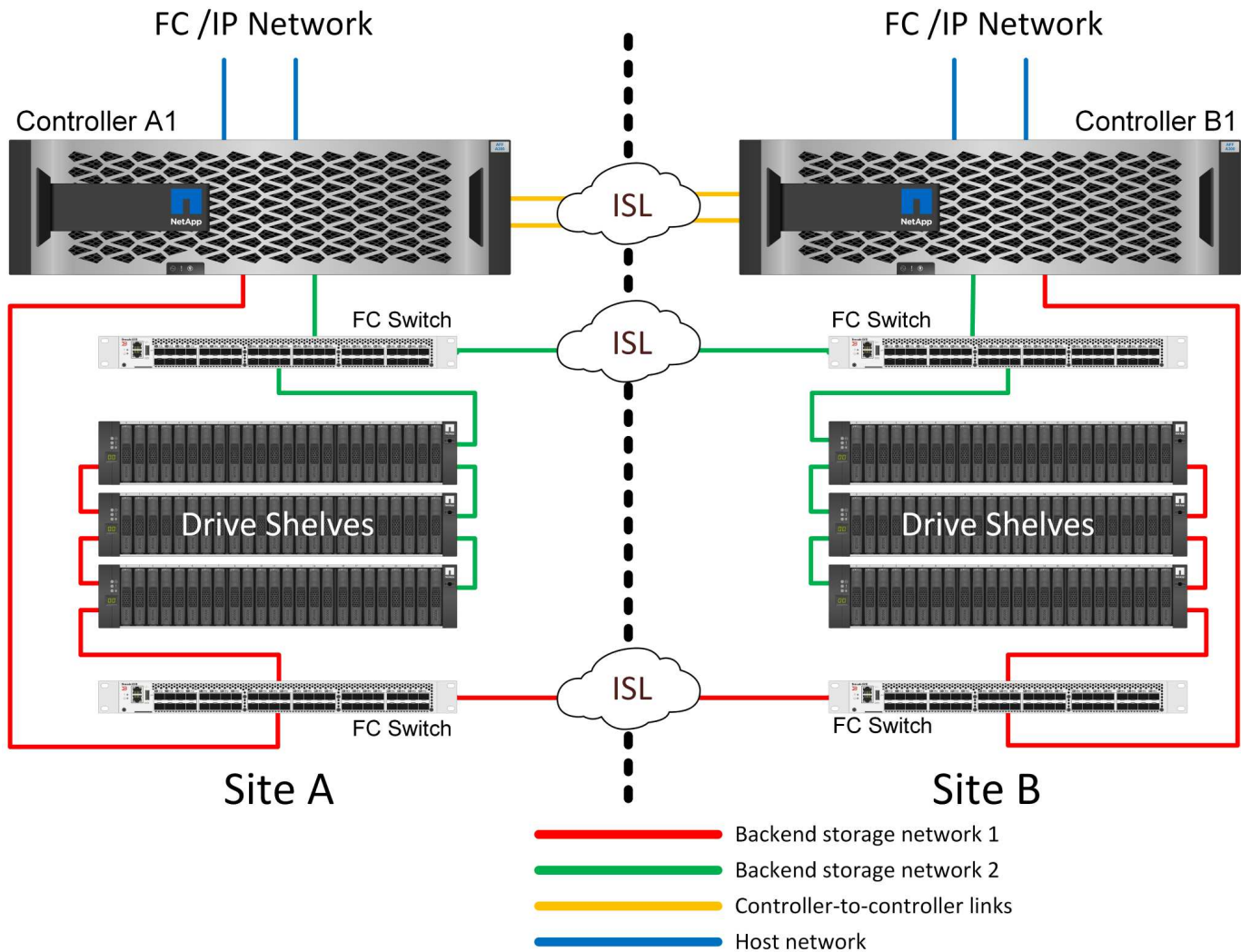


Algumas infraestruturas multisite não foram desenvolvidas para operações ativas-ativas, mas são usadas mais como local principal e local de recuperação de desastres. Nesta situação, uma opção MetroCluster de par de HA é geralmente preferível pelas seguintes razões:

- Embora um cluster de dois nós MetroCluster seja um sistema de HA, a falha inesperada de uma controladora ou a manutenção planejada exige que os serviços de dados fiquem online no local oposto. Se a conectividade de rede entre locais não puder suportar a largura de banda necessária, o desempenho é afetado. A única opção seria também falhar sobre os vários sistemas operacionais host e serviços associados ao site alternativo. O cluster de MetroCluster de par de HA elimina esse problema porque a perda de uma controladora resulta em failover simples no mesmo local.
- Algumas topologias de rede não são projetadas para acesso entre sites, mas usam sub-redes diferentes ou SANs FC isoladas. Nesses casos, o cluster MetroCluster de dois nós não funciona mais como um sistema de HA porque o controlador alternativo não pode fornecer dados aos servidores no local oposto. A opção MetroCluster de par de HA é necessária para fornecer redundância completa.
- Se uma infraestrutura de dois locais for vista como uma única infraestrutura altamente disponível, a configuração de dois nós do MetroCluster será adequada. No entanto, se o sistema precisar funcionar por um longo período de tempo após a falha do local, é preferível usar um par de HA porque ele continua fornecendo HA em um único local.

MetroCluster com conexão SAN FC de dois nós

A configuração do MetroCluster de dois nós usa apenas um nó por local. Esse design é mais simples do que a opção de par de HA, pois há menos componentes para configurar e manter. Ele também reduziu as demandas de infraestrutura em termos de cabeamento e switch FC. Finalmente, reduz custos.



O impacto óbvio desse projeto é que a falha do controlador em um único local significa que os dados estão disponíveis no local oposto. Esta restrição não é necessariamente um problema. Muitas empresas têm operações de data center multisite com redes estendidas, de alta velocidade e baixa latência, que funcionam essencialmente como uma única infraestrutura. Nesses casos, a versão de dois nós do MetroCluster é a configuração preferida. Sistemas de dois nós são usados atualmente em escala de petabyte por vários provedores de serviços.

Recursos de resiliência do MetroCluster

Não há pontos únicos de falha em uma solução MetroCluster:

- Cada controladora tem dois caminhos independentes para os compartimentos de unidades no local.
- Cada controladora tem dois caminhos independentes para o shelves de unidades no local remoto.
- Cada controlador tem dois caminhos independentes para os controladores no local oposto.
- Na configuração de par de HA, cada controladora tem dois caminhos para seu parceiro local.

Em resumo, qualquer componente na configuração pode ser removido sem comprometer a capacidade do MetroCluster de fornecer dados. A única diferença em termos de resiliência entre as duas opções é que a versão do par de HA ainda é um sistema de storage de HA geral após uma falha do local.

Arquitetura lógica

Entender como os bancos de dados Oracle operam em um ambiente MetroCluster o alsono requer alguma explicação da funcionalidade lógica de um sistema MetroCluster.

Proteção contra falha do local: NVRAM e MetroCluster

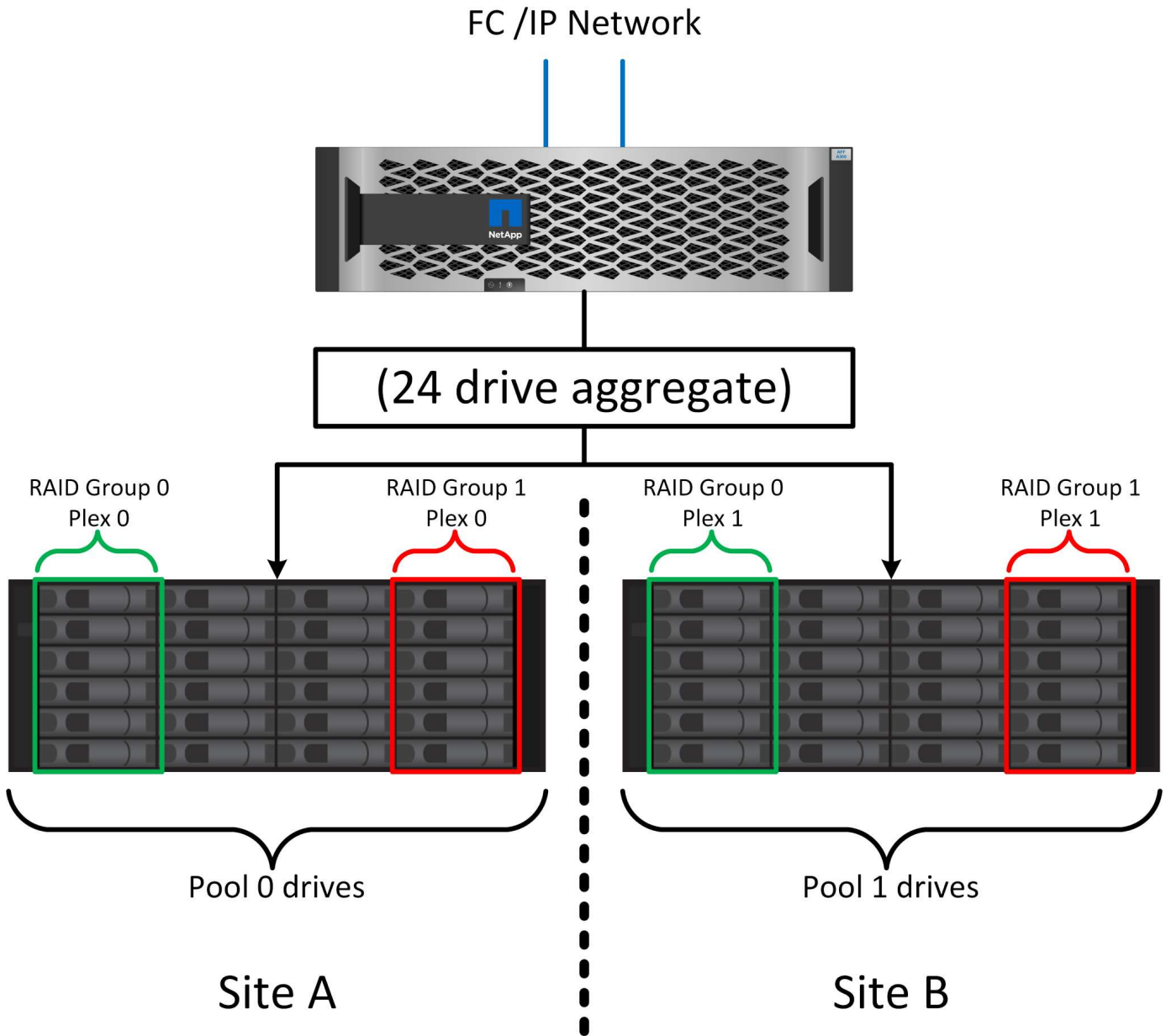
A MetroCluster estende a proteção de dados da NVRAM das seguintes maneiras:

- Em uma configuração de dois nós, os dados do NVRAM são replicados usando os ISLs (Inter-Switch Links) para o parceiro remoto.
- Em uma configuração de par de HA, os dados do NVRAM são replicados para o parceiro local e para um parceiro remoto.
- Uma gravação não é reconhecida até que seja replicada para todos os parceiros. Essa arquitetura protege a e/S em trânsito contra falhas do local replicando dados do NVRAM para um parceiro remoto. Este processo não está envolvido com a replicação de dados no nível da unidade. A controladora que detém os agregados é responsável pela replicação de dados por gravação em ambos os plexos no agregado, mas ainda deve haver proteção contra perda de e/S em trânsito em caso de perda do local. Os dados replicados do NVRAM só serão usados se um controlador do parceiro precisar assumir o controle de uma controladora com falha.

Proteção contra falhas no local e no compartimento: SyncMirror e plexos

O SyncMirror é uma tecnologia de espelhamento que aprimora, mas não substitui, o RAID DP ou o RAID-teC. Ele espelha o conteúdo de dois grupos RAID independentes. A configuração lógica é a seguinte:

1. As unidades são configuradas em dois pools com base no local. Um pool é composto por todas as unidades no local A, e o segundo pool é composto por todas as unidades no local B..
2. Um pool comum de armazenamento, conhecido como agregado, é criado com base em conjuntos espelhados de grupos RAID. Um número igual de unidades é extraído de cada local. Por exemplo, um agregado SyncMirror de 20 unidades seria composto por 10 unidades do local A e 10 unidades do local B..
3. Cada conjunto de unidades em um determinado local é configurado automaticamente como um ou mais grupos RAID DP ou RAID-teC totalmente redundantes, independentemente do uso do espelhamento. Esse uso de RAID por baixo do espelhamento fornece proteção de dados mesmo após a perda de um site.



A figura acima ilustra um exemplo de configuração do SyncMirror. Um agregado de 24 unidades foi criado na controladora com 12 unidades de um compartimento alocado no local A e 12 unidades de um compartimento alocado no local B. as unidades foram agrupadas em dois grupos RAID espelhados. RAID grupo 0 inclui um Plex de 6 unidades no local Um espelhado para um Plex de 6 unidades no local B. da mesma forma, RAID grupo 1 inclui um Plex de 6 unidades no local Um espelhado para um Plex de 6 unidades no local B.

O SyncMirror normalmente é usado para fornecer espelhamento remoto com sistemas MetroCluster, com uma cópia dos dados em cada local. Ocasionalmente, ele tem sido usado para fornecer um nível extra de redundância em um único sistema. Em particular, ele fornece redundância em nível de prateleira. Um compartimento de unidades já contém fontes de alimentação duplas e controladores e é, no geral, pouco mais do que chapas metálicas, mas em alguns casos, a proteção extra pode ser garantida. Por exemplo, um cliente da NetApp implantou o SyncMirror para uma plataforma móvel de análise em tempo real usada durante testes automotivos. O sistema foi separado em dois racks físicos fornecidos com alimentação de energia independente e sistemas UPS independentes.

Falha de redundância: NVFAIL

Como discutido anteriormente, uma gravação não é reconhecida até que ela tenha sido registrada no NVRAM local e no NVRAM em pelo menos um outro controlador. Essa abordagem garante que uma falha de hardware ou falha de energia não resulte na perda de e/S em trânsito. Se o NVRAM local falhar ou a conectividade com outros nós falhar, os dados não serão mais espelhados.

Se o NVRAM local relatar um erro, o nó será encerrado. Esse desligamento resulta em failover para uma controladora de parceiro quando os pares de HA são usados. Com o MetroCluster, o comportamento depende da configuração geral escolhida, mas pode resultar em failover automático para a nota remota. Em qualquer caso, nenhum dado é perdido porque o controlador que está tendo a falha não reconheceu a operação de gravação.

Uma falha de conectividade local a local que bloqueia a replicação do NVRAM para nós remotos é uma situação mais complicada. As gravações não são mais replicadas nos nós remotos, criando uma possibilidade de perda de dados se ocorrer um erro catastrófico em um controlador. Mais importante ainda, tentar fazer failover para um nó diferente durante essas condições resulta em perda de dados.

O fator de controle é se o NVRAM está sincronizado. Se o NVRAM estiver sincronizado, o failover de nó para nó será seguro para prosseguir sem risco de perda de dados. Em uma configuração do MetroCluster, se o NVRAM e os plexos agregados subjacentes estiverem sincronizados, é seguro prosseguir com o switchover sem risco de perda de dados.

O ONTAP não permite um failover ou switchover quando os dados estão fora de sincronia, a menos que o failover ou switchover seja forçado. Forçar uma alteração de condições desta forma reconhece que os dados podem ser deixados para trás no controlador original e que a perda de dados é aceitável.

Bancos de dados e outros aplicativos ficam especialmente vulneráveis à corrupção se um failover ou switchover for forçado, porque eles mantêm caches internos maiores de dados no disco. Se ocorrer um failover forçado ou switchover, as alterações anteriormente confirmadas serão efetivamente descartadas. O conteúdo da matriz de armazenamento salta efetivamente para trás no tempo, e o estado do cache não reflete mais o estado dos dados no disco.

Para evitar essa situação, o ONTAP permite que os volumes sejam configurados para proteção especial contra falha do NVRAM. Quando acionado, esse mecanismo de proteção resulta em um volume entrando em um estado chamado NVFAIL. Esse estado resulta em erros de e/S que causam uma falha no aplicativo. Esta falha faz com que os aplicativos sejam desligados para que eles não usem dados obsoletos. Os dados não devem ser perdidos porque quaisquer dados de transação confirmados devem estar presentes nos logs. As próximas etapas usuais são para que um administrador desligue totalmente os hosts antes de colocar manualmente os LUNs e volumes novamente on-line. Embora essas etapas possam envolver algum trabalho, essa abordagem é a maneira mais segura de garantir a integridade dos dados. Nem todos os dados exigem essa proteção, e é por isso que o comportamento do NVFAIL pode ser configurado volume a volume.

Pares DE HA e MetroCluster

O MetroCluster está disponível em duas configurações: Dois nós e par de HA. A configuração de dois nós se comporta da mesma forma que um par de HA em relação ao NVRAM. Em caso de falha repentina, o nó do parceiro pode repetir dados do NVRAM para tornar as unidades consistentes e garantir que nenhuma gravação reconhecida tenha sido perdida.

A configuração de par de HA replica NVRAM também para o nó do parceiro local. Uma simples falha do controlador resulta em uma repetição do NVRAM no nó do parceiro, como é o caso de um par de HA autônomo sem MetroCluster. Em caso de perda súbita total do local, o local remoto também tem o NVRAM necessário para tornar as unidades consistentes e começar a fornecer dados.

Um aspecto importante do MetroCluster é que os nós remotos não têm acesso aos dados do parceiro em condições operacionais normais. Cada site funciona essencialmente como um sistema independente que pode assumir a personalidade do site oposto. Esse processo é conhecido como switchover e inclui um switchover planejado no qual as operações do local são migradas sem interrupções para o local oposto. Ele também inclui situações não planejadas em que um local é perdido e um switchover manual ou automático é necessário como parte da recuperação de desastres.

Comutação e comutação

Os termos switchover e switchback referem-se ao processo de transição de volumes entre controladores remotos em uma configuração do MetroCluster. Este processo aplica-se apenas aos nós remotos. Quando o MetroCluster é usado em uma configuração de quatro volumes, o failover de nó local é o mesmo processo de aquisição e giveback descrito anteriormente.

Comutação planejada e switchback

Um switchover planejado ou switchback é semelhante a um takeover ou giveback entre nós. O processo tem várias etapas e pode parecer exigir vários minutos, mas o que está realmente acontecendo é uma transição graciosa multifásica de recursos de armazenamento e rede. O momento em que as transferências de controle ocorrem muito mais rapidamente do que o tempo necessário para que o comando completo seja executado.

A principal diferença entre o takeover/giveback e o switchover/switchback está com o efeito na conectividade FC SAN. Com a takeover local/giveback, um host sofre a perda de todos os caminhos FC para o nó local e conta com o MPIO nativo para mudar para caminhos alternativos disponíveis. As portas não são realocadas. Com o switchover e o switchback, as portas de destino FC virtual nos controladores fazem a transição para o outro local. Eles efetivamente deixam de existir na SAN por um momento e, em seguida, reaparecem em um controlador alternativo.

Tempos limite do SyncMirror

O SyncMirror é uma tecnologia de espelhamento ONTAP que fornece proteção contra falhas nas shelves. Quando as gavetas são separadas à distância, o resultado é a proteção de dados remota.

O SyncMirror não fornece espelhamento síncrono universal. O resultado é uma melhor disponibilidade. Alguns sistemas de storage usam espelhamento constante de tudo ou nada, às vezes chamado de modo domino. Essa forma de espelhamento é limitada no aplicativo porque toda atividade de gravação deve cessar se a conexão com o local remoto for perdida. Caso contrário, uma escrita existiria em um site, mas não no outro. Normalmente, esses ambientes são configurados para colocar LUNs off-line se a conectividade site-a-site for perdida por mais de um curto período (como 30 segundos).

Este comportamento é desejável para um pequeno subconjunto de ambientes. No entanto, a maioria dos aplicativos exige uma solução que ofereça replicação síncrona garantida em condições operacionais normais, mas com a capacidade de suspender a replicação. Uma perda completa de conectividade local a local é frequentemente considerada uma situação de quase desastre. Normalmente, esses ambientes são mantidos on-line e fornecem dados até que a conectividade seja reparada ou uma decisão formal seja tomada para encerrar o ambiente para proteger os dados. Um requisito para o desligamento automático do aplicativo puramente por causa de falha de replicação remota é incomum.

O SyncMirror dá suporte aos requisitos de espelhamento síncrono com a flexibilidade de um tempo limite. Se a conectividade com o telecomando e/ou Plex for perdida, um temporizador de 30 segundos começa a contagem decrescente. Quando o contador atinge 0, o processamento de e/S de escrita é retomado utilizando os dados locais. A cópia remota dos dados é utilizável, mas fica congelada no tempo até que a conectividade seja restaurada. A ressincronização utiliza snapshots em nível agregado para retornar o sistema ao modo síncrono o mais rápido possível.

Notavelmente, em muitos casos, esse tipo de replicação universal do modo dominó tudo ou nada é melhor implementado na camada de aplicativo. Por exemplo, o Oracle DataGuard inclui o modo de proteção máximo, o que garante replicação de longa instância em todas as circunstâncias. Se o link de replicação falhar por um período que excede um tempo limite configurável, os bancos de dados serão desligados.

Switchover automático sem supervisão com MetroCluster conectado à malha

O switchover automático sem supervisão (AUSO) é um recurso de MetroCluster anexado a malha que fornece uma forma de HA entre os locais. Como discutido anteriormente, o MetroCluster está disponível em dois tipos: Um único controlador em cada local ou um par de HA em cada local. A principal vantagem da opção HA é que o desligamento planejado ou não planejado do controlador ainda permite que todas as I/O sejam locais. A vantagem da opção de nó único é reduzir os custos, a complexidade e a infraestrutura.

O principal valor do AUSO é melhorar os recursos de HA dos sistemas MetroCluster conectados a malha. Cada local monitora a integridade do local oposto e, se nenhum nó permanecer para fornecer dados, o AUSO resulta em switchover rápido. Essa abordagem é especialmente útil nas configurações do MetroCluster com apenas um nó único por local, pois aproxima a configuração de um par de HA em termos de disponibilidade.

A AUSO não pode oferecer monitoramento abrangente no nível de um par de HA. Um par de HA pode fornecer disponibilidade extremamente alta porque inclui dois cabos físicos redundantes para comunicação direta de nó a nó. Além disso, ambos os nós de um par de HA têm acesso ao mesmo conjunto de discos em loops redundantes, entregando outra rota para um nó monitorar a integridade de outro.

Os clusters do MetroCluster existem em locais para os quais a comunicação nó a nó e o acesso ao disco dependem da conectividade de rede local a local. A capacidade de monitorar o batimento cardíaco do restante do cluster é limitada. AUSO tem que discriminar entre uma situação em que o outro site está realmente inativo, em vez de indisponível devido a um problema de rede.

Como resultado, uma controladora em um par de HA pode solicitar um takeover se detectar uma falha na controladora que ocorreu por um motivo específico, como pânico do sistema. Ele também pode solicitar uma aquisição se houver uma perda completa de conectividade, às vezes conhecida como batimento cardíaco perdido.

Um sistema MetroCluster só pode efetuar uma mudança automática em segurança quando é detectada uma avaria específica no local original. Além disso, a controladora que assume a propriedade do sistema de storage deve ser capaz de garantir que os dados do disco e do NVRAM estejam sincronizados. O controlador não pode garantir a segurança de uma mudança apenas porque perdeu o Contato com o local de origem, que ainda poderia estar operacional. Para obter opções adicionais para automatizar um switchover, consulte as informações sobre a solução MetroCluster tiebreaker (MCTB) na próxima seção.

Desempate MetroCluster com MetroCluster conectado à malha

<https://library.netapp.com/ecmdocs/ECMP12007400/html/GUID-3662A7CE-3AF2-4562-A11C-5C37DE0E3A87.html>["Desempate de NetApp MetroCluster"]O software pode ser executado em um terceiro local para monitorar a integridade do ambiente MetroCluster, enviar notificações e, opcionalmente, forçar um switchover em uma situação de desastre. Uma descrição completa do desempate pode ser encontrada no <http://mysupport.netapp.com>["Site de suporte da NetApp"], mas o principal objetivo do desempate do MetroCluster é detetar a perda do local. Ele também deve discriminar entre a perda do local e a perda de conectividade. Por exemplo, o switchover não deve ocorrer porque o tiebreaker não conseguiu chegar ao local principal, e é por isso que o tiebreaker também monitora a capacidade do local remoto de entrar em Contato com o local principal.

O switchover automático com AUSO também é compatível com o MCTB. O AUSO reage muito rapidamente porque foi concebido para detetar eventos de falha específicos e, em seguida, invocar o switchover apenas quando os plexos NVRAM e SyncMirror estão em sincronia.

Em contraste, o desempate está localizado remotamente e, portanto, deve esperar que um temporizador decorra antes de declarar um local morto. O tiebreaker eventualmente deteta o tipo de falha de controladora coberta pelo AUSO, mas, em geral, a AUSO já iniciou o switchover e possivelmente concluiu o switchover antes que o tiebreaker atue. O segundo comando de comutação resultante vindo do tiebreaker seria rejeitado.

*Atenção: *O software MCTB não verifica se o NVRAM estava e/ou os plexos estão sincronizados ao forçar um switchover. O switchover automático, se configurado, deve ser desativado durante atividades de manutenção que resultem na perda de sincronização para NVRAM ou SyncMirror plexes.

Além disso, o MCTB pode não resolver um desastre contínuo que leva à seguinte sequência de eventos:

1. A conectividade entre locais é interrompida durante mais de 30 segundos.
2. O tempo de replicação do SyncMirror expirou e as operações continuam no local principal, deixando a réplica remota obsoleta.
3. O site principal é perdido. O resultado é a presença de alterações não replicadas no site principal. Uma mudança pode então ser indesejável por uma série de razões, incluindo o seguinte:
 - Dados críticos podem estar presentes no site principal e esses dados podem eventualmente ser recuperáveis. Um switchover que permitiu que o aplicativo continuasse operando descartaria efetivamente esses dados críticos.
 - Um aplicativo no site que estava usando recursos de armazenamento no site principal no momento da perda do site pode ter dados em cache. Um switchover introduziria uma versão obsoleta dos dados que não corresponde ao cache.
 - Um sistema operacional no site sobrevivente que estava usando recursos de armazenamento no site principal no momento da perda do site pode ter dados em cache. Um switchover introduziria uma versão obsoleta dos dados que não corresponde ao cache. A opção mais segura é configurar o tiebreaker para enviar um alerta se ele detetar falha no local e, em seguida, fazer com que uma pessoa tome uma decisão sobre se deve forçar um switchover. Os aplicativos e/ou sistemas operacionais podem precisar primeiro ser desligados para limpar os dados armazenados em cache. Além disso, as configurações NVFAIL podem ser usadas para adicionar mais proteção e ajudar a simplificar o processo de failover.

Mediador ONTAP com MetroCluster IP

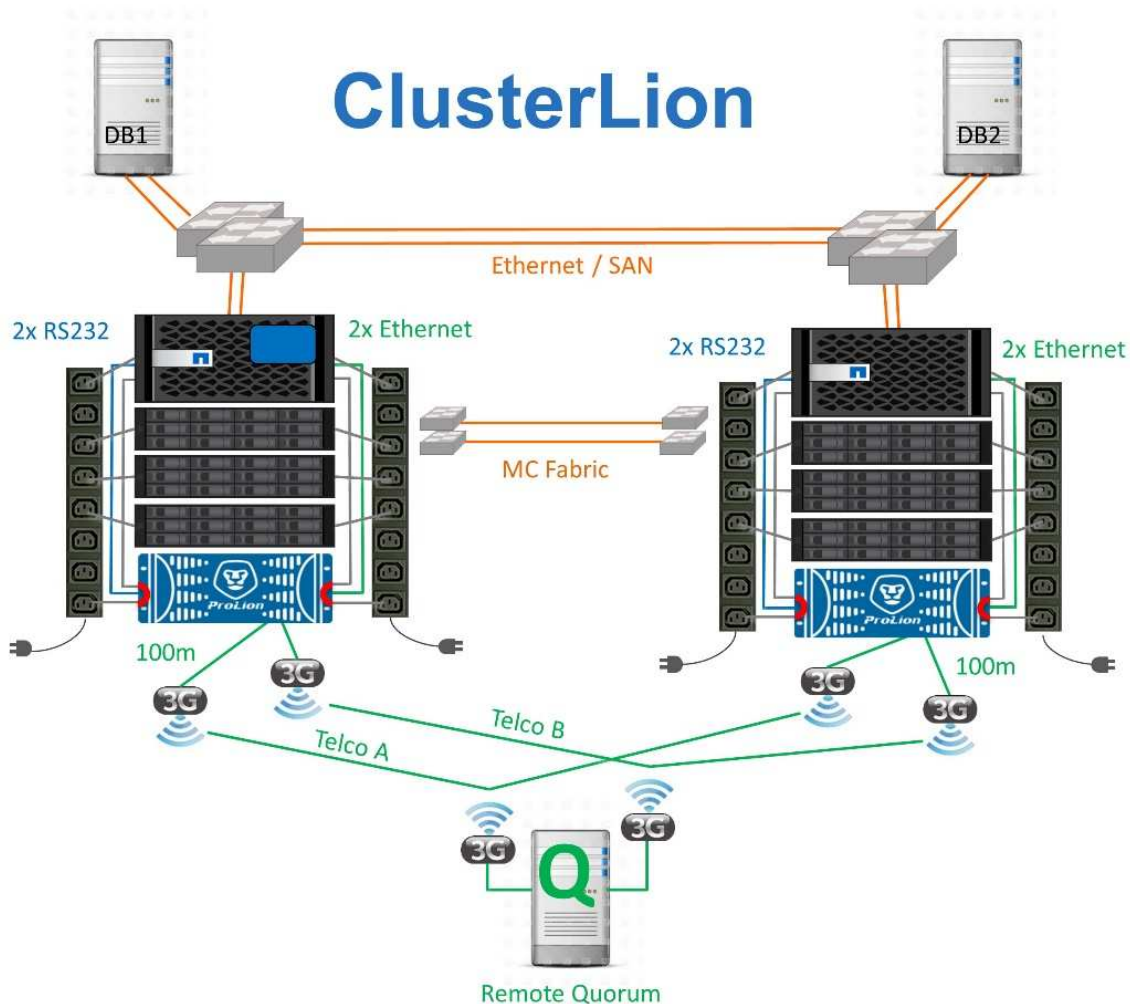
O Mediador ONTAP é usado com MetroCluster IP e outras soluções ONTAP. Ele funciona como um serviço de desempate tradicional, assim como o software de desempate do MetroCluster discutido acima, mas também inclui um recurso crítico: Executar switchover automatizado sem supervisão.

Um MetroCluster conectado à malha tem acesso direto aos dispositivos de storage no local oposto. Isso permite que um controlador MetroCluster monitore a integridade dos outros controladores lendo dados de batimentos cardíacos das unidades. Isso permite que um controlador reconheça a falha de outro controlador e execute um switchover.

Em contraste, a arquitetura IP do MetroCluster roteia todas as I/O exclusivamente através da conexão controlador-controlador; não há acesso direto a dispositivos de armazenamento no local remoto. Isso limita a capacidade de um controlador detectar falhas e executar um switchover. O Mediador ONTAP é, portanto, necessário como um dispositivo de desempate para detectar a perda do local e executar automaticamente um switchover.

Terceiro site virtual com ClusterLion

O ClusterLion é um dispositivo avançado de monitoramento MetroCluster que funciona como um terceiro site virtual. Essa abordagem permite que o MetroCluster seja implantado com segurança em uma configuração de dois locais com recurso de switchover totalmente automatizado. Além disso, o ClusterLion pode executar um monitor de nível de rede adicional e executar operações pós-switchover. A documentação completa está disponível no ProLion.



- Os dispositivos ClusterLion monitoram a integridade dos controladores com cabos Ethernet e seriais conectados diretamente.
- Os dois aparelhos estão conectados entre si com conexões sem fio redundantes de 3G GHz.
- A alimentação para o controlador ONTAP é direcionada através de relés internos. No caso de uma falha no local, o ClusterLion, que contém um sistema interno de UPS, corta as conexões de energia antes de chamar uma mudança. Este processo garante que nenhuma condição de divisão cerebral ocorra.
- O ClusterLion executa um switchover dentro do tempo limite de 30 segundos do SyncMirror ou não.
- O ClusterLion não executa uma mudança a menos que os estados dos plexes NVRAM e SyncMirror estejam sincronizados.
- Como o ClusterLion só executa um switchover se o MetroCluster estiver totalmente sincronizado, o NVFAIL não é necessário. Essa configuração permite que ambientes que abrangem o local, como um Oracle RAC estendido, permaneçam on-line, mesmo durante um switchover não planejado.
- O suporte inclui MetroCluster conectado à malha e MetroCluster IP

SyncMirror

A base da proteção de dados Oracle com um sistema MetroCluster é o SyncMirror, uma tecnologia de espelhamento síncrono com escalabilidade horizontal e alta performance.

Proteção de dados com o SyncMirror

No nível mais simples, a replicação síncrona significa que qualquer alteração deve ser feita em ambos os lados do storage espelhado antes que seja reconhecida. Por exemplo, se um banco de dados estiver escrevendo um log ou se um convidado VMware estiver sendo corrigido, uma gravação nunca deve ser perdida. Como um nível de protocolo, o sistema de storage não deve reconhecer a gravação até que ela tenha sido comprometida com a Mídia não volátil em ambos os locais. Só então é seguro prosseguir sem o risco de perda de dados.

O uso de uma tecnologia de replicação síncrona é a primeira etapa no projeto e gerenciamento de uma solução de replicação síncrona. A consideração mais importante é entender o que poderia acontecer durante vários cenários de falha planejados e não planejados. Nem todas as soluções de replicação síncrona oferecem os mesmos recursos. Se você precisa de uma solução que forneça um objetivo de ponto de restauração (RPO) zero, o que significa perda de dados zero, é necessário considerar todos os cenários de falha. Em particular, qual é o resultado esperado quando a replicação é impossível devido à perda de conectividade entre sites?

Disponibilidade de dados do SyncMirror

A replicação do MetroCluster é baseada na tecnologia NetApp SyncMirror, projetada para entrar e sair do modo síncrono com eficiência. Essa funcionalidade atende aos requisitos dos clientes que exigem replicação síncrona, mas que também precisam de alta disponibilidade para seus serviços de dados. Por exemplo, se a conectividade a um local remoto for cortada, geralmente é preferível que o sistema de armazenamento continue operando em um estado não replicado.

Muitas soluções de replicação síncrona só são capazes de operar no modo síncrono. Esse tipo de replicação tudo ou nada é às vezes chamado de modo domino. Esses sistemas de storage param de fornecer dados em vez de permitir que cópias locais e remotas dos dados fiquem não sincronizadas. Se a replicação for violada à força, a ressincronização pode ser extremamente demorada e pode deixar um cliente exposto à perda completa de dados durante o tempo em que o espelhamento é restabelecido.

O SyncMirror não só pode alternar facilmente do modo síncrono se o local remoto não estiver acessível, como também pode sincronizar rapidamente para um estado RPO de 0 quando a conectividade é restaurada. A cópia

obsoleta dos dados no local remoto também pode ser preservada em um estado utilizável durante a resincronização, o que garante que cópias locais e remotas dos dados existam em todos os momentos.

Quando o modo domino é necessário, o NetApp oferece SnapMirror Synchronous (SM-S). Opções de nível de aplicativo também existem, como o Oracle DataGuard ou o SQL Server Always On Availability Groups. O espelhamento de disco no nível DO SO pode ser uma opção. Consulte sua equipe de conta do NetApp ou do parceiro para obter informações e opções adicionais.

MetroCluster e NVFAIL

O NVFAIL é um recurso de integridade de dados geral do ONTAP projetado para maximizar a proteção da integridade de dados com bancos de dados.



Esta seção expande a explicação do NVFAIL básico do ONTAP para cobrir tópicos específicos do MetroCluster.

Com o MetroCluster, uma gravação não é reconhecida até que ela tenha sido registrada no NVRAM local e no NVRAM em pelo menos um outro controlador. Essa abordagem garante que uma falha de hardware ou falha de energia não resulte na perda de e/S em trânsito. Se o NVRAM local falhar ou a conectividade com outros nós falhar, os dados não serão mais espelhados.

Se o NVRAM local relatar um erro, o nó será encerrado. Esse desligamento resulta em failover para uma controladora de parceiro quando os pares de HA são usados. Com o MetroCluster, o comportamento depende da configuração geral escolhida, mas pode resultar em failover automático para a nota remota. Em qualquer caso, nenhum dado é perdido porque o controlador que está tendo a falha não reconheceu a operação de gravação.

Uma falha de conectividade local a local que bloqueia a replicação do NVRAM para nós remotos é uma situação mais complicada. As gravações não são mais replicadas nos nós remotos, criando uma possibilidade de perda de dados se ocorrer um erro catastrófico em um controlador. Mais importante ainda, tentar fazer failover para um nó diferente durante essas condições resulta em perda de dados.

O fator de controle é se o NVRAM está sincronizado. Se o NVRAM estiver sincronizado, o failover de nó para nó será seguro para prosseguir sem o risco de perda de dados. Em uma configuração do MetroCluster, se o NVRAM e os plexos agregados subjacentes estiverem sincronizados, é seguro prosseguir com o switchover sem o risco de perda de dados.

O ONTAP não permite um failover ou switchover quando os dados estão fora de sincronia, a menos que o failover ou switchover seja forçado. Forçar uma alteração de condições desta forma reconhece que os dados podem ser deixados para trás no controlador original e que a perda de dados é aceitável.

Os bancos de dados são especialmente vulneráveis à corrupção se um failover ou switchover for forçado porque os bancos de dados mantêm caches internos maiores de dados no disco. Se ocorrer um failover forçado ou switchover, as alterações anteriormente confirmadas serão efetivamente descartadas. O conteúdo da matriz de armazenamento salta efetivamente para trás no tempo, e o estado do cache do banco de dados não reflete mais o estado dos dados no disco.

Para proteger aplicativos contra essa situação, o ONTAP permite que volumes sejam configurados para proteção especial contra falha do NVRAM. Quando acionado, esse mecanismo de proteção resulta em um volume entrando em um estado chamado NVFAIL. Esse estado resulta em erros de e/S que causam o desligamento de um aplicativo para que eles não usem dados obsoletos. Os dados não devem ser perdidos porque quaisquer gravações reconhecidas ainda estão presentes no sistema de armazenamento e, com bancos de dados, quaisquer dados de transações confirmadas devem estar presentes nos logs.

As próximas etapas usuais são para que um administrador desligue totalmente os hosts antes de colocar manualmente os LUNs e volumes novamente on-line. Embora essas etapas possam envolver algum trabalho, essa abordagem é a maneira mais segura de garantir a integridade dos dados. Nem todos os dados exigem essa proteção, e é por isso que o comportamento do NVFAIL pode ser configurado volume a volume.

NVFAIL forçado manualmente

A opção mais segura para forçar um switchover com um cluster de aplicativos (incluindo VMware, Oracle RAC e outros) que é distribuído entre locais é especificando `-force-nvfail-all` na linha de comando. Essa opção está disponível como uma medida de emergência para garantir que todos os dados em cache sejam limpos. Se um host estiver usando recursos de armazenamento localizados originalmente no local afetado por desastre, ele receberá erros de e/S ou um (`ESTALE` erro de identificador de arquivo obsoleto). Os bancos de dados Oracle falham e os sistemas de arquivos ficam totalmente offline ou mudam para o modo somente leitura.

Após a conclusão do switchover, o `in-nvfailed-state` sinalizador precisa ser limpo e os LUNs precisam ser colocados on-line. Depois de concluir esta atividade, a base de dados pode ser reiniciada. Essas tarefas podem ser automatizadas para reduzir o rto.

dr-force-nvfail

Como medida geral de segurança, defina o `dr-force-nvfail` sinalizador em todos os volumes que possam ser acessados de um local remoto durante operações normais, o que significa que são atividades usadas antes do failover. O resultado desta definição é que os volumes remotos selecionados ficam indisponíveis quando entram `in-nvfailed-state` durante um switchover. Após a conclusão do switchover, o `in-nvfailed-state` sinalizador deve ser limpo e os LUNs devem ser colocados on-line. Depois de concluir estas atividades, as aplicações podem ser reiniciadas. Essas tarefas podem ser automatizadas para reduzir o rto.

O resultado é como usar a `-force-nvfail-all` bandeira para switchovers manuais. No entanto, o número de volumes afetados pode ser limitado apenas aos volumes que devem ser protegidos de aplicativos ou sistemas operacionais com caches obsoletos.

Há dois requisitos essenciais para um ambiente que não é usado `dr-force-nvfail` em volumes de aplicações:

- Um switchover forçado não deve ocorrer mais de 30 segundos após a perda do local principal.
- Um switchover não deve ocorrer durante as tarefas de manutenção ou quaisquer outras condições em que os plexos SyncMirror ou a replicação NVRAM estejam fora de sincronia. O primeiro requisito pode ser atendido usando o software tiebreaker configurado para executar um switchover em 30 segundos após uma falha no local. Esse requisito não significa que o switchover deve ser executado dentro de 30 segundos após a detecção de uma falha no local. Isso significa que não é mais seguro forçar uma mudança se tiverem decorrido 30 segundos desde que um local foi confirmado como operacional.

O segundo requisito pode ser parcialmente atendido desativando todos os recursos de switchover automatizado quando a configuração do MetroCluster estiver fora de sincronia. Uma opção melhor é ter uma solução de desempate que possa monitorar a integridade da replicação do NVRAM e dos plexes do SyncMirror. Se o cluster não estiver totalmente sincronizado, o desempate não deverá acionar um switchover.

O software MCTB da NetApp não consegue monitorizar o estado da sincronização, pelo que deve ser desativado quando o MetroCluster não está sincronizado por qualquer motivo. O ClusterLion inclui recursos de monitoramento NVRAM e monitoramento Plex e pode ser configurado para não acionar o switchover, a menos que o sistema MetroCluster seja confirmado como totalmente sincronizado.

Instância única Oracle

Como dito anteriormente, a presença de um sistema MetroCluster não necessariamente adiciona ou altera quaisquer práticas recomendadas para operar um banco de dados. A maioria dos bancos de dados atualmente em execução em sistemas MetroCluster cliente é uma instância única e segue as recomendações na documentação do Oracle On ONTAP.

Failover com um SO pré-configurado

O SyncMirror fornece uma cópia síncrona dos dados no local de recuperação de desastre, mas disponibilizar esses dados requer um sistema operacional e as aplicações associadas. A automação básica pode melhorar significativamente o tempo de failover do ambiente geral. Os produtos Clusterware, como o Veritas Cluster Server (VCS), são frequentemente usados para criar um cluster nos sites e, em muitos casos, o processo de failover pode ser conduzido com scripts simples.

Se os nós primários forem perdidos, o clusterware (ou scripts) é configurado para colocar os bancos de dados on-line no site alternativo. Uma opção é criar servidores de reserva pré-configurados para os recursos NFS ou SAN que compõem o banco de dados. Se o site principal falhar, a alternativa clusterware ou scripted executa uma sequência de ações semelhantes às seguintes:

1. Forçar um switchover do MetroCluster
2. Realizando a descoberta de FC LUNs (somente SAN)
3. Montagem de sistemas de arquivos e/ou montagem de grupos de discos ASM
4. Iniciando o banco de dados

O principal requisito dessa abordagem é um sistema operacional em execução no local remoto. Ele deve ser pré-configurado com binários Oracle, o que também significa que tarefas como patches Oracle devem ser executadas no site primário e em espera. Como alternativa, os binários Oracle podem ser espelhados para o local remoto e montados se um desastre for declarado.

O procedimento de ativação real é simples. Comandos como o reconhecimento LUN requerem apenas alguns comandos por porta FC. A montagem do sistema de arquivos não é mais do que um `mount` comando, e os bancos de dados e ASM podem ser iniciados e parados na CLI com um único comando. Se os volumes e os sistemas de arquivos não estiverem em uso no local de recuperação de desastres antes do switchover, não há requisito de definir `dr-force- nvfail` os volumes.

Failover com um sistema operacional virtualizado

O failover de ambientes de banco de dados pode ser estendido para incluir o próprio sistema operacional. Em teoria, esse failover pode ser feito com LUNs de inicialização, mas na maioria das vezes é feito com um sistema operacional virtualizado. O procedimento é semelhante aos seguintes passos:

1. Forçar um switchover do MetroCluster
2. Montagem dos armazenamentos de dados que hospedam as máquinas virtuais do servidor de banco de dados
3. Iniciar as máquinas virtuais
4. Iniciando bancos de dados manualmente ou configurando as máquinas virtuais para iniciar automaticamente os bancos de dados, por exemplo, um cluster ESX pode abranger sites. Em caso de desastre, as máquinas virtuais podem ser colocadas on-line no local de recuperação de desastres após o switchover. Desde que os armazenamentos de dados que hospedam os servidores de banco de dados

virtualizados não estejam em uso no momento do desastre, não há nenhum requisito para definir `dr-force-nvfail` os volumes associados.

Oracle Extended RAC

Muitos clientes otimizam seu rto alongando um cluster do Oracle RAC entre locais, gerando uma configuração totalmente ativo-ativo. O projeto geral se torna mais complicado porque deve incluir o gerenciamento de quórum do Oracle RAC. Além disso, os dados são acessados de ambos os sites, o que significa que uma mudança forçada pode levar ao uso de uma cópia desatualizada dos dados.

Embora uma cópia dos dados esteja presente em ambos os sites, apenas o controlador que atualmente possui um agregado pode fornecer dados. Portanto, com clusters RAC estendidos, os nós remotos devem executar e/S em uma conexão local a local. O resultado é uma latência de e/S adicionada, mas essa latência geralmente não é um problema. A rede de interconexão RAC também deve ser estendida entre locais, o que significa que uma rede de alta velocidade e baixa latência é necessária de qualquer maneira. Se a latência adicionada causar um problema, o cluster pode ser operado de forma ativo-passivo. As operações com uso intenso de e/S precisariam então ser direcionadas para os nós RAC que são locais para a controladora que possui os agregados. Os nós remotos executam então operações de e/S mais leves ou são usados puramente como servidores de espera quentes.

Se o RAC estendido ativo-ativo for necessário, a sincronização ativa do SnapMirror deve ser considerada no lugar do MetroCluster. A replicação SM-as permite que uma réplica específica dos dados seja preferida. Portanto, um cluster RAC estendido pode ser construído no qual todas as leituras ocorrem localmente. A I/O de leitura nunca cruza sites, o que proporciona a menor latência possível. Todas as atividades de gravação ainda devem transitar a conexão entre locais, mas esse tráfego é inevitável com qualquer solução de espelhamento síncrono.



Se os LUNs de inicialização, incluindo discos de inicialização virtualizados, forem usados com o Oracle RAC, o `misscount` parâmetro pode precisar ser alterado. Para obter mais informações sobre os parâmetros de tempo limite do RAC, "[Oracle RAC com ONTAP](#)" consulte .

Configuração de dois locais

Uma configuração RAC estendida de dois locais pode fornecer serviços de banco de dados ativo-ativo que podem sobreviver a muitos, mas não todos, cenários de desastre sem interrupções.

Ficheiros de votação RAC

A primeira consideração ao implantar o RAC estendido no MetroCluster deve ser o gerenciamento de quórum. O Oracle RAC tem dois mecanismos para gerenciar quórum: O batimento cardíaco do disco e o batimento cardíaco da rede. O heartbeat do disco monitora o acesso ao armazenamento usando os arquivos de votação. Com uma configuração RAC de local único, um único recurso de votação é suficiente, desde que o sistema de armazenamento subjacente ofereça recursos de HA.

Em versões anteriores do Oracle, os arquivos de votação foram colocados em dispositivos de armazenamento físico, mas nas versões atuais do Oracle os arquivos de votação são armazenados em grupos de discos ASM.



O Oracle RAC é compatível com NFS. Durante o processo de instalação da grade, um conjunto de processos ASM é criado para apresentar o local NFS usado para arquivos de grade como um grupo de discos ASM. O processo é quase transparente para o usuário final e não requer gerenciamento contínuo do ASM após a conclusão da instalação.

O primeiro requisito em uma configuração de dois locais é garantir que cada local possa sempre acessar mais da metade dos arquivos de votação de forma que garanta um processo de recuperação de desastres sem interrupções. Essa tarefa era simples antes que os arquivos de votação fossem armazenados em grupos de discos ASM, mas hoje os administradores precisam entender os princípios básicos da redundância ASM.

Os grupos de discos ASM têm três opções para redundância `external`, `normal` e `high`. Em outras palavras, sem espelhamento, espelhado e espelhado de 3 vias. Uma opção mais recente chamada `Flex` também está disponível, mas raramente usada. O nível de redundância e o posicionamento dos dispositivos redundantes controlam o que acontece em cenários de falha. Por exemplo:

- Colocar os arquivos de votação em um `diskgroup` recurso com `external` redundância garante despejo de um site se a conectividade entre sites for perdida.
- Colocar os arquivos de votação em um `diskgroup` com `normal` redundância com apenas um disco ASM por site garante despejo de nó em ambos os sites se a conectividade entre sites for perdida porque nenhum dos sites teria quórum de maioria.
- Colocar os arquivos de votação em um `diskgroup high` com redundância com dois discos em um local e um único disco no outro local permite operações ativas-ativas quando ambos os sites estão operacionais e mutuamente acessíveis. No entanto, se o site de disco único for isolado da rede, então esse site é despejado.

Batimento cardíaco da rede RAC

O batimento cardíaco da rede do Oracle RAC monitora a acessibilidade dos nós na interconexão de cluster. Para permanecer no cluster, um nó deve ser capaz de entrar em Contato com mais da metade dos outros nós. Em uma arquitetura de dois locais, esse requisito cria as seguintes opções para a contagem de nós RAC:

- O posicionamento de um número igual de nós por local resulta em despejo em um local caso a conectividade de rede seja perdida.
- O posicionamento de nós N em um local e nós N-1 no outro local garante que a perda de conectividade entre locais resulta no local com o maior número de nós restantes no quórum de rede e o local com menos nós despejando.

Antes do Oracle 12cR2, não era viável controlar qual lado experimentaria um despejo durante a perda do local. Quando cada local tem um número igual de nós, o despejo é controlado pelo nó principal, que em geral é o primeiro nó RAC a inicializar.

O Oracle 12cR2 introduz a capacidade de ponderação de nós. Essa capacidade dá a um administrador mais controle sobre como a Oracle resolve condições de `split-brain`. Como um exemplo simples, o comando a seguir define a preferência de um nó específico em um 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.
```

Após reiniciar o Oracle High-Availability Services, a configuração será a seguinte:

```
[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
```

O nó `host-a` agora é designado como o servidor crítico. Se os dois nós RAC estiverem isolados, `host-a` sobrevive e `host-b` é despejado.



Para obter detalhes completos, consulte o white paper da Oracle "Visão geral técnica do Oracle Clusterware 12c versão 2. "

Para versões do Oracle RAC anteriores ao 12cR2, o nó principal pode ser identificado verificando os logs do CRS da seguinte forma:

```
[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
```

Este log indica que o nó principal é 2 e o nó `host-a` tem uma ID de 1. Esse fato significa que `host-a` não é o nó principal. A identidade do nó principal pode ser confirmada com o comando `olsnodes -n`.

```
[root@host-a ~]# /grid/bin/olsnodes -n
host-a 1
host-b 2
```

O nó com uma ID de 2 é `host-b`, que é o nó principal. Em uma configuração com números iguais de nós em cada site, o site com `host-b` é o site que sobrevive se os dois conjuntos perderem a conectividade de rede por qualquer motivo.

É possível que a entrada de log que identifica o nó mestre possa ficar fora do sistema. Nesta situação, os carimbos de data/hora dos backups do Oracle Cluster Registry (OCR) podem ser usados.

```
[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
```

Este exemplo mostra que o nó principal é `host-b`. Ele também indica uma mudança no nó mestre de `host-a` para `host-b` algum lugar entre 2:05 e 21:39 em 4 de maio. Este método de identificação do nó principal só é seguro se os logs do CRS também tiverem sido verificados porque é possível que o nó principal tenha sido alterado desde o backup OCR anterior. Se essa alteração tiver ocorrido, ela deverá estar visível nos logs do OCR.

A maioria dos clientes escolhe um único grupo de discos de votação que atende todo o ambiente e um número igual de nós RAC em cada local. O grupo de discos deve ser colocado no site que contém o banco de dados. O resultado é que a perda de conectividade resulta em despejo no local remoto. O site remoto não teria mais quórum, nem teria acesso aos arquivos do banco de dados, mas o site local continua sendo executado como de costume. Quando a conectividade é restaurada, a instância remota pode ser colocada online novamente.

Em caso de desastre, é necessário um switchover para colocar os arquivos do banco de dados e o grupo de discos de votação on-line no local sobrevivente. Se o desastre permitir que o AUSO acione o switchover, o NVFAIL não será acionado porque o cluster é conhecido por estar em sincronia e os recursos de storage ficam online normalmente. AUSO é uma operação muito rápida e deve ser concluída antes que o `disktimeout` período expire.

Como existem apenas dois locais, não é possível usar qualquer tipo de software de quebra de informações externo automatizado, o que significa que o switchover forçado deve ser uma operação manual.

Configurações de três locais

Um cluster RAC estendido é muito mais fácil de arquitetar com três locais. Os dois sites que hospedam cada metade do sistema MetroCluster também dão suporte aos workloads de banco de dados, enquanto o terceiro local serve como desempate para o banco de dados e para o sistema MetroCluster. A configuração do Oracle tiebreaker pode ser tão simples quanto colocar um membro do grupo de discos ASM usado para votar em um site 3rd e também pode incluir uma instância operacional no site 3rd para garantir que haja um número ímpar de nós no cluster RAC.



Consulte a documentação da Oracle sobre "grupo de falha de quórum" para obter informações importantes sobre o uso do NFS em uma configuração RAC estendida. Em resumo, as opções de montagem NFS podem precisar ser modificadas para incluir a opção de software para garantir que a perda de conectividade com os recursos de quórum de hospedagem de sites 3rd não pendure os servidores Oracle primários ou os processos Oracle RAC.

Sincronização ativa do SnapMirror

Visão geral

O SnapMirror ativo Sync permite que você crie ambientes de banco de dados Oracle de alta disponibilidade, onde LUNs estão disponíveis em dois clusters de storage diferentes.

Com a sincronização ativa do SnapMirror, não há cópia "primária" e "secundária" dos dados. Cada cluster pode fornecer leitura de e/S de sua cópia local dos dados, e cada cluster replicará uma gravação para seu parceiro. O resultado é um comportamento de IO simétrico.

Entre outras opções, isso permite que você execute o Oracle RAC como um cluster estendido com instâncias operacionais em ambos os sites. Como alternativa, você pode criar clusters de banco de dados ativo-passivo RPO igual a 0 em que bancos de dados de instância única podem ser movidos por locais durante uma falha no local. Esse processo pode ser automatizado por meio de produtos como pacemaker ou VMware HA. A base para todas essas opções é a replicação síncrona gerenciada pela sincronização ativa do SnapMirror.

Replicação síncrona

Em operação normal, o SnapMirror active Sync fornece réplica síncrona RPO igual a 0 em todos os momentos, com uma exceção. Se os dados não puderem ser replicados, o ONTAP cumprirá o requisito de replicar dados e retomar a distribuição de I/O em um local, enquanto os LUNs no outro local ficam offline.

Hardware de storage

Ao contrário de outras soluções de recuperação de desastres de storage, o SnapMirror active Sync oferece flexibilidade assimétrica de plataforma. O hardware em cada local não precisa ser idêntico. Esse recurso permite dimensionar corretamente o hardware usado para suportar a sincronização ativa do SnapMirror. O sistema de storage remoto pode ser idêntico ao local principal se precisar dar suporte a uma carga de trabalho de produção completa, mas se um desastre resultar em e/S reduzida, do que um sistema menor no local remoto pode ser mais econômico.

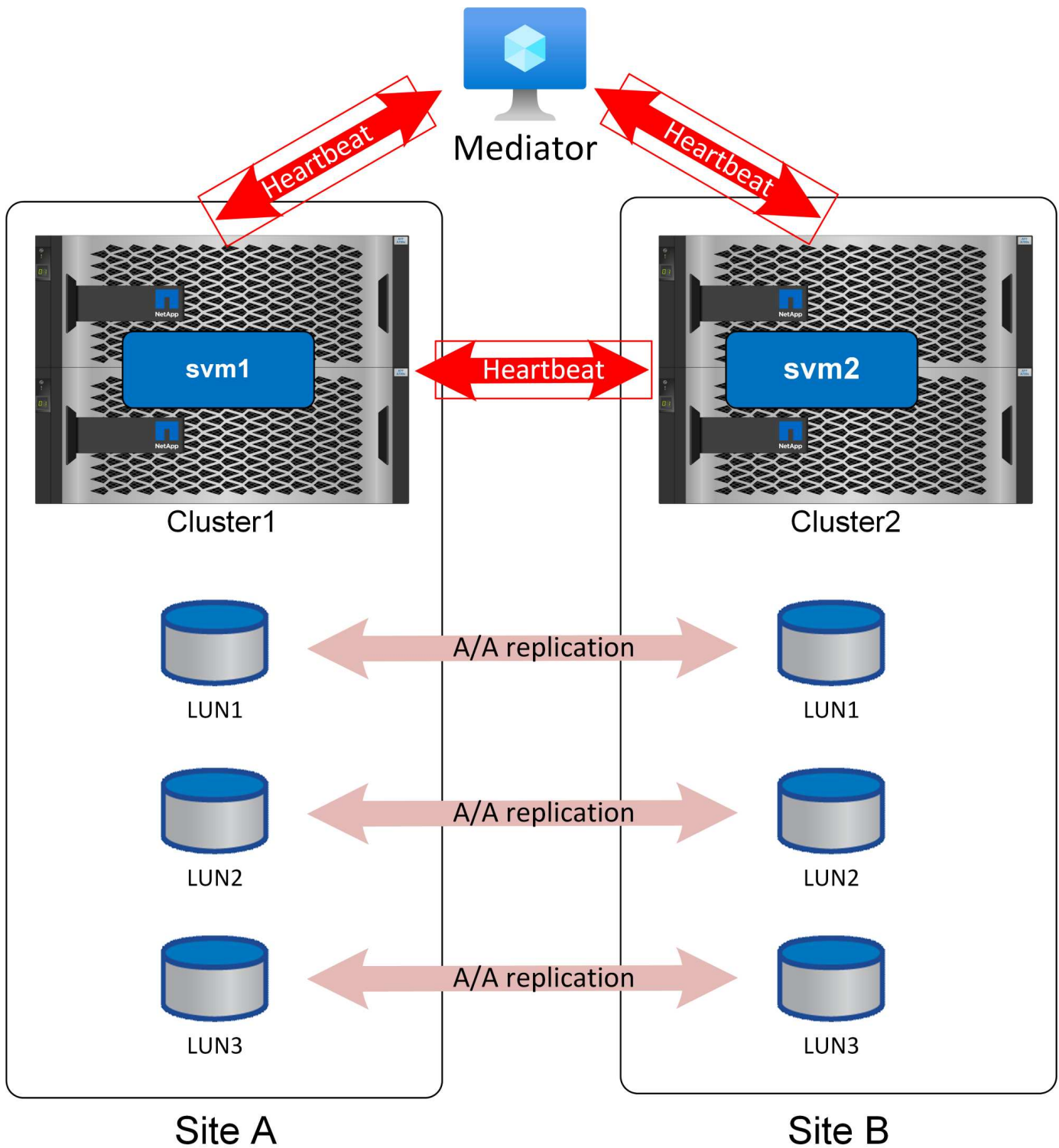
Mediador do ONTAP

O Mediador ONTAP é um aplicativo de software que é baixado do suporte do NetApp e normalmente é implantado em uma pequena máquina virtual. O Mediador ONTAP não é um tiebreaker quando usado com a sincronização ativa do SnapMirror. É um canal de comunicação alternativo para os dois clusters que participam da replicação de sincronização ativa do SnapMirror. As operações automatizadas são orientadas pelo ONTAP com base nas respostas recebidas do parceiro por meio de conexões diretas e por meio do mediador.

ONTAP Mediador

O mediador é necessário para automatizar o failover com segurança. Idealmente, ele seria colocado em um local 3rd independente, mas ainda pode funcionar para a maioria das necessidades se colocasse em um dos clusters que participam da replicação.

O mediador não é realmente um tiebreaker, embora essa seja efetivamente a função que ele fornece. Ele não realiza nenhuma ação; em vez disso, fornece um canal de comunicação alternativo para comunicação de cluster para cluster.



O desafio nº 1 com failover automatizado é o problema de split-brain, e esse problema surge se os dois locais perderem a conexão entre si. O que deve acontecer? Você não quer que dois sites diferentes se designem como as cópias sobreviventes dos dados, mas como um único site pode dizer a diferença entre a perda real do site oposto e a incapacidade de se comunicar com o site oposto?

É aqui que o mediador entra na imagem. Se for colocado em um site 3rd e cada site tiver uma conexão de rede separada com esse site, então você terá um caminho adicional para cada site validar a integridade do outro. Olhe para a imagem acima novamente e considere os seguintes cenários.

- O que acontece se o mediador falhar ou não estiver acessível a partir de um ou de ambos os sites?
 - Os dois clusters ainda podem se comunicar entre si pelo mesmo link usado para serviços de replicação.
 - Os dados ainda são servidos com proteção RPO igual a 0
- O que acontece se o Site A falhar?
 - O local B verá ambos os canais de comunicação diminuírem.
 - O local B assumirá os serviços de dados, mas sem o espelhamento RPO igual a 0
- O que acontece se o local B falhar?
 - O local A verá ambos os canais de comunicação diminuírem.
 - O local A assumirá os serviços de dados, mas sem o espelhamento do RPO igual a 0

Há um outro cenário a considerar: Perda do link de replicação de dados. Se o link de replicação entre locais for perdido, o espelhamento RPO 0 obviamente será impossível. O que deve acontecer então?

Isso é controlado pelo status do site preferido. Em uma relação SM-as, um dos locais é secundário ao outro. Isso não tem efeito nas operações normais, e todo o acesso aos dados é simétrico, mas se a replicação for interrompida, o empate terá que ser quebrado para retomar as operações. O resultado é que o local preferido continuará as operações sem espelhamento, e o local secundário interromperá o processamento de e/S até que a comunicação de replicação seja restaurada.

Site preferido para sincronização ativa do SnapMirror

O comportamento de sincronização ativa do SnapMirror é simétrico, com uma exceção importante - configuração de site preferida.

A sincronização ativa do SnapMirror considerará um site a "fonte" e o outro o "destino". Isso implica uma relação de replicação unidirecional, mas isso não se aplica ao comportamento de IO. A replicação é bidirecional e simétrica, e os tempos de resposta de e/S são os mesmos em ambos os lados do espelho.

A `source` designação é controla o local preferido. Se o link de replicação for perdido, os caminhos de LUN na cópia de origem continuarão a servir dados enquanto os caminhos de LUN na cópia de destino ficarão indisponíveis até que a replicação seja restabelecida e o SnapMirror reinsira um estado síncrono. Os caminhos irão então retomar a veiculação de dados.

A configuração de origem/destino pode ser visualizada através do SystemManager:

The screenshot shows the 'Relationships' section in SystemManager, with the 'Local sources' tab selected. The table below displays a single relationship between two local sources.

Source	Destination	Policy type
jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	Synchronous

Ou na 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
```

O segredo é que a fonte é o SVM no cluster1. Como mencionado acima, os termos "fonte" e "destino" não descrevem o fluxo de dados replicados. Ambos os sites podem processar uma gravação e replicá-la para o site oposto. Com efeito, ambos os clusters são fontes e destinos. O efeito de designar um cluster como uma fonte simplesmente controla qual cluster sobrevive como um sistema de armazenamento de leitura e gravação se o link de replicação for perdido.

Topologia de rede

Acesso uniforme

Rede de acesso uniforme significa que os hosts são capazes de acessar caminhos em ambos os sites (ou domínios de falha dentro do mesmo site).

Um recurso importante do SM-as é a capacidade de configurar os sistemas de storage para saber onde os hosts estão localizados. Quando você mapeia os LUNs para um determinado host, você pode indicar se eles são ou não proximais a um determinado sistema de armazenamento.

Definições de proximidade

Proximidade refere-se a uma configuração por cluster que indica que um determinado host WWN ou ID de iniciador iSCSI pertence a um host local. É uma segunda etapa opcional para configurar o acesso LUN.

O primeiro passo é a configuração usual do igroup. Cada LUN deve ser mapeado para um grupo que contenha as IDs WWN/iSCSI dos hosts que precisam de acesso a esse LUN. Isso controla qual host tem *access* para um LUN.

A segunda etapa opcional é configurar a proximidade do host. Isso não controla o acesso, ele controla *Priority*.

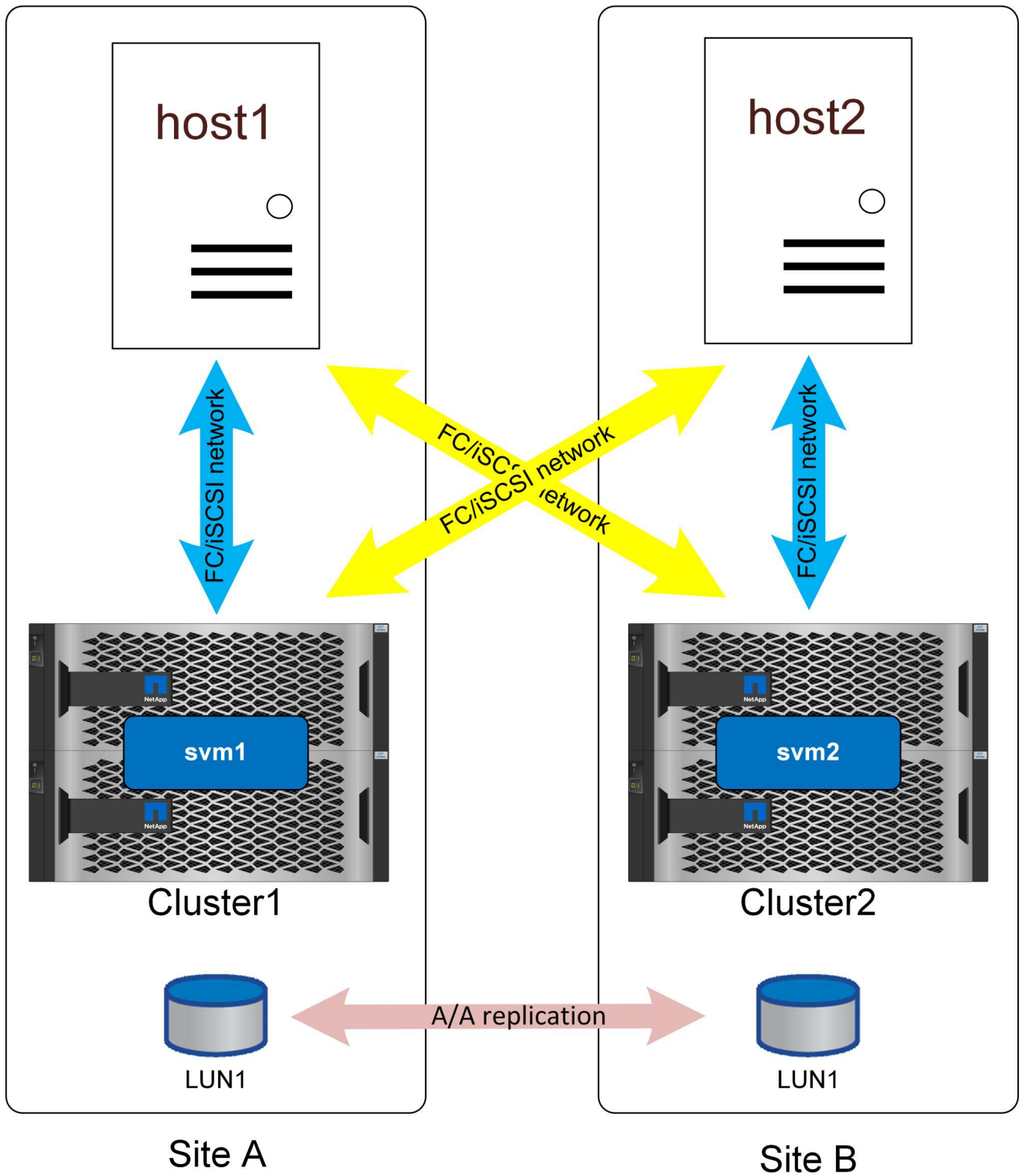
Por exemplo, um host no local A pode ser configurado para acessar um LUN que é protegido pela sincronização ativa do SnapMirror e, como a SAN é estendida entre sites, há caminhos disponíveis para esse LUN usando armazenamento no local A ou armazenamento no local B.

Sem configurações de proximidade, esse host usará ambos os sistemas de storage igualmente porque ambos os sistemas de storage anunciarão caminhos ativos/otimizados. Se a latência da SAN e/ou a largura de banda entre locais for limitada, isso pode não ser desejado e você pode querer garantir que, durante a operação normal, cada host utilize preferencialmente caminhos para o sistema de armazenamento local. Isso é

configurado adicionando o ID WWN/iSCSI do host ao cluster local como um host proximal. Isso pode ser feito na CLI ou no SystemManager.

AFF

Com um sistema AFF, os caminhos aparecerão como mostrado abaixo quando a proximidade do host for configurada.



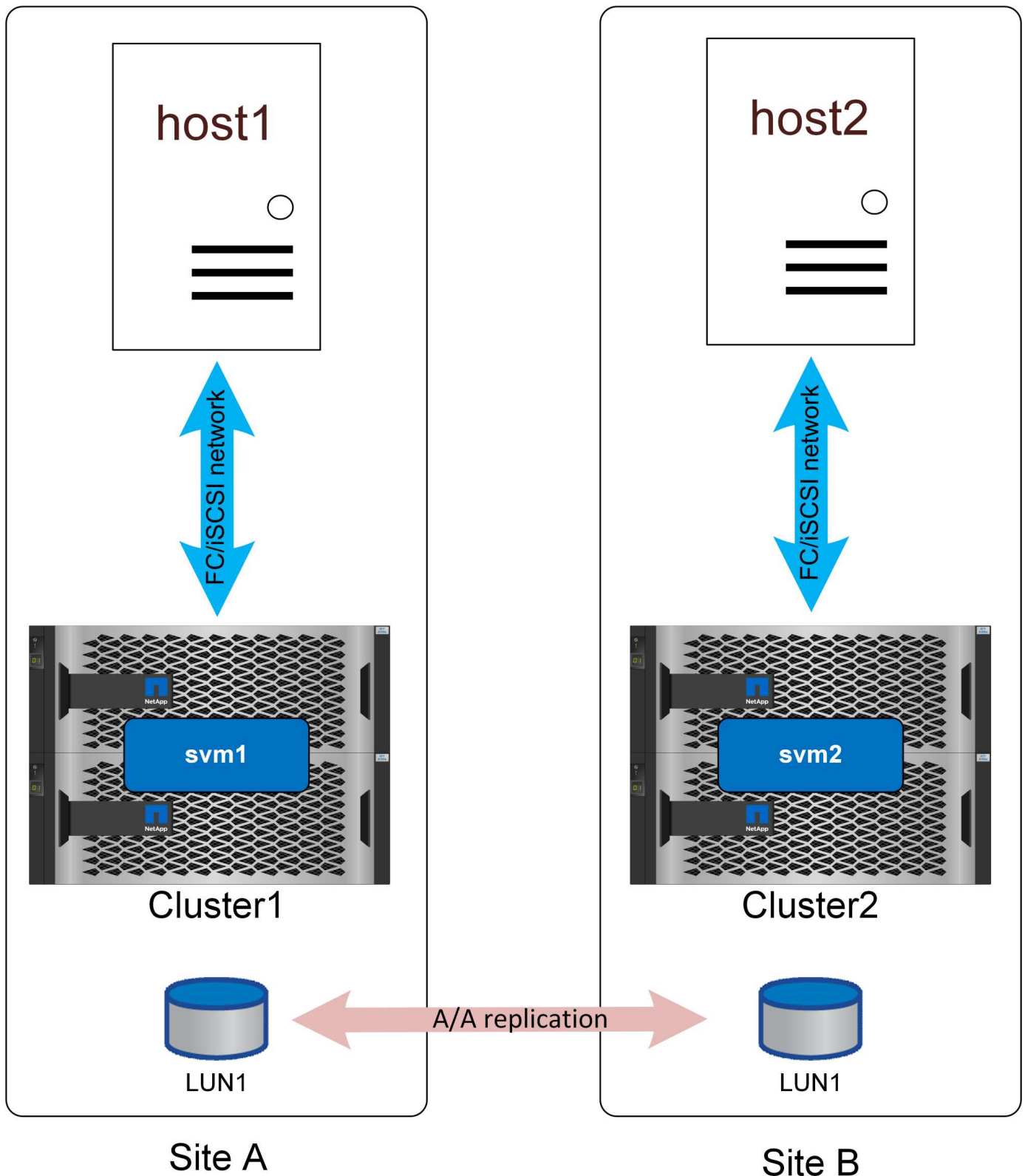
Em operação normal, todo o IO é IO local. Leituras e gravações são atendidas a partir do storage array local. É claro que o write IO também precisará ser replicado pelo controlador local para o sistema remoto antes de ser reconhecido, mas todas as IO de leitura serão atendidas localmente e não incorrerão latência extra ao atravessar o link SAN entre locais.

A única vez que os caminhos não otimizados serão usados é quando todos os caminhos ativos/otimizados forem perdidos. Por exemplo, se todo o array no local perder energia, os hosts no local A ainda poderão acessar caminhos para o array no local B e, portanto, permanecer operacionais, embora estejam com maior latência.

Há caminhos redundantes pelo cluster local que não são mostrados nesses diagramas por uma questão de simplicidade. Os sistemas de storage da ONTAP estão HA, portanto, uma falha da controladora não deve resultar em falha do local. Deve apenas resultar em uma mudança na qual os caminhos locais são usados no site afetado.

ASA

Os sistemas NetApp ASA oferecem multipathing ativo-ativo em todos os caminhos em um cluster. Isso também se aplica às configurações SM-as.



Active/Optimized Path

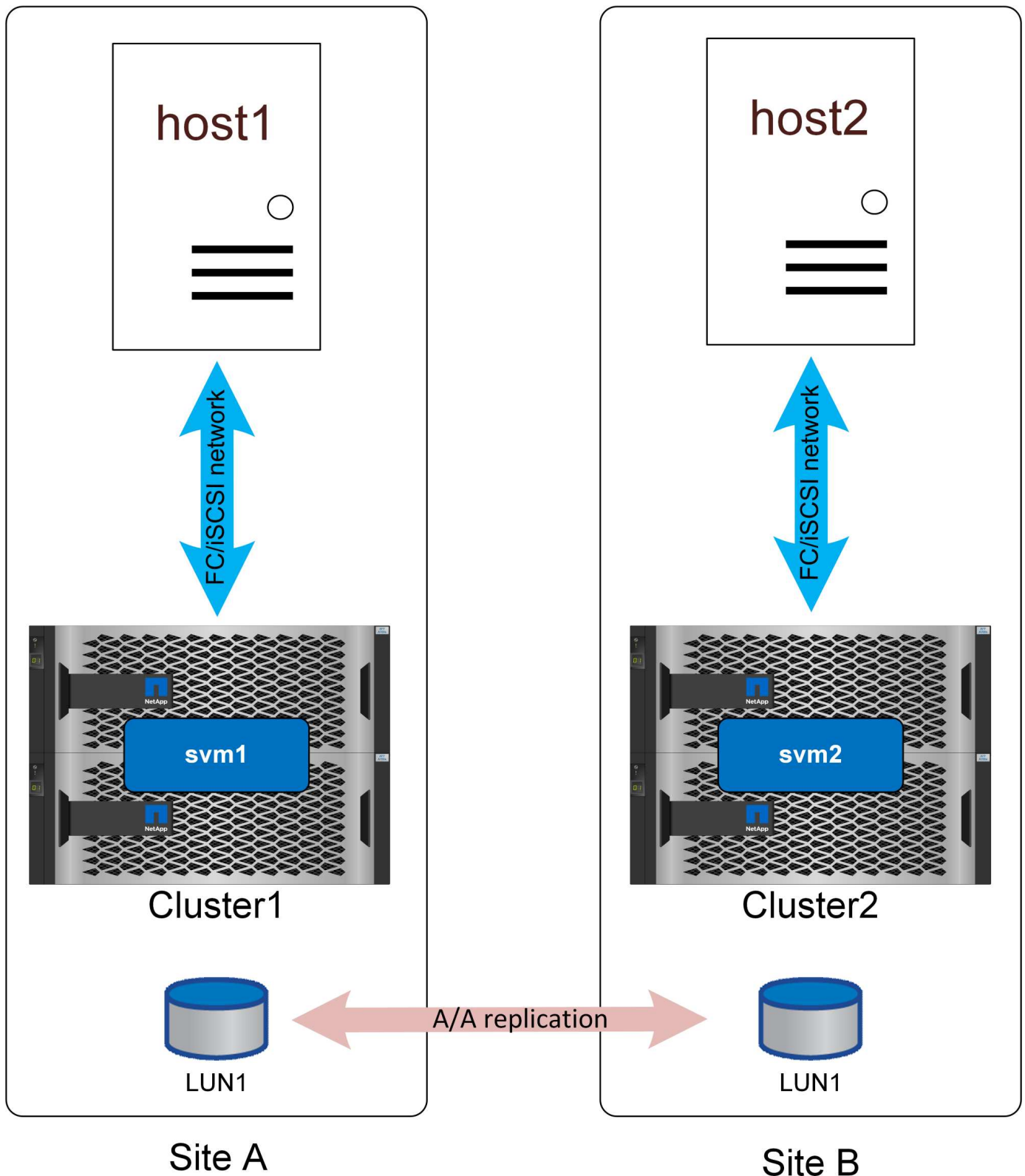
Uma configuração ASA com acesso não uniforme funcionaria em grande parte da mesma forma que faria com o AFF. Com acesso uniforme, o IO estaria atravessando a WAN. Isto pode ou não ser desejável.

Se os dois locais estivessem a 100 metros de distância com conectividade de fibra, não deveria haver latência adicional detectável cruzando a WAN, mas se os locais estivessem a uma distância longa, então o desempenho de leitura sofreria em ambos os locais. Em contraste, com o AFF, esses caminhos de cruzamento de WAN só seriam usados se não houvesse caminhos locais disponíveis e o desempenho do dia-a-dia seria melhor porque todo o IO seria IO local. O ASA com rede de acesso não uniforme seria uma opção para obter os benefícios de custo e recursos do ASA sem incorrer em uma penalidade de acesso à latência entre locais.

O ASA com SM-as em uma configuração de baixa latência oferece dois benefícios interessantes. Primeiro, ele basicamente **dobra** a performance de qualquer host porque a e/S pode ser atendida por duas vezes mais controladoras usando o dobro de caminhos. Em segundo lugar, em um ambiente de local único, ele oferece disponibilidade extrema porque todo um sistema de storage pode ser perdido sem interromper o acesso de host.

Acesso não uniforme

A rede de acesso não uniforme significa que cada host só tem acesso às portas no sistema de storage local. A SAN não é estendida entre sites (ou domínios de falha dentro do mesmo site).



Active/Optimized Path

O principal benefício dessa abordagem é a simplicidade da SAN - você elimina a necessidade de estender uma SAN pela rede. Alguns clientes não têm conectividade de baixa latência suficiente entre locais ou não têm

infraestrutura para túnel do tráfego SAN FC em uma rede entre locais.

A desvantagem para o acesso não uniforme é que certos cenários de falha, incluindo a perda do link de replicação, resultarão em alguns hosts perdendo acesso ao armazenamento. Os aplicativos que são executados como instâncias únicas, como um banco de dados não agrupado, que inerentemente está sendo executado apenas em um único host em qualquer montagem, falharão se a conectividade de armazenamento local for perdida. Os dados ainda seriam protegidos, mas o servidor de banco de dados não teria mais acesso. Ele precisaria ser reiniciado em um local remoto, de preferência através de um processo automatizado. Por exemplo, o VMware HA pode detectar uma situação de todos os caminhos em um servidor e reiniciar uma VM em outro servidor onde os caminhos estão disponíveis.

Em contraste, um aplicativo em cluster, como o Oracle RAC, pode fornecer um serviço que está disponível simultaneamente em dois locais diferentes. Perder um site não significa perda do serviço do aplicativo como um todo. As instâncias ainda estão disponíveis e em execução no local sobrevivente.

Em muitos casos, a sobrecarga de latência adicional de um aplicativo que acessa o storage em um link local a local seria inaceitável. Isso significa que a disponibilidade aprimorada de redes uniformes é mínima, uma vez que a perda de armazenamento em um local levaria à necessidade de encerrar serviços nesse local com falha de qualquer maneira.

Há caminhos redundantes pelo cluster local que não são mostrados nesses diagramas por uma questão de simplicidade. Os sistemas de storage da ONTAP estão HA, portanto, uma falha da controladora não deve resultar em falha do local. Deve apenas resultar em uma mudança na qual os caminhos locais são usados no site afetado.

Configurações Oracle

Visão geral

O uso da sincronização ativa do SnapMirror não necessariamente adiciona ou altera quaisquer práticas recomendadas para operar um banco de dados.

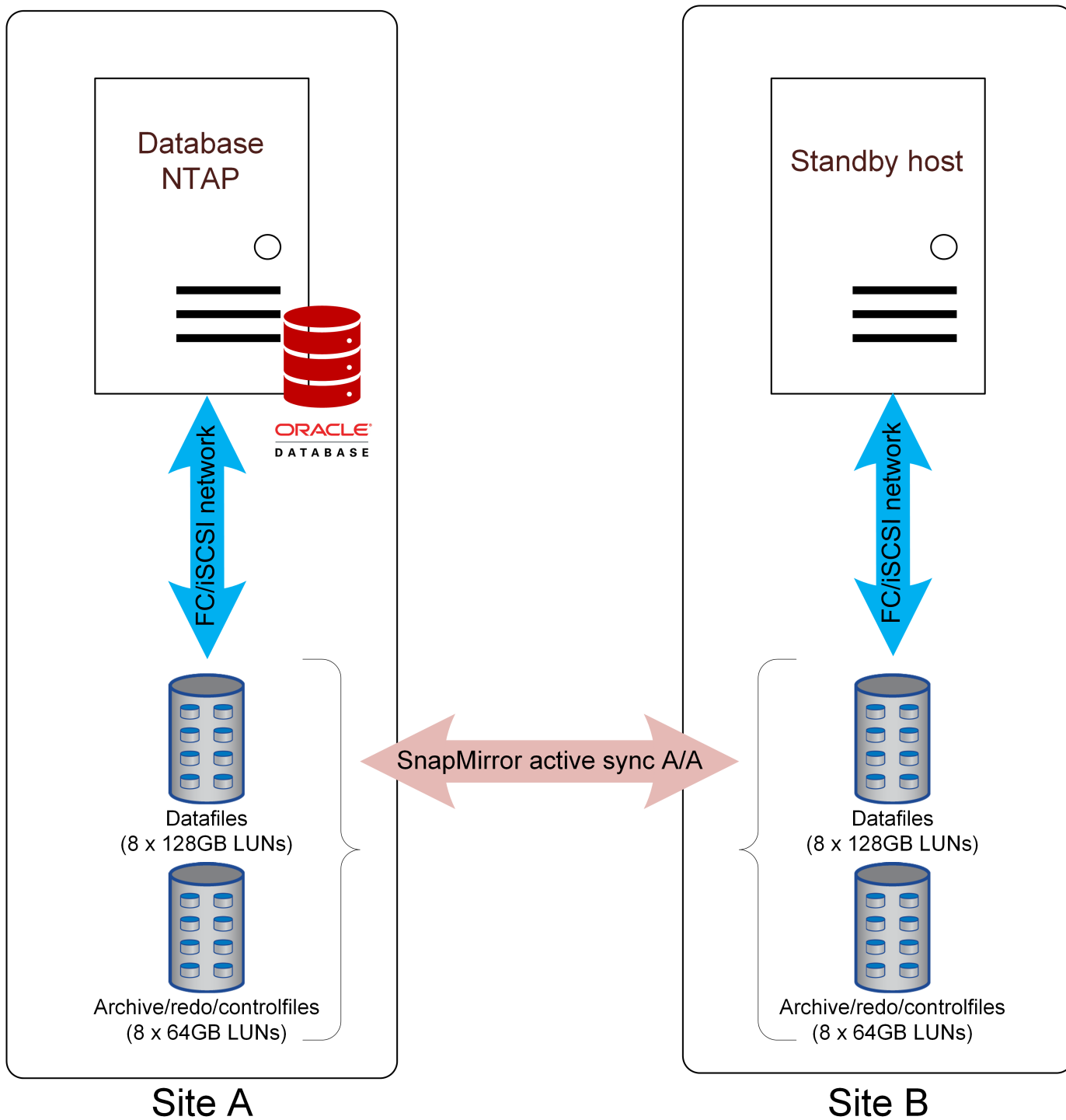
A melhor arquitetura depende dos requisitos de negócios. Por exemplo, se o objetivo é ter proteção RPO igual a 0 contra a perda de dados, mas o rto estiver relaxado, o uso de bancos de dados de Instância única Oracle e a replicação dos LUNs com SM-as pode ser suficiente e menos caro de um padrão de licenciamento Oracle. A falha do local remoto não interromperia as operações e a perda do local principal resultaria em LUNs no local sobrevivente que estão on-line e prontos para serem usados.

Se o rto fosse mais rigoroso, a automação ativo-passivo básica por meio de scripts ou clusterware, como pacemaker ou Ansible, melhoraria o tempo de failover. Por exemplo, o VMware HA pode ser configurado para detectar falha de VM no local principal e ativar a VM no local remoto.

Finalmente, para um failover extremamente rápido, o Oracle RAC poderia ser implantado em todos os locais. O rto seria essencialmente zero porque o banco de dados estaria online e disponível em ambos os sites em todos os momentos.

Instância única Oracle

Os exemplos explicados abaixo mostram algumas das muitas opções para implantar bancos de dados de Instância única Oracle com replicação de sincronização ativa do SnapMirror.



Failover com um SO pré-configurado

O SnapMirror active Sync fornece uma cópia síncrona dos dados no local de recuperação de desastres, mas disponibilizar esses dados requer um sistema operacional e as aplicações associadas. A automação básica pode melhorar significativamente o tempo de failover do ambiente geral. Os produtos Clusterware, como o pacemaker, costumam ser usados para criar um cluster nos sites e, em muitos casos, o processo de failover pode ser conduzido com scripts simples.

Se os nós primários forem perdidos, o clusterware (ou scripts) colocará os bancos de dados on-line no site alternativo. Uma opção é criar servidores em espera pré-configurados para os recursos SAN que compõem o banco de dados. Se o site principal falhar, a alternativa clusterware ou scripted executa uma sequência de

ações semelhantes às seguintes:

1. Detectar falha do local principal
2. Realize a descoberta de LUNs FC ou iSCSI
3. Montagem de sistemas de arquivos e/ou montagem de grupos de discos ASM
4. Iniciando o banco de dados

O principal requisito dessa abordagem é um sistema operacional em execução no local remoto. Ele deve ser pré-configurado com binários Oracle, o que também significa que tarefas como patches Oracle devem ser executadas no site primário e em espera. Como alternativa, os binários Oracle podem ser espelhados para o local remoto e montados se um desastre for declarado.

O procedimento de ativação real é simples. Comandos como o reconhecimento LUN requerem apenas alguns comandos por porta FC. A montagem do sistema de arquivos não é mais do que um `mount` comando, e os bancos de dados e ASM podem ser iniciados e parados na CLI com um único comando.

Failover com um sistema operacional virtualizado

O failover de ambientes de banco de dados pode ser estendido para incluir o próprio sistema operacional. Em teoria, esse failover pode ser feito com LUNs de inicialização, mas na maioria das vezes é feito com um sistema operacional virtualizado. O procedimento é semelhante aos seguintes passos:

1. Detectar falha do local principal
2. Montagem dos armazenamentos de dados que hospedam as máquinas virtuais do servidor de banco de dados
3. Iniciar as máquinas virtuais
4. Iniciando bancos de dados manualmente ou configurando as máquinas virtuais para iniciar automaticamente os bancos de dados.

Por exemplo, um cluster ESX pode abranger locais. Em caso de desastre, as máquinas virtuais podem ser colocadas on-line no local de recuperação de desastres após o switchover.

Proteção contra falha de storage

O diagrama acima mostra o uso "[acesso não uniforme](#)" do , em que a SAN não é estendida nos locais. Isso pode ser mais simples de configurar e, em alguns casos, pode ser a única opção, dada a capacidade de SAN atual, mas também significa que a falha do sistema de storage primário causaria uma interrupção do banco de dados até que o aplicativo fosse failover.

Para obter resiliência adicional, a solução poderia ser implantada com "[acesso uniforme](#)"o . Isso permitiria que os aplicativos continuassem operando usando os caminhos anunciados a partir do site oposto.

Oracle Extended RAC

Muitos clientes otimizam seu rto alongando um cluster do Oracle RAC entre locais, gerando uma configuração totalmente ativo-ativo. O projeto geral se torna mais complicado porque deve incluir o gerenciamento de quórum do Oracle RAC.

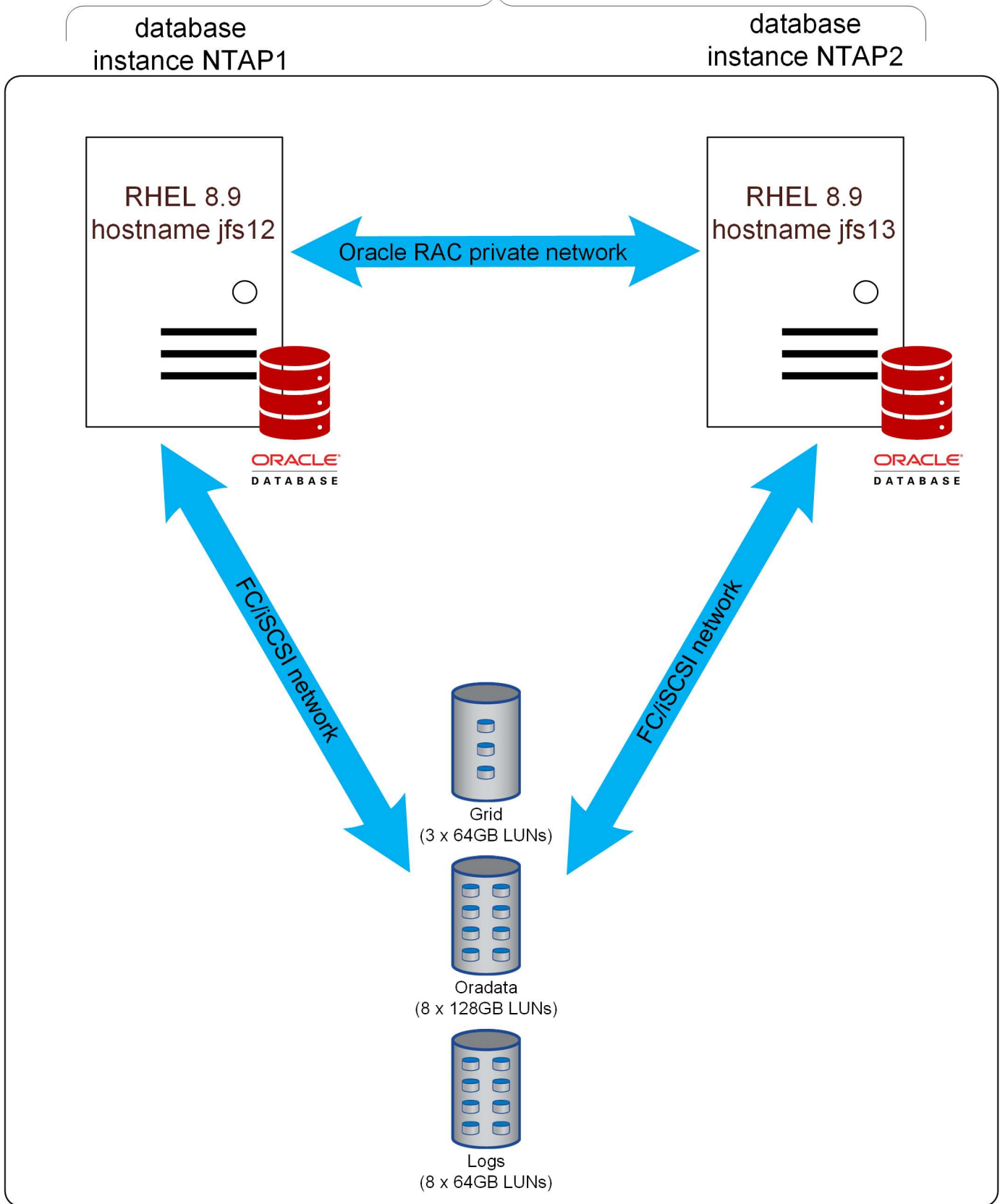
O RAC estendido tradicional em cluster contou com o espelhamento ASM para fornecer proteção de dados. Essa abordagem funciona, mas também requer muitas etapas de configuração manual e impõe sobrecarga na infraestrutura de rede. Em contraste, permitir que o SnapMirror ative Sync assuma a responsabilidade pela replicação de dados simplifica significativamente a solução. Operações como sincronização, ressincronização

após interrupções, failovers e gerenciamento de quórum são mais fáceis, e a SAN não precisa ser distribuída entre locais, o que simplifica o design e o gerenciamento da SAN.

Replicação

A chave para entender a funcionalidade RAC no SnapMirror ativo Sync é visualizar o armazenamento como um único conjunto de LUNs hospedados no armazenamento espelhado. Por exemplo:

Database NTAP



Não há cópia primária ou cópia espelhada. Logicamente, há apenas uma única cópia de cada LUN e esse LUN está disponível em caminhos SAN localizados em dois sistemas de armazenamento diferentes. Do ponto de vista do host, não há failovers de storage; em vez disso, há alterações de caminho. Vários eventos de falha

podem levar à perda de certos caminhos para o LUN, enquanto outros caminhos permanecem on-line. A sincronização ativa do SnapMirror garante que os mesmos dados estejam disponíveis em todos os caminhos operacionais.

Configuração de armazenamento

Neste exemplo de configuração, os discos ASM são configurados da mesma forma que seriam em qualquer configuração RAC de local único no storage empresarial. Como o sistema de armazenamento fornece proteção de dados, a redundância externa ASM seria usada.

Uniforme vs acesso não informado

A consideração mais importante com o Oracle RAC na sincronização ativa do SnapMirror é usar acesso uniforme ou não uniforme.

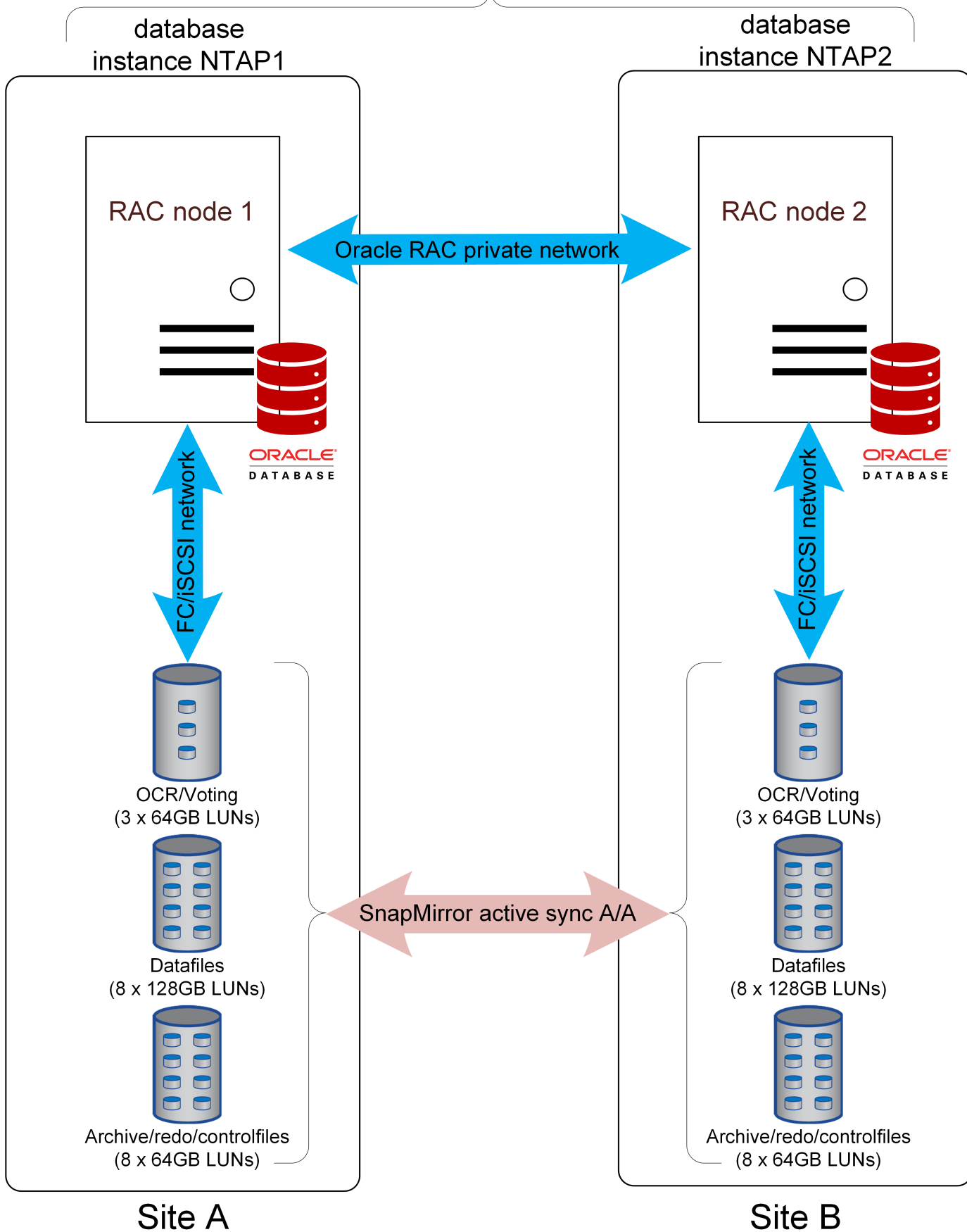
O acesso uniforme significa que cada host pode ver caminhos em ambos os clusters. O acesso não uniforme significa que os hosts só podem ver caminhos para o cluster local.

Nenhuma das opções é especificamente recomendada ou desencorajada. Alguns clientes têm fibra escura prontamente disponível para conectar sites, outros não têm essa conectividade ou sua infraestrutura de SAN não oferece suporte a um ISL de longa distância.

Acesso não uniforme

O acesso não uniforme é mais simples de configurar do ponto de vista da SAN.

Database NTAP



A principal desvantagem "[acesso não uniforme](#)" da abordagem é que a perda de conectividade ONTAP site a site ou a perda de um sistema de storage resultará na perda de instâncias de banco de dados em um local. Isso obviamente não é desejável, mas pode ser um risco aceitável em troca de uma configuração SAN mais simples.

Acesso uniforme

O acesso uniforme requer a extensão da SAN entre os locais. O principal benefício é que a perda de um sistema de storage não resultará na perda de uma instância de banco de dados. Em vez disso, isso resultaria em uma mudança multipathing em que os caminhos estão atualmente em uso.

Existem várias maneiras de configurar o acesso não uniforme.

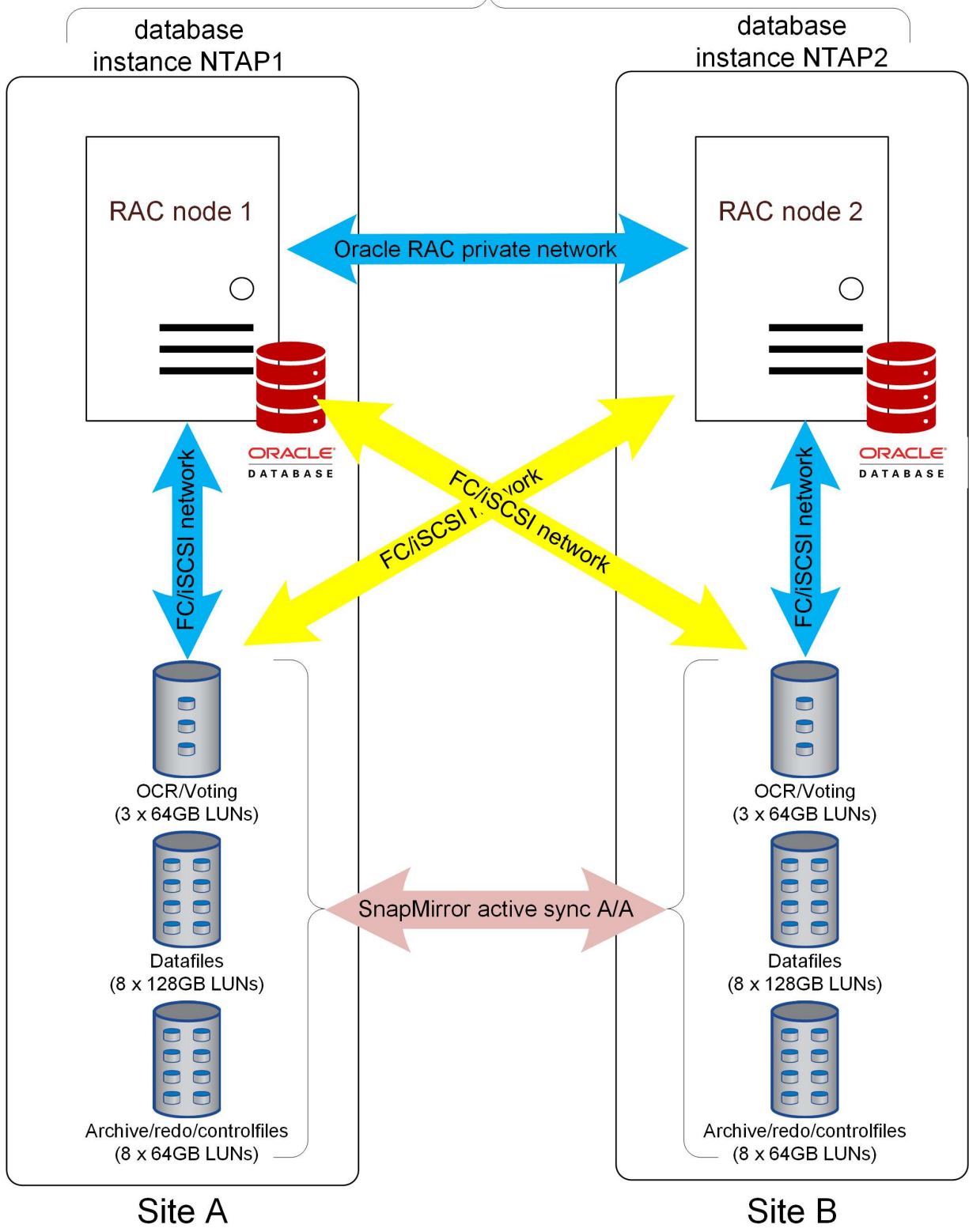


Nos diagramas abaixo, há também caminhos ativos, mas não otimizados, que seriam usados durante falhas simples do controlador, mas esses caminhos não são mostrados no interesse de simplificar os diagramas.

AFF com definições de proximidade

Se houver latência significativa entre sites, os sistemas AFF podem ser configurados com configurações de proximidade do host. Isso permite que cada sistema de armazenamento esteja ciente de quais hosts são locais e quais são remotos e atribua prioridades de caminho adequadamente.

Database NTAP



Active/Optimized Path

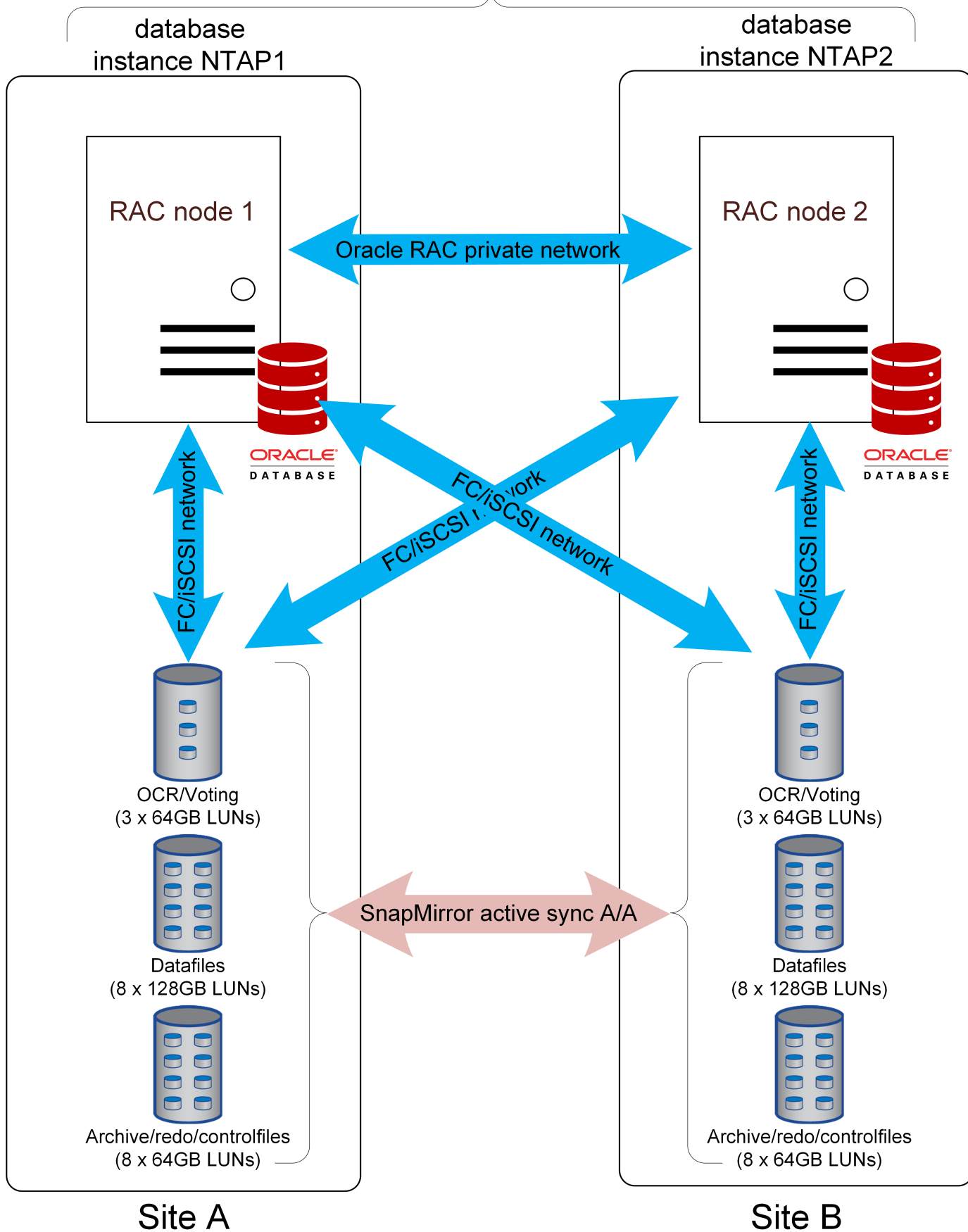
Active Path

Na operação normal, cada instância do Oracle usaria preferencialmente os caminhos ativos/otimizados locais. O resultado é que todas as leituras seriam atendidas pela cópia local dos blocos. Isso produz a menor latência possível. O write IO é enviado de forma semelhante para o controlador local. O IO ainda deve ser replicado antes de ser reconhecido e, portanto, ainda incorreria na latência adicional de cruzar a rede local a local, mas isso não pode ser evitado em uma solução de replicação síncrona.

ASA / AFF sem definições de proximidade

Se não houver latência significativa entre sites, os sistemas AFF podem ser configurados sem configurações de proximidade do host ou ASA podem ser usados.

Database NTAP



Cada host poderá usar todos os caminhos operacionais em ambos os sistemas de storage. Isso potencialmente melhora o desempenho de maneira significativa, permitindo que cada host aproveite o potencial de desempenho de dois clusters, e não apenas um.

Com o ASA, não só todos os caminhos para ambos os clusters seriam considerados ativos e otimizados, como também os caminhos nos controladores do parceiro estariam ativos. O resultado seria caminhos SAN all-ativos em todo o cluster, o tempo todo.



Os sistemas ASA também podem ser usados em uma configuração de acesso não uniforme. Uma vez que não existem caminhos entre locais, não haveria impactos no desempenho resultante do IO cruzando o ISL.

Desempate do RAC

Embora o RAC estendido usando o SnapMirror ative Sync seja uma arquitetura simétrica com relação ao IO, há uma exceção que está conetada ao gerenciamento de split-brain.

O que acontece se o link de replicação for perdido e nenhum dos sites tiver quórum? O que deve acontecer? Esta pergunta se aplica ao comportamento do Oracle RAC e do ONTAP. Se as alterações não puderem ser replicadas nos sites e você quiser retomar as operações, um dos sites terá que sobreviver e o outro site terá que ficar indisponível.

O "ONTAP Mediador" atende a esse requisito na camada ONTAP. Existem várias opções para quebra de binário RAC.

Tiebreakers Oracle

O melhor método para gerenciar os riscos do Oracle RAC dividido é usar um número ímpar de nós RAC, de preferência pelo uso de um desempate de site 3rd. Se um site 3rd não estiver disponível, a instância tiebreaker pode ser colocada em um local dos dois sites, designando-o efetivamente um local sobrevivente preferido.

Oracle e CSS_critical

Com um número par de nós, o comportamento padrão do Oracle RAC é que um dos nós no cluster será considerado mais importante do que os outros nós. O local com esse nó de maior prioridade sobreviverá ao isolamento do local, enquanto os nós do outro local serão despejados. A priorização é baseada em vários fatores, mas você também pode controlar esse comportamento usando a `css_critical` configuração.

Na "exemplo" arquitetura, os nomes de host para os nós RAC são `jfs12` e `jfs13`. As definições atuais para `css_critical` são as seguintes:

```
[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 você quiser que o site com `jfs12` seja o site preferido, altere esse valor para `sim` em um nó. De site e reinicie os serviços.

```

[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.

```

Cenários de falha

Visão geral

Planejar uma arquitetura completa de aplicativos de sincronização ativa do SnapMirror requer entender como o SM-as responderá em vários cenários de failover planejados e não planejados.

Para os exemplos a seguir, suponha que o site A esteja configurado como o site preferido.

Perda de conectividade de replicação

Se a replicação SM-as for interrompida, não é possível concluir a e/S de gravação porque seria impossível que um cluster replique alterações no local oposto.

Local A (local preferido)

O resultado da falha do link de replicação no site preferido será uma pausa de aproximadamente 15 segundos no processamento de e/S de gravação, à medida que o ONTAP tenta novamente as operações de gravação replicadas antes de determinar que o link de replicação é genuinamente inacessível. Após os 15 segundos decorridos, o Site Um sistema retoma o processamento de e/S de leitura e escrita. Os caminhos de SAN não serão alterados e os LUNs permanecerão online.

Local B

Como o local B não é o site preferido de sincronização ativa do SnapMirror, seus caminhos de LUN ficarão indisponíveis após cerca de 15 segundos.

Falha do sistema de storage

O resultado de uma falha do sistema de armazenamento é quase idêntico ao resultado da perda do link de replicação. O local sobrevivente deve experimentar uma pausa de IO de aproximadamente 15 segundos. Uma vez decorrido esse período de 15 segundos, o IO será retomado nesse site como de costume.

Perda do mediador

O serviço mediador não controla diretamente as operações de storage. Ele funciona como um caminho de controle alternativo entre clusters. Ele existe principalmente para automatizar o failover sem o risco de um cenário de divisão cerebral. Em operação normal, cada cluster replica alterações em seu parceiro, e cada cluster pode verificar se o cluster do parceiro está on-line e fornecendo dados. Se o link de replicação falhar, a replicação cessaria.

O motivo pelo qual um mediador é necessário para o failover automatizado seguro é porque, de outra forma, seria impossível que um cluster de storage pudesse determinar se a perda de comunicação bidirecional foi o resultado de uma interrupção da rede ou falha real do storage.

O mediador fornece um caminho alternativo para cada cluster para verificar a integridade de seu parceiro. Os cenários são os seguintes:

- Se um cluster puder entrar em Contato diretamente com seu parceiro, os serviços de replicação estarão operacionais. Nenhuma ação necessária.
- Se um site preferido não puder entrar em Contato diretamente com seu parceiro ou por meio do mediador, ele assumirá que ele está realmente indisponível ou foi isolado e que levou seus caminhos LUN off-line. O site preferido continuará lançando o estado RPO/0 e continuará processando e/S de leitura e gravação.
- Se um site não preferencial não puder entrar em Contato diretamente com seu parceiro, mas puder contatá-lo por meio do mediador, ele tomará seus caminhos off-line e aguardará o retorno da conexão de replicação.
- Se um site não preferencial não puder entrar em Contato com seu parceiro diretamente ou por meio de um mediador operacional, ele assumirá que o parceiro está realmente indisponível ou foi isolado e que tomou seus caminhos LUN off-line. O site não preferencial lançará o estado RPO 0 e continuará processando e/S de leitura e gravação. Ele assumirá o papel da fonte de replicação e se tornará o novo site preferido.

Se o mediador não estiver totalmente disponível:

- A falha dos serviços de replicação por qualquer motivo, incluindo a falha do sistema de storage ou local não preferido, resultará no lançamento do estado RPO/0 e no reinício do processamento de e/S de leitura e gravação. O site não preferencial tomará seus caminhos off-line.
- A falha do site preferido resultará em uma falha porque o site não-preferido não será capaz de verificar se o site oposto está realmente off-line e, portanto, não seria seguro para o site não-preferido retomar os serviços.

Restauração de serviços

Depois que uma falha é resolvida, como restaurar a conectividade site-a-site ou ligar um sistema com falha, os pontos de extremidade de sincronização ativa do SnapMirror detetarão automaticamente a presença de uma relação de replicação com defeito e o devolverão ao estado RPO-0. Uma vez que a replicação síncrona for restabelecida, os caminhos com falha ficarão online novamente.

Em muitos casos, os aplicativos em cluster detetarão automaticamente o retorno de caminhos com falha, e esses aplicativos também voltarão online. Em outros casos, pode ser necessária uma análise SAN no nível do

host ou os aplicativos podem precisar ser colocados online manualmente. Depende do aplicativo e como ele é configurado e, em geral, essas tarefas podem ser facilmente automatizadas. O próprio ONTAP é com autorrecuperação e não deve exigir a intervenção do usuário para retomar as operações de storage RPO de 0.

Failover manual

Alterar o local preferido requer uma operação simples. A e/S pausa por um segundo ou dois como autoridade sobre os switches de comportamento de replicação entre clusters, mas a e/S não é afetada.

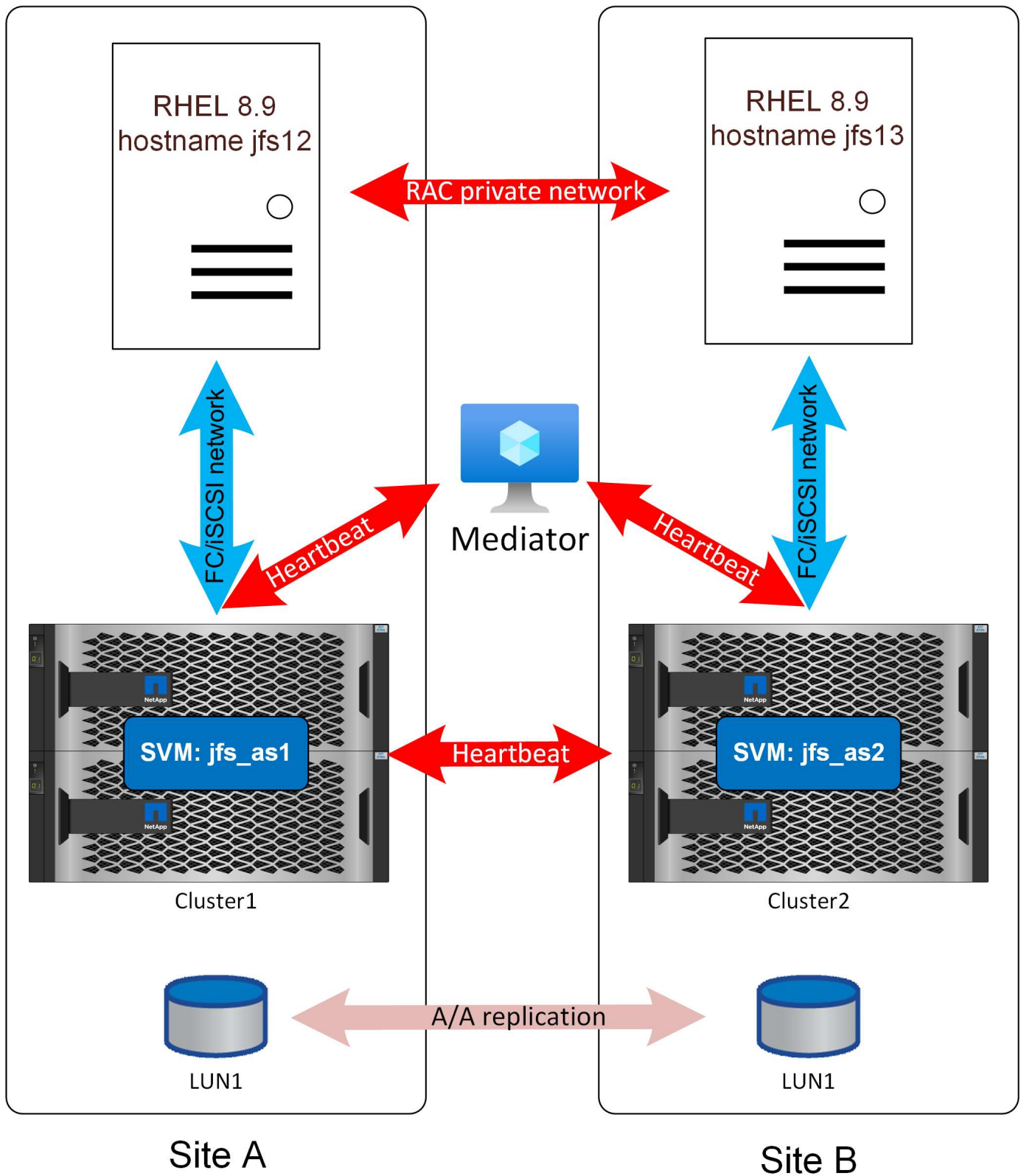
Arquitetura de amostra

Os exemplos de falha detalhados mostrados nestas seções são baseados na arquitetura mostrada abaixo.



Esta é apenas uma das muitas opções para bancos de dados Oracle na sincronização ativa do SnapMirror. Este design foi escolhido porque ilustra alguns dos cenários mais complicados.

Neste design, suponha que o local A esteja definido no "[site preferido](#)".



Falha na interconexão RAC

A perda do link de replicação do Oracle RAC produzirá um resultado semelhante à perda de conectividade do SnapMirror, exceto que os tempos limite serão menores por padrão. Sob as configurações padrão, um nó Oracle RAC aguardará 200 segundos após a perda

de conectividade de armazenamento antes de despejar, mas só aguardará 30 segundos após a perda do batimento cardíaco da rede RAC.

As mensagens CRS são semelhantes às mostradas abaixo. Você pode ver o intervalo de tempo limite de 30 segundos. Uma vez que `CSS_critical` foi definido em `jfs12`, localizado no local A, que será o local para sobreviver e `jfs13` no local B será despejado.

```
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'.
```

Falha de comunicação do SnapMirror

Se o link de replicação de sincronização ativa do SnapMirror, a e/S de gravação não puder ser concluída porque seria impossível que um cluster replique alterações no local oposto.

Local A

O resultado de uma falha no link de replicação no local A será uma pausa de aproximadamente 15 segundos no processamento de e/S de gravação, já que o ONTAP tenta replicar gravações antes de determinar que o link de replicação está genuinamente inoperável. Após os 15 segundos decorridos, o cluster do ONTAP no local A retoma o processamento de e/S de leitura e gravação. Os caminhos de SAN não serão alterados e os LUNs permanecerão online.

Local B

Como o local B não é o site preferido de sincronização ativa do SnapMirror, seus caminhos de LUN ficarão indisponíveis após cerca de 15 segundos.

O link de replicação foi cortado no carimbo de data/hora 15:19:44. O primeiro aviso do Oracle RAC chega 100 segundos depois, quando o tempo limite de 200 segundos (controlado pelo parâmetro Oracle RAC `disktimeout`) se aproxima.

```
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 CRS
resources.
```

Assim que o tempo limite do disco de votação de 200 segundos for atingido, esse nó Oracle RAC se despejará do cluster e reiniciará.

Falha total de interconetividade de rede

Se o link de replicação entre sites for completamente perdido, a sincronização ativa do SnapMirror e a conectividade do Oracle RAC serão interrompidas.

A detecção de split-brain do Oracle RAC tem uma dependência do batimento cardíaco do armazenamento do Oracle RAC. Se a perda de conectividade local a local resultar na perda simultânea dos serviços de replicação de armazenamento e batimento cardíaco da rede RAC, o resultado será que os sites RAC não poderão se comunicar entre locais através da interconexão RAC ou dos discos de votação RAC. O resultado em um conjunto de nós com dormência uniforme pode ser despejo de ambos os sites sob configurações padrão. O comportamento exato dependerá da sequência de eventos e do tempo das pesquisas de batimento cardíaco da rede RAC e do disco.

O risco de uma interrupção no 2 local pode ser resolvido de duas maneiras. Primeiro, uma **"desempate"** configuração pode ser usada.

Se um site 3rd não estiver disponível, esse risco pode ser resolvido ajustando o parâmetro `misscount` no cluster RAC. Sob os padrões, o tempo limite do batimento cardíaco da rede RAC é de 30 segundos. Isso normalmente é usado pelo RAC para identificar os nós RAC com falha e removê-los do cluster. Ele também

tem uma conexão com o heartbeat do disco de votação.

Se, por exemplo, o canal que transporta tráfego intersite para Oracle RAC e serviços de replicação de armazenamento for cortado por uma retroescavadeira, a contagem regressiva de 30 segundos de misscount começará. Se o nó do local preferencial RAC não puder restabelecer o Contato com o local oposto dentro de 30 segundos, e ele também não puder usar os discos de votação para confirmar que o local oposto está para baixo dentro dessa mesma janela de 30 segundos, os nós do local preferido também serão despejados. O resultado é uma interrupção completa do banco de dados.

Dependendo de quando a polling do misscount ocorre, 30 segundos podem não ser tempo suficiente para que a sincronização ativa do SnapMirror expire e permita que o armazenamento no site preferido retome os serviços antes que a janela de 30 segundos expire. Esta janela de 30 segundos pode ser aumentada.

```
[root@jfs12 ~]# /grid/bin/crsctl set css misscount 100
CRS-4684: Successful set of parameter misscount to 100 for Cluster
Synchronization Services.
```

Esse valor permite que o sistema de armazenamento no site preferido retome as operações antes que o tempo limite de contagem de erros expire. O resultado será, então, despejo apenas dos nós no local onde os caminhos LUN foram removidos. Exemplo abaixo:

```
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
```

O Oracle Support desencoraja fortemente a alteração com os parâmetros misscount ou disktimeout para resolver problemas de configuração. A alteração desses parâmetros pode, no entanto, ser garantida e inevitável em muitos casos, incluindo configurações de inicialização por SAN, virtualizadas e replicação de

armazenamento. Se, por exemplo, você tiver problemas de estabilidade com uma rede SAN ou IP que resultasse em despejos RAC, você deve corrigir o problema subjacente e não cobrar os valores do misscount ou disktimeout. Alterar tempos limite para resolver erros de configuração está mascarando um problema, não resolvendo um problema. Alterar esses parâmetros para configurar adequadamente um ambiente RAC com base em aspetos de design da infraestrutura subjacente é diferente e é consistente com as declarações de suporte Oracle. Com a inicialização SAN, é comum ajustar o misscount até 200 para corresponder ao disktimeout. ["este link"](#)Consulte para obter informações adicionais.

Falha do local

O resultado de uma falha do sistema de armazenamento ou do local é quase idêntico ao resultado da perda do link de replicação. O site sobrevivente deve experimentar uma pausa de IO de cerca de 15 segundos nas gravações. Uma vez decorrido esse período de 15 segundos, o IO será retomado nesse site como de costume.

Se apenas o sistema de armazenamento tiver sido afetado, o nó Oracle RAC no site com falha perderá os serviços de armazenamento e entrará na mesma contagem regressiva de tempo limite de disco de 200 segundos antes da remoção e reinicialização subsequente.

```

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 CRS
resources.

```

O estado do caminho SAN no nó RAC que perdeu os serviços de armazenamento é assim:

```

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

```

O host linux detetou a perda dos caminhos muito mais rápido do que 200 segundos, mas de uma perspetiva de banco de dados as conexões do cliente com o host no site com falha ainda serão congeladas por 200 segundos sob as configurações padrão do Oracle RAC. As operações de banco de dados completo só serão retomadas após a conclusão do despejo.

Enquanto isso, o nó Oracle RAC no local oposto registrará a perda do outro nó RAC. De outra forma, continua a funcionar como de costume.

```
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
```

Falha do mediador

O serviço mediador não controla diretamente as operações de storage. Ele funciona como um caminho de controle alternativo entre clusters. Ele existe principalmente para automatizar o failover sem o risco de um cenário de divisão cerebral.

Em operação normal, cada cluster replica alterações em seu parceiro, e cada cluster pode verificar se o cluster do parceiro está on-line e fornecendo dados. Se o link de replicação falhar, a replicação cessaria.

O motivo pelo qual um mediador é necessário para operações automatizadas seguras é porque, de outra forma, seria impossível que os clusters de storage pudessem determinar se a perda de comunicação bidirecional foi o resultado de uma interrupção da rede ou falha real do storage.

O mediador fornece um caminho alternativo para cada cluster para verificar a integridade de seu parceiro. Os cenários são os seguintes:

- Se um cluster puder entrar em Contato diretamente com seu parceiro, os serviços de replicação estarão operacionais. Nenhuma ação necessária.
- Se um site preferido não puder entrar em Contato diretamente com seu parceiro ou por meio do mediador, ele assumirá que ele está realmente indisponível ou foi isolado e que levou seus caminhos LUN off-line. O site preferido continuará lançando o estado RPO/0 e continuará processando e/S de leitura e gravação.
- Se um site não preferencial não puder entrar em Contato diretamente com seu parceiro, mas puder contatá-lo por meio do mediador, ele tomará seus caminhos off-line e aguardará o retorno da conexão de replicação.
- Se um site não preferencial não puder entrar em Contato com seu parceiro diretamente ou por meio de um mediador operacional, ele assumirá que o parceiro está realmente indisponível ou foi isolado e que

tomou seus caminhos LUN off-line. O site não preferencial lançará o estado RPO 0 e continuará processando e/S de leitura e gravação. Ele assumirá o papel da fonte de replicação e se tornará o novo site preferido.

Se o mediador não estiver totalmente disponível:

- A falha dos serviços de replicação por qualquer motivo resultará no lançamento do estado RPO 0 e no reinício do processamento de e/S de leitura e gravação. O site não preferencial tomará seus caminhos off-line.
- A falha do site preferido resultará em uma falha porque o site não-preferido não será capaz de verificar se o site oposto está realmente off-line e, portanto, não seria seguro para o site não-preferido retomar os serviços.

Restauração do serviço

SnapMirror é auto-cura. O SnapMirror active Sync detetará automaticamente a presença de uma relação de replicação com defeito e o levará de volta ao estado RPO igual a 0. Uma vez que a replicação síncrona for restabelecida, os caminhos ficarão online novamente.

Em muitos casos, os aplicativos em cluster detetarão automaticamente o retorno de caminhos com falha, e esses aplicativos também voltarão online. Em outros casos, pode ser necessária uma análise SAN no nível do host ou os aplicativos podem precisar ser colocados online manualmente.

Depende do aplicativo e de como ele é configurado e, em geral, essas tarefas podem ser facilmente automatizadas. O SnapMirror active Sync em si é auto-consertado e não deve exigir a intervenção do usuário para retomar as operações de storage RPO 0 depois que a energia e a conectividade forem restauradas.

Failover manual

O termo "failover" não se refere à direção da replicação com a sincronização ativa do SnapMirror porque é uma tecnologia de replicação bidirecional. Em vez disso, "failover" refere-se a qual sistema de armazenamento será o local preferido em caso de falha.

Por exemplo, você pode querer executar um failover para alterar o local preferido antes de encerrar um local para manutenção ou antes de executar um teste de DR.

Alterar o local preferido requer uma operação simples. A e/S pausa por um segundo ou dois como autoridade sobre os switches de comportamento de replicação entre clusters, mas a e/S não é afetada.

Exemplo de GUI:

Relationships

Local destinations

Local sources

Search Download Show/hide Filter

Source	Destination	Policy type
▼ jfs_as1:/cg/jfsAA	⋮ jfs_as2:/cg/jfsAA	Synchronous

- Edit
- Update
- Delete
- Failover

Exemplo de alterá-lo de volta através da 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 Path	Destination Path	Type	Status	start-time	end-time	Error Reason
jfs_as1:/cg/jfsAA	jfs_as2:/cg/jfsAA	planned	completed	9/11/2024 09:29:22	9/11/2024 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
```

Migração de banco de dados Oracle

Visão geral

Aproveitar os recursos de uma nova plataforma de storage tem um requisito inevitável; os dados devem ser colocados no novo sistema de storage. O ONTAP simplifica o processo de migração, incluindo migrações e atualizações do ONTAP para o ONTAP, importações de LUN estrangeiras e procedimentos para o uso do sistema operacional host ou do software de banco de dados Oracle diretamente.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-4534: Migração de bancos de dados Oracle para sistemas de armazenamento NetApp*

No caso de um novo projeto de banco de dados, isso não é uma preocupação porque os ambientes de banco de dados e aplicativos são construídos no lugar. A migração, no entanto, apresenta desafios especiais em relação à interrupção dos negócios, ao tempo necessário para a conclusão da migração, aos conjuntos de

habilidades necessários e à minimização dos riscos.

Scripts

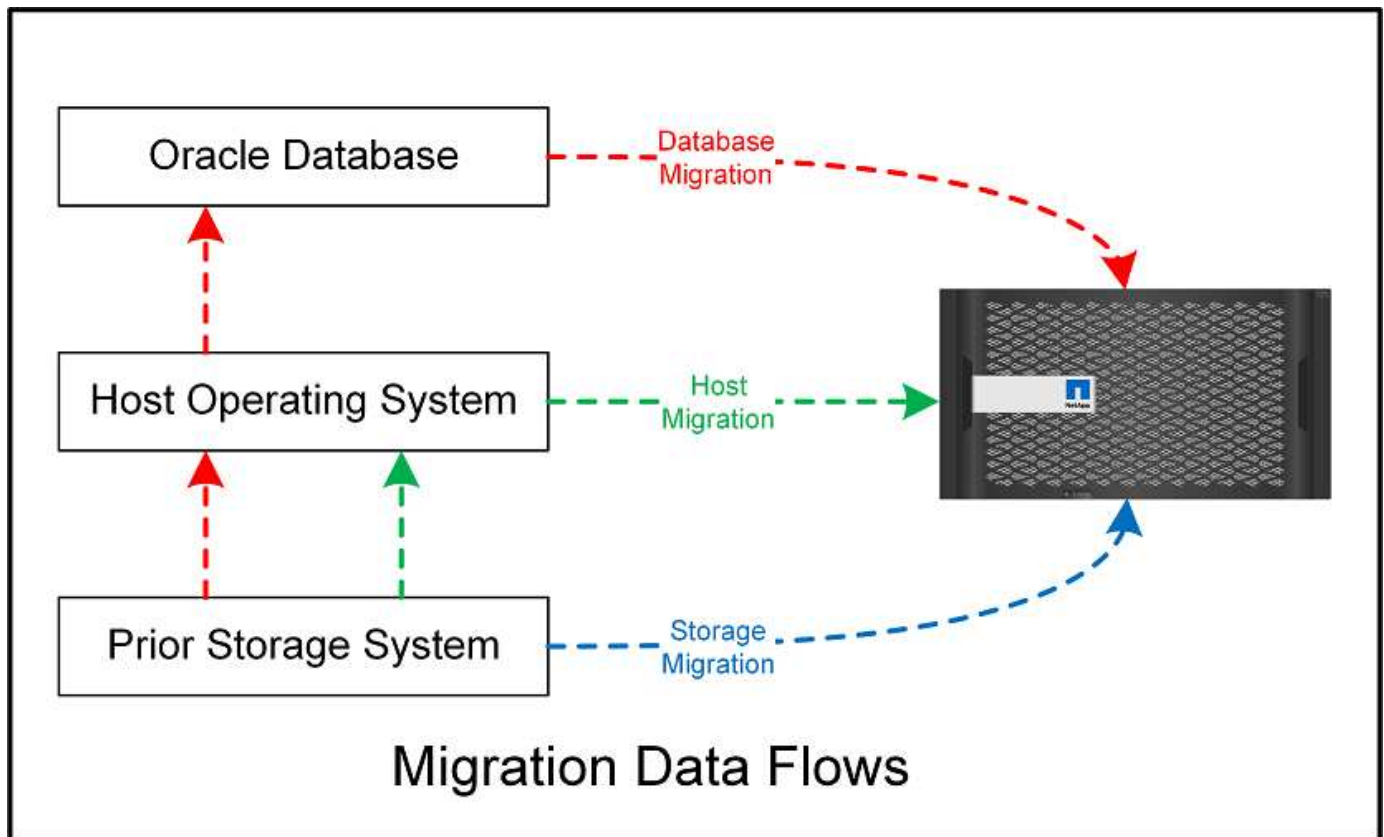
Exemplos de scripts são fornecidos nesta documentação. Esses scripts fornecem métodos de exemplo para automatizar vários aspectos da migração para reduzir a chance de erros do usuário. Os scripts podem reduzir as demandas gerais da equipe DE TI responsável por uma migração e acelerar o processo geral. Todos esses scripts são extraídos de projetos de migração reais executados pelos Serviços profissionais da NetApp e parceiros da NetApp. Exemplos de seu uso são mostrados ao longo desta documentação.

Planejamento de migração

A migração de dados Oracle pode ocorrer em um de três níveis: O banco de dados, o host ou o storage array.

As diferenças estão em qual componente da solução geral é responsável pela movimentação de dados: O banco de dados, o sistema operacional host ou o sistema de storage.

A figura abaixo mostra um exemplo dos níveis de migração e do fluxo de dados. No caso da migração no nível do banco de dados, os dados são movidos do sistema de storage original pelas camadas do host e do banco de dados para o novo ambiente. A migração em nível de host é semelhante, mas os dados não passam pela camada de aplicativo e são gravados no novo local usando processos de host. Finalmente, com a migração no nível do storage, um array como um sistema NetApp FAS é responsável pela movimentação de dados.



Uma migração no nível do banco de dados geralmente se refere ao uso do Oracle log enviado por meio de um banco de dados de reserva para concluir uma migração na camada Oracle. As migrações em nível de host são executadas usando a capacidade nativa da configuração do sistema operacional do host. Essa configuração inclui operações de cópia de arquivos usando comandos como cp, tar e Oracle Recovery Manager (RMAN) ou usando um gerenciador de volume lógico (LVM) para realocar os bytes subjacentes de

um sistema de arquivos. O Oracle Automatic Storage Management (ASM) é categorizado como um recurso no nível do host, porque é executado abaixo do nível do aplicativo de banco de dados. O ASM ocupa o lugar do gerenciador de volume lógico usual em um host. Finalmente, os dados podem ser migrados no nível de storage array, o que significa que estão abaixo do nível do sistema operacional.

Considerações de Planejamento

A melhor opção para migração depende de uma combinação de fatores, incluindo a escala do ambiente a ser migrado, a necessidade de evitar tempo de inatividade e o esforço geral necessário para realizar a migração. Bancos de dados grandes obviamente exigem mais tempo e esforço para a migração, mas a complexidade dessa migração é mínima. Pequenos bancos de dados podem ser migrados rapidamente, mas, se houver milhares para serem migrados, a escala do esforço pode criar complicações. Finalmente, quanto maior o banco de dados, maior a probabilidade de ser crítico para os negócios, o que dá origem a uma necessidade de minimizar o tempo de inatividade, preservando um caminho de back-out.

Algumas das considerações para o Planejamento de uma estratégia de migração são discutidas aqui.

Tamanho dos dados

Os tamanhos dos bancos de dados a serem migrados obviamente afetam o Planejamento da migração, embora o tamanho não afete necessariamente o tempo de transição. Quando uma grande quantidade de dados precisa ser migrada, a principal consideração é a largura de banda. As operações de cópia são geralmente realizadas com e/S sequenciais eficientes. Como uma estimativa conservadora, suponha 50% de utilização da largura de banda de rede disponível para operações de cópia. Por exemplo, uma porta FC de 8GB GB pode transferir cerca de 800MBps GB teoricamente. Assumindo 50% de utilização, um banco de dados pode ser copiado a uma taxa de cerca de 400Mbps. Portanto, um banco de dados 10TB pode ser copiado em cerca de sete horas a essa taxa.

A migração em distâncias maiores geralmente requer uma abordagem mais criativa, como o processo de envio de logs explicado em "[Movimentação online do arquivo de dados](#)". As redes IP de longa distância raramente têm largura de banda em qualquer lugar perto das velocidades LAN ou SAN. Em um caso, o NetApp ajudou com a migração de longa distância de um banco de dados 220TB com taxas de geração de arquivos log muito altas. A abordagem escolhida para transferência de dados foi o envio diário de fitas, pois esse método oferecia a máxima largura de banda possível.

Contagem da base de dados

Em muitos casos, o problema de mover uma grande quantidade de dados não é o tamanho dos dados, mas sim a complexidade da configuração que suporta o banco de dados. Simplesmente saber que 50TB de bancos de dados devem ser migrados não é informação suficiente. Pode ser um único banco de dados de missão crítica 50TB, uma coleção de 4, 000 bancos de dados legados ou uma combinação de dados de produção e não produção. Em alguns casos, grande parte dos dados consiste em clones de um banco de dados de origem. Esses clones não precisam ser migrados porque podem ser facilmente recriados, especialmente quando a nova arquitetura foi projetada para utilizar volumes NetApp FlexClone.

Para Planejar a migração, você deve entender quantos bancos de dados estão no escopo e como eles devem ser priorizados. À medida que o número de bancos de dados aumenta, a opção de migração preferida tende a ser menor e menor na pilha. Por exemplo, copiar um único banco de dados pode ser facilmente executado com RMAN e uma pequena interrupção. Essa é a replicação no nível do host.

Se houver bancos de dados 50, pode ser mais fácil evitar a configuração de uma nova estrutura de sistema de arquivos para receber uma cópia RMAN e, em vez disso, mover os dados no lugar. Esse processo pode ser feito aproveitando a migração LVM baseada em host para realocar dados de LUNs antigos para novos LUNs. Isso move a responsabilidade da equipe do administrador do banco de dados (DBA) para a equipe do sistema operacional e, como resultado, os dados são migrados de forma transparente em relação ao banco de dados.

A configuração do sistema de arquivos não foi alterada.

Por fim, se for necessário migrar bancos de dados 500 em servidores 200, opções baseadas em armazenamento, como o recurso de importação de LUN estrangeiro (FLI) do ONTAP, podem ser usadas para realizar uma migração direta dos LUNs.

Requisitos de rearquitetura

Normalmente, um layout de arquivo de banco de dados deve ser alterado para aproveitar os recursos do novo storage array; no entanto, isso nem sempre é o caso. Por exemplo, os recursos dos all-flash arrays EF-Series são direcionados principalmente à performance de SAN e à confiabilidade de SAN. Na maioria dos casos, os bancos de dados podem ser migrados para um array da EF-Series sem considerações especiais para o layout de dados. Os únicos requisitos são alto IOPS, baixa latência e confiabilidade robusta. Embora existam práticas recomendadas relacionadas a fatores como configuração RAID ou Dynamic Disk Pools, os projetos EF-Series raramente exigem alterações significativas na arquitetura de storage geral para utilizar esses recursos.

Em contraste, a migração para o ONTAP geralmente requer mais consideração do layout do banco de dados para garantir que a configuração final forneça o valor máximo. Por si só, o ONTAP oferece muitos recursos para um ambiente de banco de dados, mesmo sem nenhum esforço de arquitetura específico. O mais importante é que ele oferece a capacidade de migrar sem interrupções para um novo hardware quando o hardware atual atinge seu fim de vida útil. De um modo geral, uma migração para o ONTAP é a última migração que você precisaria executar. O hardware subsequente é atualizado e os dados são migrados para novas Mídias sem interrupções.

Com algum Planejamento, ainda mais benefícios estão disponíveis. As considerações mais importantes envolvem o uso de instantâneos. Os snapshots são a base para realizar backups, restaurações e operações de clonagem quase instantâneos. Como exemplo do poder dos snapshots, o maior uso conhecido é com um único banco de dados de 996TB GB executado em cerca de 250 LUNs em controladores de 6. Esse banco de dados pode ser feito em 2 minutos, restaurado em 2 minutos e clonado em 15 minutos. Os benefícios adicionais incluem a capacidade de mover dados pelo cluster em resposta a alterações no workload e a aplicação de controles de qualidade do serviço (QoS) para fornecer performance boa e consistente em um ambiente de vários bancos de dados.

Tecnologias como controles de QoS, realocação de dados, snapshots e clonagem funcionam em praticamente qualquer configuração. No entanto, alguns pensamentos geralmente são necessários para maximizar os benefícios. Em alguns casos, os layouts de armazenamento de banco de dados podem exigir alterações de design para maximizar o investimento no novo storage array. Essas alterações no design podem afetar a estratégia de migração porque as migrações baseadas em host ou baseadas em storage replicam o layout de dados original. Etapas adicionais podem ser necessárias para concluir a migração e fornecer um layout de dados otimizado para o ONTAP. Os procedimentos mostrados em "[Visão geral dos procedimentos de migração Oracle](#)" e mais tarde demonstram alguns dos métodos para não apenas migrar um banco de dados, mas para migrá-lo para o layout final ideal com o mínimo de esforço.

Tempo de redução

A interrupção máxima permitida de serviço durante a transição deve ser determinada. É um erro comum supor que todo o processo de migração causa interrupções. Muitas tarefas podem ser concluídas antes do início de qualquer interrupção de serviço e muitas opções permitem a conclusão da migração sem interrupção ou interrupção. Mesmo quando a interrupção é inevitável, você ainda deve definir a interrupção máxima de serviço permitida, pois a duração do tempo de transição varia de procedimento para procedimento.

Por exemplo, copiar um banco de dados 10TB normalmente requer aproximadamente sete horas para ser concluído. Se as necessidades empresariais permitirem uma interrupção de sete horas, a cópia de arquivos é uma opção fácil e segura para migração. Se cinco horas forem inaceitáveis, um processo simples de envio de

logs (consulte "[Oracle log de envio](#)") pode ser configurado com o mínimo de esforço para reduzir o tempo de transferência para aproximadamente 15 minutos. Durante esse tempo, um administrador de banco de dados pode concluir o processo. Se 15 minutos forem inaceitáveis, o processo de transição final pode ser automatizado por meio de script para reduzir o tempo de transição para apenas alguns minutos. Você sempre pode acelerar uma migração, mas isso acontece com o custo de tempo e esforço. As metas de tempo de transição devem ser baseadas no que é aceitável para o negócio.

Caminho de back-out

Nenhuma migração é completamente livre de riscos. Mesmo que a tecnologia funcione perfeitamente, há sempre a possibilidade de erro do usuário. O risco associado a um caminho de migração escolhido deve ser considerado juntamente com as consequências de uma migração falhada. Por exemplo, o recurso transparente de migração de armazenamento on-line do Oracle ASM é um de seus principais recursos, e esse método é um dos mais confiáveis conhecidos. No entanto, os dados estão sendo copiados irreversivelmente com este método. No caso altamente improvável de que um problema ocorra com ASM, não há um caminho de back-out fácil. A única opção é restaurar o ambiente original ou usar o ASM para reverter a migração de volta para os LUNs originais. O risco pode ser minimizado, mas não eliminado, executando um backup do tipo snapshot no sistema de storage original, supondo que o sistema seja capaz de executar tal operação.

Ensaio

Alguns procedimentos de migração devem ser totalmente verificados antes da execução. A necessidade de migração e ensaio do processo de transição é uma solicitação comum com bancos de dados de missão crítica para os quais a migração deve ser bem-sucedida e o tempo de inatividade deve ser minimizado. Além disso, os testes de aceitação do usuário são frequentemente incluídos como parte do trabalho de pós-migração, e o sistema geral pode ser devolvido à produção somente após a conclusão desses testes.

Se houver necessidade de ensaio, vários recursos do ONTAP podem tornar o processo muito mais fácil. Em particular, os snapshots podem redefinir um ambiente de teste e criar rapidamente várias cópias com uso eficiente de espaço de um ambiente de banco de dados.

Procedimentos

Visão geral

Muitos procedimentos estão disponíveis para o banco de dados de migração Oracle. O certo depende das necessidades do seu negócio.

Em muitos casos, os administradores de sistema e DBAs têm seus próprios métodos preferidos de realocação de dados de volume físico, espelhamento e desirritações, ou de aproveitamento do Oracle RMAN para copiar dados.

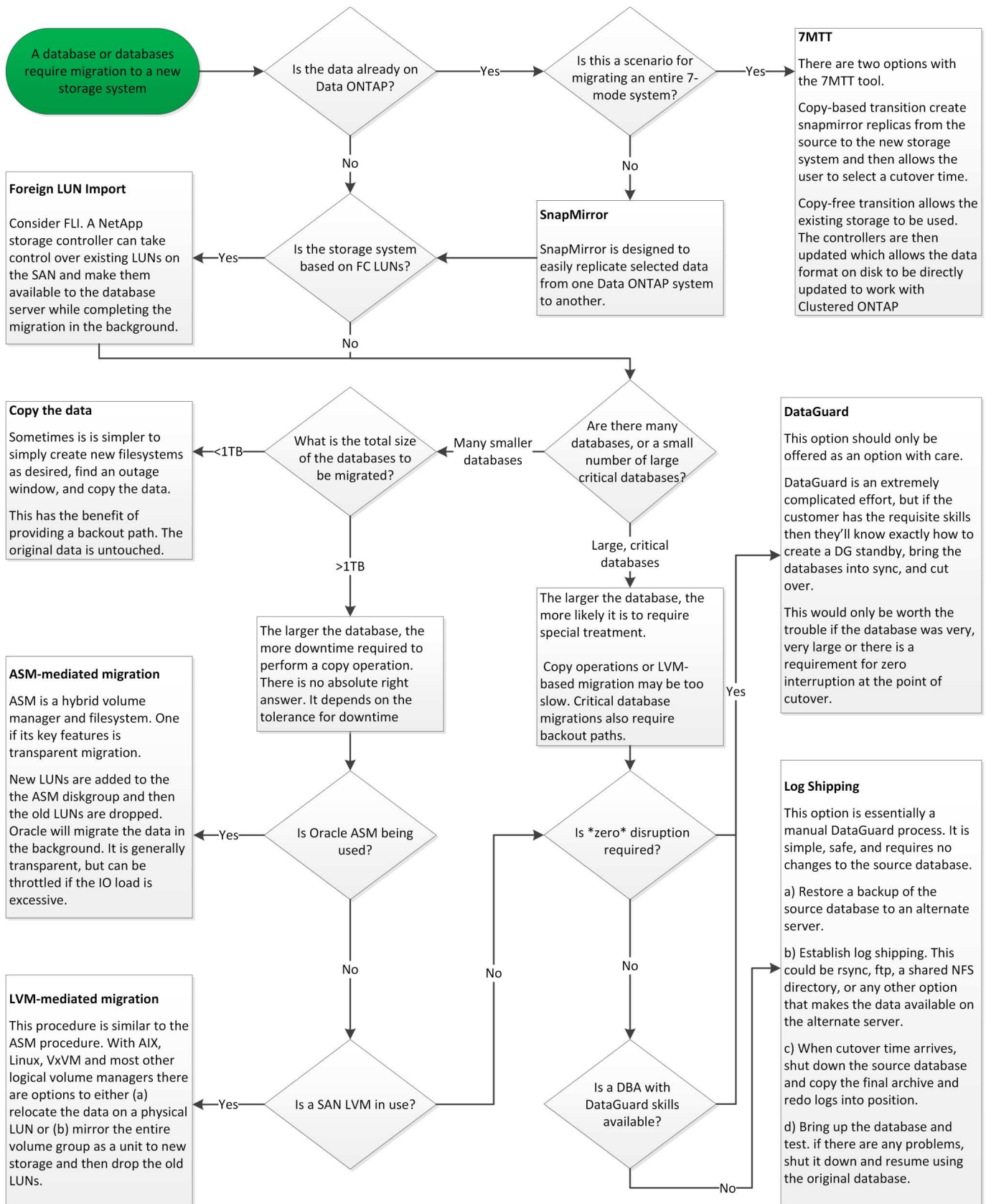
Esses procedimentos são fornecidos principalmente como orientação para a equipe DE TI menos familiarizada com algumas das opções disponíveis. Além disso, os procedimentos ilustram as tarefas, os requisitos de tempo e as demandas do conjunto de habilidades para cada abordagem de migração. Isso permite que outras partes, como NetApp e serviços profissionais de parceiros ou gerenciamento DE TI, apreciem mais plenamente os requisitos de cada procedimento.

Não existe uma prática recomendada única para a criação de uma estratégia de migração. A criação de um plano requer primeiro compreender as opções de disponibilidade e, em seguida, selecionar o método que melhor se adapta às necessidades do negócio. A figura abaixo ilustra as considerações básicas e conclusões típicas feitas pelos clientes, mas não é universalmente aplicável a todas as situações.

Por exemplo, uma etapa levanta a questão do tamanho total do banco de dados. A próxima etapa depende se

o banco de dados é mais ou menos do que 1TB. As etapas recomendadas são exatamente isso: Recomendações baseadas em práticas típicas do cliente. A maioria dos clientes não usaria o DataGuard para copiar um pequeno banco de dados, mas alguns podem. A maioria dos clientes não tentaria copiar um banco de dados 50TB devido ao tempo necessário, mas alguns podem ter uma janela de manutenção suficientemente grande para permitir tal operação.

O fluxograma abaixo mostra os tipos de considerações sobre qual caminho de migração é melhor. Você pode clicar com o botão direito na imagem e abri-la em uma nova guia para melhorar a legibilidade.



Movimentação online do arquivo de dados

Oracle 12cR1 e superior incluem a capacidade de mover um arquivo de dados enquanto o banco de dados permanece online. Ele também funciona entre diferentes tipos de sistema de arquivos. Por exemplo, um

arquivo de dados pode ser relocado de um sistema de arquivos xfs para ASM. Esse método geralmente não é usado em escala devido ao número de operações individuais de movimentação de arquivos de dados que seriam necessárias, mas é uma opção que vale a pena considerar com bancos de dados menores com menos datafiles.

Além disso, simplesmente mover um arquivo de dados é uma boa opção para migrar partes de bancos de dados existentes. Por exemplo, datafiles menos ativos podem ser relocados para um storage mais econômico, como um volume FabricPool que pode armazenar blocos ociosos no armazenamento de objetos.

Migração em nível de banco de dados

Migração no nível do banco de dados significa permitir que o banco de dados relocate os dados. Especificamente, isso significa o envio de logs. Tecnologias como RMAN e ASM são produtos Oracle, mas, para fins de migração, elas operam no nível do host, onde copiam arquivos e gerenciam volumes.

Registrar envio

A base para a migração no nível do banco de dados é o log de arquivo Oracle, que contém um log de alterações no banco de dados. Na maioria das vezes, um log de arquivamento faz parte de uma estratégia de backup e recuperação. O processo de recuperação começa com a restauração de um banco de dados e, em seguida, a reprodução de um ou mais logs de arquivo para trazer o banco de dados para o estado desejado. Essa mesma tecnologia básica pode ser usada para realizar uma migração com pouca ou nenhuma interrupção das operações. Mais importante ainda, essa tecnologia permite a migração ao mesmo tempo em que deixa o banco de dados original intocado, preservando um caminho de back-out.

O processo de migração começa com a restauração de um backup de banco de dados para um servidor secundário. Você pode fazer isso de várias maneiras, mas a maioria dos clientes usa seu aplicativo de backup normal para restaurar os arquivos de dados. Depois que os arquivos de dados são restaurados, os usuários estabelecem um método para o envio de log. O objetivo é criar um feed constante de logs de arquivo gerados pelo banco de dados principal e reproduzi-los no banco de dados restaurado para mantê-los ambos perto do mesmo estado. Quando o tempo de transição chega, o banco de dados de origem é completamente desligado e os Registros finais do arquivo e, em alguns casos, os logs de refazer são copiados e reproduzidos. É fundamental que os logs de refazer também sejam considerados porque eles podem conter algumas das transações finais confirmadas.

Depois que esses logs foram transferidos e reproduzidos, ambos os bancos de dados são consistentes uns com os outros. Neste ponto, a maioria dos clientes realiza alguns testes básicos. Se algum erro for feito durante o processo de migração, a repetição do log deve relatar erros e falhar. Ainda é aconselhável realizar alguns testes rápidos com base em consultas conhecidas ou atividades orientadas por aplicativos para verificar se a configuração é ideal. Também é uma prática comum criar uma tabela de teste final antes de encerrar o banco de dados original para verificar se ele está presente no banco de dados migrado. Esta etapa garante que não foram feitos erros durante a sincronização final do log.

Uma simples migração de envio de logs pode ser configurada fora da banda em relação ao banco de dados original, o que o torna particularmente útil para bancos de dados de missão crítica. Nenhuma alteração de configuração é necessária para o banco de dados de origem, e a restauração e configuração inicial do ambiente de migração não têm efeito sobre as operações de produção. Depois que o envio de log é configurado, ele coloca algumas demandas de e/S nos servidores de produção. No entanto, o envio de logs consiste em leituras sequenciais simples dos logs do arquivo, o que é improvável que tenha qualquer efeito no desempenho do banco de dados de produção.

O envio de logs provou ser particularmente útil para projetos de migração de longa distância e alta taxa de mudança. Em um caso, um único banco de dados 220TB foi migrado para um novo local a aproximadamente 500 km de distância. A taxa de mudança foi extremamente alta e as restrições de segurança impediram o uso de uma conexão de rede. O envio de log foi realizado usando fita adesiva e correio. Uma cópia do banco de

dados de origem foi inicialmente restaurada usando os procedimentos descritos abaixo. Os logs foram então enviados semanalmente pelo correio até o momento da transição quando o conjunto final de fitas foi entregue e os logs foram aplicados ao banco de dados de réplica.

Oracle DataGuard

Em alguns casos, um ambiente DataGuard completo é garantido. É incorreto usar o termo DataGuard para se referir a qualquer envio de log ou configuração de banco de dados em espera. O Oracle DataGuard é uma estrutura abrangente para gerenciar a replicação de banco de dados, mas não é uma tecnologia de replicação. O principal benefício de um ambiente DataGuard completo em um esforço de migração é o switchover transparente de um banco de dados para outro. O DataGuard também permite um switchover transparente de volta para o banco de dados original se um problema for descoberto, como um problema de desempenho ou conectividade de rede com o novo ambiente. Um ambiente DataGuard totalmente configurado requer a configuração não apenas da camada de banco de dados, mas também de aplicativos para que os aplicativos sejam capazes de detectar uma alteração na localização do banco de dados primário. Em geral, não é necessário usar o DataGuard para concluir uma migração, mas alguns clientes têm uma vasta experiência do DataGuard internamente e já dependem dele para o trabalho de migração.

Rearquitetura

Como discutido anteriormente, aproveitar os recursos avançados de storage arrays às vezes requer a alteração do layout do banco de dados. Além disso, uma alteração no protocolo de armazenamento, como a mudança de ASM para um sistema de arquivos NFS, necessariamente altera o layout do sistema de arquivos.

Uma das principais vantagens dos métodos de envio de log, incluindo o DataGuard, é que o destino de replicação não precisa corresponder à origem. Não há problemas com o uso de uma abordagem de envio de logs para migrar do ASM para um sistema de arquivos regular ou vice-versa. O layout preciso dos arquivos de dados pode ser alterado no destino para otimizar o uso da tecnologia PDB (Pluggable Database) ou para definir controles de QoS seletivamente em determinados arquivos. Em outras palavras, um processo de migração baseado no envio de logs permite otimizar o layout de armazenamento de banco de dados de forma fácil e segura.

Recursos do servidor

Uma limitação à migração no nível do banco de dados é a necessidade de um segundo servidor. Há duas maneiras de usar este segundo servidor:

1. Você pode usar o segundo servidor como uma nova casa permanente para o banco de dados.
2. Você pode usar o segundo servidor como um servidor de teste temporário. Depois que a migração de dados para o novo storage array for concluída e testada, os sistemas de arquivos LUN ou NFS são desconectados do servidor de teste e reconectados ao servidor original.

A primeira opção é a mais fácil, mas usá-la pode não ser viável em ambientes muito grandes que exigem servidores muito poderosos. A segunda opção requer trabalho extra para realocar os sistemas de arquivos de volta para o local original. Essa pode ser uma operação simples na qual o NFS é usado como protocolo de armazenamento, pois os sistemas de arquivos podem ser desmontados do servidor de teste e remontados no servidor original.

Os sistemas de arquivos baseados em blocos exigem trabalho extra para atualizar o zoneamento FC ou iniciadores iSCSI. Com a maioria dos gerenciadores lógicos de volume (incluindo ASM), os LUNs são detetados e colocados on-line automaticamente depois que são disponibilizados no servidor original. No entanto, algumas implementações de sistema de arquivos e LVM podem exigir mais trabalho para exportar e importar os dados. O procedimento preciso pode variar, mas geralmente é fácil estabelecer um procedimento simples e repetível para concluir a migração e realojar os dados no servidor original.

Embora seja possível configurar o envio de log e replicar um banco de dados em um único ambiente de servidor, a nova instância deve ter um SID de processo diferente para reproduzir os logs. É possível abrir temporariamente o banco de dados em um conjunto diferente de IDs de processo com um SID diferente e alterá-lo mais tarde. No entanto, isso pode levar a muitas atividades de gerenciamento complicadas, e coloca o ambiente de banco de dados em risco de erro do usuário.

Migração em nível de host

Migrar dados no nível do host significa usar o sistema operacional host e os utilitários associados para concluir a migração. Esse processo inclui qualquer utilitário que copia dados, incluindo Oracle RMAN e Oracle ASM.

Cópia de dados

O valor de uma operação de cópia simples não deve ser subestimado. Infraestruturas de rede modernas podem mover dados a taxas medidas em gigabytes por segundo, e as operações de cópia de arquivos são baseadas em e/S de leitura e gravação sequencial eficiente. Mais interrupções são inevitáveis com uma operação de cópia de host quando comparada ao envio de logs, mas uma migração é mais do que apenas a movimentação de dados. Geralmente inclui alterações na rede, no tempo de reinicialização do banco de dados e nos testes de pós-migração.

O tempo real necessário para copiar dados pode não ser significativo. Além disso, uma operação de cópia preserva um caminho de back-out garantido porque os dados originais permanecem intocados. Se algum problema for encontrado durante o processo de migração, os sistemas de arquivos originais com os dados originais podem ser reativados.

Replatforming

Replatforming refere-se a uma alteração no tipo de CPU. Quando um banco de dados é migrado de uma plataforma tradicional Solaris, AIX ou HP-UX para o Linux x86, os dados devem ser reformatados devido a alterações na arquitetura da CPU. SPARC, IA64 e CPUs DE ENERGIA são conhecidos como processadores big endian, enquanto as arquiteturas x86 e x86_64 são conhecidas como little endian. Como resultado, alguns dados dentro dos arquivos de dados Oracle são ordenados de forma diferente, dependendo do processador em uso.

Tradicionalmente, os clientes usam o DataPump para replicar dados entre plataformas. DataPump é um utilitário que cria um tipo especial de exportação de dados lógicos que pode ser mais rapidamente importado no banco de dados de destino. Como ele cria uma cópia lógica dos dados, o DataPump deixa as dependências da endianness do processador para trás. O DataPump ainda é usado por alguns clientes para replatforming, mas uma opção mais rápida se tornou disponível com o Oracle 11g: Tablespaces transportáveis entre plataformas. Este avanço permite que um espaço de tablespace seja convertido para um formato de endian diferente no lugar. Esta é uma transformação física que oferece melhor desempenho do que uma exportação DataPump, que deve converter bytes físicos em dados lógicos e depois converter de volta para bytes físicos.

Uma discussão completa sobre DataPump e tablespaces transportáveis está além da documentação do Scope NetApp, mas o NetApp tem algumas recomendações com base em nossa experiência ajudando os clientes durante a migração para um novo log de storage array com uma nova arquitetura de CPU:

- Se o DataPump estiver sendo usado, o tempo necessário para concluir a migração deve ser medido em um ambiente de teste. Às vezes, os clientes ficam surpresos no momento necessário para concluir a migração. Este tempo de inatividade adicional inesperado pode causar interrupções.
- Muitos clientes acreditam erroneamente que as tabelas transportáveis entre plataformas não exigem conversão de dados. Quando uma CPU com um endian diferente é usada, uma operação RMAN

`convert` deve ser executada nos arquivos de dados de antemão. Esta não é uma operação instantânea. Em alguns casos, o processo de conversão pode ser acelerado por ter vários threads operando em diferentes arquivos de dados, mas o processo de conversão não pode ser evitado.

Migração lógica orientada pelo Gerenciador de volumes

Os LVMs funcionam tomando um grupo de uma ou mais LUNs e dividindo-os em pequenas unidades geralmente chamadas de extensões. O conjunto de extensões é então usado como uma fonte para criar volumes lógicos que são essencialmente virtualizados. Essa camada de virtualização agrega valor de várias maneiras:

- Volumes lógicos podem usar extensões desenhadas a partir de vários LUNs. Quando um sistema de arquivos é criado em um volume lógico, ele pode usar todas as funcionalidades de performance de todos os LUNs. Ele também promove o carregamento uniforme de todos os LUNs no grupo de volumes, fornecendo performance mais previsível.
- Os volumes lógicos podem ser redimensionados adicionando e, em alguns casos, removendo extensões. Redimensionar um sistema de arquivos em um volume lógico geralmente não causa interrupções.
- Os volumes lógicos podem ser migrados sem interrupções ao mover as extensões subjacentes.

A migração usando um LVM funciona de duas maneiras: Mover uma extensão ou espelhar/desirritar uma extensão. A migração LVM usa e/S sequenciais de grandes blocos eficientes e raramente cria preocupações de desempenho. Se isso se tornar um problema, geralmente há opções para limitar a taxa de e/S. Isso aumenta o tempo necessário para concluir a migração e, ao mesmo tempo, reduz a sobrecarga de e/S nos sistemas de host e storage.

Espelho e demirror

Alguns gerenciadores de volume, como o AIX LVM, permitem que o usuário especifique o número de cópias para cada extensão e controle quais dispositivos hospedam cada cópia. A migração é feita pegando um volume lógico existente, espelhando as extensões subjacentes aos novos volumes, aguardando a sincronização das cópias e, em seguida, deixando cair a cópia antiga. Se um caminho de saída for desejado, um instantâneo dos dados originais pode ser criado antes do ponto em que a cópia espelhada é descartada. Como alternativa, o servidor pode ser encerrado brevemente para mascarar LUNs originais antes de excluir forçosamente as cópias espelhadas contidas. Isso preserva uma cópia recuperável dos dados em seu local original.

Migração de extensão

Quase todos os gerenciadores de volume permitem que extensões sejam migradas e, às vezes, existem várias opções. Por exemplo, alguns gerenciadores de volume permitem que um administrador relocate as extensões individuais para um volume lógico específico do armazenamento antigo para o novo. Os gerenciadores de volume, como o Linux LVM2, oferecem o `pvmove` comando, que realocaliza todas as extensões no dispositivo LUN especificado para um novo LUN. Depois que o LUN antigo é evacuado, ele pode ser removido.



O principal risco para as operações é a remoção de LUNs antigos e não utilizados da configuração. É preciso ter muito cuidado ao alterar o zoneamento FC e remover dispositivos LUN obsoletos.

Gerenciamento automático de armazenamento Oracle

O Oracle ASM é um gerenciador de volumes lógicos e um sistema de arquivos combinados. Em um alto nível, o Oracle ASM toma uma coleção de LUNs, os divide em pequenas unidades de alocação e os apresenta

como um único volume conhecido como um grupo de discos ASM. O ASM também inclui a capacidade de espelhar o grupo de discos definindo o nível de redundância. Um volume pode ser sem espelhamento (redundância externa), espelhado (redundância normal) ou espelhado em três vias (redundância alta). É necessário ter cuidado ao configurar o nível de redundância porque não pode ser alterado após a criação.

O ASM também fornece a funcionalidade do sistema de arquivos. Embora o sistema de arquivos não seja visível diretamente do host, o banco de dados Oracle pode criar, mover e excluir arquivos e diretórios em um grupo de discos ASM. Além disso, a estrutura pode ser navegada usando o utilitário `asmcmd`.

Assim como em outras implementações de LVM, o Oracle ASM otimiza a performance de e/S por meio da distribuição e balanceamento de carga da e/S de cada arquivo em todas as LUNs disponíveis. Em segundo lugar, as extensões subjacentes podem ser relocadas para permitir o redimensionamento do grupo de discos ASM, bem como a migração. O Oracle ASM automatiza o processo por meio da operação de rebalanceamento. Novos LUNs são adicionados a um grupo de discos ASM e os LUNs antigos são descartados, o que aciona a realocação de extensão e a subsequente queda do LUN evacuado do grupo de discos. Esse processo é um dos métodos de migração mais comprovados, e a confiabilidade do ASM na entrega de migração transparente é possivelmente sua característica mais importante.



Como o nível de espelhamento do Oracle ASM é fixo, ele não pode ser usado com o método de migração `mirror` e `Demirror`.

Migração no nível de storage

Migração no nível de storage significa realizar a migração abaixo do nível de aplicativo e do sistema operacional. No passado, isso às vezes significava usar dispositivos especializados que copiavam LUNs no nível da rede, mas esses recursos agora são encontrados nativamente no ONTAP.

SnapMirror

A migração de bancos de dados entre sistemas NetApp é quase universalmente realizada com o software de replicação de dados NetApp SnapMirror. O processo envolve a configuração de uma relação de espelho para os volumes a serem migrados, permitindo que eles sincronizem e, em seguida, aguardando a janela de transição. Quando ele chega, o banco de dados de origem é desligado, uma atualização final do espelho é executada e o espelho é quebrado. Os volumes de réplica ficam prontos para uso, seja pela montagem de um diretório de sistema de arquivos NFS contido ou descobrindo os LUNs contidos e iniciando o banco de dados.

A realocação de volumes em um único cluster do ONTAP não é considerada migração, mas sim uma operação de rotina `volume move`. O SnapMirror é usado como o mecanismo de replicação de dados no cluster. Este processo é totalmente automatizado. Não há etapas adicionais de migração a serem executadas quando os atributos do volume, como mapeamento LUN ou permissões de exportação NFS, são movidos com o próprio volume. A realocação não causa interrupções nas operações de host. Em alguns casos, o acesso à rede deve ser atualizado para garantir que os dados recém-relocados sejam acessados da maneira mais eficiente possível, mas essas tarefas também não causam interrupções.

Importação de LUN estrangeiro (FLI)

O FLI é um recurso que permite que um sistema Data ONTAP executando 8,3 ou superior migre um LUN existente de outro storage array. O procedimento é simples: O sistema ONTAP está localizado no storage array existente como se fosse qualquer outro host SAN. O Data ONTAP então assume o controle dos LUNs legados desejados e migra os dados subjacentes. Além disso, o processo de importação usa as configurações de eficiência do novo volume à medida que os dados são migrados, o que significa que os dados podem ser compactados e deduplicados em linha durante o processo de migração.

A primeira implementação do FLI no Data ONTAP 8.3 permitiu apenas migração off-line. Esta foi uma

transferência extremamente rápida, mas ainda significava que os dados LUN estavam indisponíveis até que a migração fosse concluída. A migração online foi introduzida no Data ONTAP 8.3,1. Esse tipo de migração minimiza a interrupção ao permitir que o ONTAP atenda dados LUN durante o processo de transferência. Há uma breve interrupção, enquanto o host é rezoneado para usar os LUNs por meio do ONTAP. No entanto, assim que essas alterações forem feitas, os dados serão novamente acessíveis e permanecem acessíveis durante todo o processo de migração.

A leitura de e/S é suportada através do ONTAP até que a operação de cópia esteja concluída, enquanto a escrita de e/S é escrita de forma síncrona para o LUN externo e ONTAP. As duas cópias LUN são mantidas em sincronia dessa maneira até que o administrador execute uma transição completa que libera o LUN estrangeiro e não replica mais gravações.

O FLI foi projetado para funcionar com FC, mas se houver um desejo de mudar para iSCSI, o LUN migrado pode ser facilmente remapeado como um iSCSI LUN após a conclusão da migração.

Entre as características da FLI está a detecção e ajuste automáticos do alinhamento. Neste contexto, o termo alinhamento refere-se a uma partição em um dispositivo LUN. O desempenho ideal requer que a e/S seja alinhada a 4K blocos. Se uma partição for colocada em um deslocamento que não é um múltiplo de 4K, o desempenho sofre.

Há um segundo aspecto do alinhamento que não pode ser corrigido ajustando um deslocamento de partição - o tamanho do bloco do sistema de arquivos. Por exemplo, um sistema de arquivos ZFS geralmente tem um tamanho de bloco interno de 512 bytes. Outros clientes que usam AIX criaram ocasionalmente sistemas de arquivos jfs2 com um tamanho de bloco de 512 ou 1,024 bytes. Embora o sistema de arquivos possa estar alinhado a um limite 4K, os arquivos criados dentro desse sistema de arquivos não são e o desempenho sofre.

O FLI não deve ser usado nessas circunstâncias. Embora os dados estejam acessíveis após a migração, o resultado são sistemas de arquivos com sérias limitações de desempenho. Como princípio geral, qualquer sistema de arquivos que suporte uma carga de trabalho de substituição aleatória no ONTAP deve usar um tamanho de bloco 4K. Isso é principalmente aplicável a workloads como arquivos de dados de banco de dados e implantações de VDI. O tamanho do bloco pode ser identificado usando os comandos relevantes do sistema operacional do host.

Por exemplo, no AIX, o tamanho do bloco pode ser visualizado com `lsfs -q`o .` Com Linux, ``xfs_info` e `tune2fs` pode ser usado para `xfs` e `ext3/ext4`, respectivamente. `zfs`Com ,` o comando é ``zdb -C.`

O parâmetro que controla o tamanho do bloco é e geralmente o padrão é `ashift 9`, o que significa 2⁹, ou 512 bytes. Para um desempenho ideal, o `ashift` valor deve ser 12 (2¹² 4K). Esse valor é definido no momento em que o `zpool` é criado e não pode ser alterado, o que significa que os `zpool`s de dados com `ashift` outro que não o 12 devem ser migrados copiando dados para um `zpool` recém-criado.

O Oracle ASM não tem um tamanho de bloco fundamental. O único requisito é que a partição na qual o disco ASM é construído deve estar alinhada corretamente.

Ferramenta de transição de 7 modos

A 7-Mode Transition Tool (7MTT) é um utilitário de automação usado para migrar grandes configurações de 7 modos para o ONTAP. A maioria dos clientes de bancos de dados encontra outros métodos mais fáceis, em parte porque eles geralmente migram seus ambientes de banco de dados por banco de dados, em vez de realocar todo o espaço físico do storage. Além disso, os bancos de dados costumam fazer parte apenas de um ambiente de storage maior. Portanto, os bancos de dados geralmente são migrados individualmente e, em seguida, o ambiente restante pode ser movido com 7MTT.

Há um número pequeno, mas significativo de clientes que têm sistemas de storage dedicados a ambientes de banco de dados complicados. Esses ambientes podem conter muitos volumes, snapshots e vários detalhes de configuração, como permissões de exportação, grupos de iniciadores LUN, permissões de usuário e configuração do Lightweight Directory Access Protocol. Nesses casos, as habilidades de automação do 7MTT podem simplificar uma migração.

7MTT pode operar em um de dois modos:

- **Transição baseada em cópia (CBT).** O 7MTT com CBT configura volumes SnapMirror de um sistema de 7 modos existente no novo ambiente. Depois que os dados estiverem sincronizados, o 7MTT orquestrará o processo de transição.
- **Transição livre de cópia (CFT).** O 7MTT com CFT é baseado na conversão no local de prateleiras de disco existentes de 7 modos. Nenhum dado é copiado e os compartimentos de disco existentes podem ser reutilizados. Preserva a proteção de dados e a configuração de eficiência de storage existentes.

A principal diferença entre essas duas opções é que a transição livre de cópias é uma abordagem de grande impacto na qual todos os compartimentos de disco conectados ao par de HA de 7 modos original devem ser relocados para o novo ambiente. Não há opção de mover um subconjunto de prateleiras. A abordagem baseada em cópia permite que os volumes selecionados sejam movidos. Também há potencialmente uma janela de transição mais longa com transição livre de cópias por causa do laço necessário para reciclar compartimentos de disco e converter metadados. Com base na experiência de campo, a NetApp recomenda permitir 1 hora para realocação e desativação das gavetas de disco e entre 15 minutos e 2 horas para conversão de metadados.

Migração de arquivos de dados

Datafiles Oracle individuais podem ser movidos com um único comando.

Por exemplo, o comando a seguir move o arquivo de dados IOPST.dbf do sistema de arquivos para o sistema de /oradata3 arquivos /oradata2 .

```
SQL> alter database move datafile '/oradata2/NTAP/IOPS002.dbf' to  
'/oradata3/NTAP/IOPS002.dbf';  
Database altered.
```

Mover um arquivo de dados com esse método pode ser lento, mas normalmente não deve produzir e/S suficiente para interferir com as cargas de trabalho diárias do banco de dados. Em contraste, a migração via rebalanceamento do ASM pode ser executada muito mais rápido, mas à custa de diminuir a velocidade do banco de dados geral enquanto os dados estão sendo movidos.

O tempo necessário para mover arquivos de dados pode ser facilmente medido criando um arquivo de dados de teste e, em seguida, movendo-o. O tempo decorrido para a operação é gravado nos dados da sessão:

```

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

```

Neste exemplo, o arquivo que foi movido foi o arquivo de dados 8, que era de 21GB MB de tamanho e exigiu cerca de 6 minutos para migrar. O tempo necessário obviamente depende dos recursos do sistema de storage, da rede de armazenamento e da atividade geral do banco de dados que ocorre no momento da migração.

Registrar envio

O objetivo de uma migração usando o envio de log é criar uma cópia dos arquivos de dados originais em um novo local e, em seguida, estabelecer um método de envio de alterações no novo ambiente.

Uma vez estabelecido, o envio de log e a repetição podem ser automatizados para manter o banco de dados de réplica em grande parte em sincronia com a fonte. Por exemplo, um trabalho cron pode ser programado para (a) copiar os logs mais recentes para o novo local e (b) reproduzi-los a cada 15 minutos. Isso proporciona uma interrupção mínima no momento da transição, porque não mais de 15 minutos de logs de arquivo devem ser reproduzidos.

O procedimento mostrado abaixo também é essencialmente uma operação clone do banco de dados. A lógica mostrada é semelhante ao mecanismo no NetApp SnapManager para Oracle (SMO) e no plug-in NetApp SnapCenter Oracle. Alguns clientes usaram o procedimento mostrado em scripts ou fluxos de trabalho WFA para operações de clonagem personalizadas. Embora esse procedimento seja mais manual do que o uso de SMO ou SnapCenter, ele ainda é facilmente programado e as APIs de gerenciamento de dados no ONTAP simplificam ainda mais o processo.

Envio de log - sistema de arquivos para sistema de arquivos

Este exemplo demonstra a migração de um banco de dados chamado WAFFLE de um sistema de arquivos comum para outro sistema de arquivos comum localizado em um servidor diferente. Ele também ilustra o uso do SnapMirror para fazer uma cópia rápida dos arquivos de dados, mas isso não é parte integrante do procedimento geral.

Criar cópia de segurança da base de dados

O primeiro passo é criar um backup de banco de dados. Especificamente, este procedimento requer um conjunto de arquivos de dados que podem ser usados para a repetição do log de arquivamento.

Ambiente

Neste exemplo, o banco de dados de origem está em um sistema ONTAP. O método mais simples para criar um backup de um banco de dados é usando um snapshot. O banco de dados é colocado no modo hot backup por alguns segundos enquanto uma `snapshot create` operação é executada no volume que hospeda os arquivos de dados.

```
SQL> alter database begin backup;  
Database altered.
```

```
Cluster01::*> snapshot create -vserver vserver1 -volume jfsc1_oradata  
hotbackup  
Cluster01::*>
```

```
SQL> alter database end backup;  
Database altered.
```

O resultado é um instantâneo no disco chamado `hotbackup` que contém uma imagem dos arquivos de dados enquanto no modo de backup automático. Quando combinados com os logs de arquivo apropriados para tornar os arquivos de dados consistentes, os dados nesse snapshot podem ser usados como base de uma restauração ou um clone. Neste caso, ele é replicado para o novo servidor.

Restauração de um novo ambiente

Agora, o backup precisa ser restaurado no novo ambiente. Isso pode ser feito de várias maneiras, incluindo Oracle RMAN, restauração de um aplicativo de backup como NetBackup, ou uma operação de cópia simples de arquivos de dados que foram colocados no modo de backup ativo.

Neste exemplo, o SnapMirror é usado para replicar o `hotbackup` instantâneo para um novo local.

1. Crie um novo volume para receber os dados instantâneos. Inicialize o espelhamento de `jfsc1_oradata` para `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. Depois que o estado é definido pelo SnapMirror, indicando que a sincronização está concluída, atualize o espelho com base especificamente no instantâneo desejado.

```
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. A sincronização bem-sucedida pode ser verificada visualizando o `newest-snapshot` campo no volume do espelho.

```
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. O espelho pode então ser quebrado.

```
Cluster01:::> snapmirror break -destination-path vserver1:vol_oradata
Operation succeeded: snapmirror break for destination
"vserver1:vol_oradata".
Cluster01:::>
```

5. Monte o novo sistema de arquivos. Com sistemas de arquivos baseados em blocos, os procedimentos precisos variam de acordo com a LVM em uso. O zoneamento FC ou as conexões iSCSI devem ser configuradas. Depois que a conectividade com os LUNs for estabelecida, comandos como o Linux `pvscan` podem ser necessários para descobrir quais grupos de volume ou LUNs precisam ser configurados

corretamente para serem descobertos pelo ASM.

Neste exemplo, um sistema de arquivos NFS simples é usado. Este sistema de arquivos pode ser montado diretamente.

```
fas8060-nfs1:/vol_oradata      19922944    1639360    18283584    9%  
/oradata  
fas8060-nfs1:/vol_logs        9961472     128        9961344     1%  
/logs
```

Criar modelo de criação de controlfile

Em seguida, você deve criar um modelo de controlfile. O `backup controlfile to trace` comando cria comandos de texto para recriar um controlfile. Esta função pode ser útil para restaurar um banco de dados de backup em algumas circunstâncias, e é frequentemente usada com scripts que executam tarefas como clonagem de banco de dados.

1. A saída do comando a seguir é usada para recriar os controlfiles para o banco de dados migrado.

```
SQL> alter database backup controlfile to trace as '/tmp/waffle.ctrl';  
Database altered.
```

2. Depois que os arquivos de controle tiverem sido criados, copie o arquivo para o novo servidor.

```
[oracle@jpsc3 tmp]$ scp oracle@jpsc1:/tmp/waffle.ctrl /tmp/  
oracle@jpsc1's password:  
waffle.ctrl                                100% 5199  
5.1KB/s   00:00
```

Ficheiro de parâmetros de cópia de segurança

Um arquivo de parâmetro também é necessário no novo ambiente. O método mais simples é criar um pfile a partir do spfile ou pfile atual. Neste exemplo, o banco de dados de origem está usando um spfile.

```
SQL> create pfile='/tmp/waffle.tmp.pfile' from spfile;  
File created.
```

Crie uma entrada oratab

A criação de uma entrada oratab é necessária para o bom funcionamento de utilitários como oraenv. Para criar uma entrada oratab, execute a seguinte etapa.

```
WAFFLE:/orabin/product/12.1.0/dbhome_1:N
```


Prepare a estrutura do diretório

Se os diretórios necessários ainda não estavam presentes, você deve criá-los ou o procedimento de inicialização do banco de dados falhar. Para preparar a estrutura de diretórios, preencha os seguintes requisitos mínimos.

```
[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
```

Atualizações do arquivo de parâmetros

1. Para copiar o arquivo de parâmetro para o novo servidor, execute os seguintes comandos. O local padrão é o \$ORACLE_HOME/dbs diretório. Neste caso, o pfile pode ser colocado em qualquer lugar. Ele só está sendo usado como um passo intermediário no processo de migração.

```
[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. Edite o arquivo conforme necessário. Por exemplo, se a localização do log do arquivo foi alterada, o arquivo pfile deve ser alterado para refletir o novo local. Neste exemplo, apenas os controlfiles estão sendo relocados, em parte para distribuí-los entre os sistemas de arquivos de log e dados.

```

[root@jfscl 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/control02.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. Depois que as edições estiverem concluídas, crie um spfile baseado nesse pfile.

```

SQL> create spfile from pfile='waffle.tmp.pfile';
File created.

```

Recrie controlfiles

Em uma etapa anterior, a saída do backup controlfile to trace foi copiada para o novo servidor. A parte específica da saída necessária é o controlfile recreation comando. Esta informação pode ser encontrada no ficheiro na secção marcada Set #1. NORESETLOGS. Começa com a linha create controlfile reuse database e deve incluir a palavra noresetlogs. Termina com o caractere ponto e vírgula (;).

1. Neste procedimento de exemplo, o arquivo lê o seguinte.

```
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. Edite este script como desejado para refletir a nova localização dos vários arquivos. Por exemplo, certos arquivos de dados conhecidos por oferecer suporte a e/S alta podem ser redirecionados para um sistema de arquivos em uma camada de storage de alto desempenho. Em outros casos, as alterações podem ser puramente por razões de administrador, como isolar os arquivos de dados de um determinado PDB em volumes dedicados.
3. Neste exemplo, a DATAFILE estrofe permanece inalterada, mas os logs de refazer são movidos para um novo local em /redo vez de compartilhar espaço com logs de arquivo no /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 algum arquivo estiver perdido ou os parâmetros estiverem mal configurados, são gerados erros que indicam o que deve ser corrigido. O banco de dados está montado, mas ainda não está aberto e não pode ser aberto porque os arquivos de dados em uso ainda estão marcados como estando no modo hot backup. Os logs de arquivamento devem primeiro ser aplicados para tornar o banco de dados consistente.

Replicação inicial do log

Pelo menos uma operação de resposta de log é necessária para tornar os arquivos de dados consistentes. Muitas opções estão disponíveis para reproduzir logs. Em alguns casos, o local do log do arquivo original no servidor original pode ser compartilhado por NFS e a resposta do log pode ser feita diretamente. Em outros casos, os logs do arquivo devem ser copiados.

Por exemplo, uma operação simples `scp` pode copiar todos os logs atuais do servidor de origem para o servidor de migração:

```
[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
```

Reprodução inicial do registro

Depois que os arquivos estão no local do log de arquivamento, eles podem ser reproduzidos emitindo o comando `recover database until cancel` seguido da resposta `AUTO` para reproduzir automaticamente todos os logs disponíveis.

```

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

```

A resposta final do log do arquivo relata um erro, mas isso é normal. O log indica que sqlplus estava procurando um arquivo de log específico e não o encontrou. A razão é, muito provavelmente, que o arquivo log ainda não existe.

Se o banco de dados de origem puder ser desligado antes de copiar logs de arquivo, esta etapa deve ser executada apenas uma vez. Os logs de arquivo são copiados e reproduzidos e, em seguida, o processo pode continuar diretamente para o processo de transição que replica os logs críticos de refazer.

Replicação e repetição de registros incrementais

Na maioria dos casos, a migração não é realizada imediatamente. Pode ser dias ou mesmo semanas antes que o processo de migração seja concluído, o que significa que os logs devem ser enviados continuamente para o banco de dados de réplica e reproduzidos. Portanto, quando a transição chega, os dados mínimos devem ser transferidos e reproduzidos.

Fazer isso pode ser script de várias maneiras, mas um dos métodos mais populares é usar rsync, um utilitário comum de replicação de arquivos. A maneira mais segura de usar este utilitário é configurá-lo como um daemon. Por exemplo, o rsyncd.conf arquivo a seguir mostra como criar um recurso waffle.arch chamado que é acessado com credenciais de usuário Oracle e é mapeado para '/logs/WAFFLE/arch' o . Mais importante ainda, o recurso é definido como somente leitura, o que permite que os dados de produção sejam lidos, mas não alterados.

```
[root@jfscl arch]# cat /etc/rsyncd.conf
[waffle.arch]
  uid=oracle
  gid=dba
  path=/logs/WAFFLE/arch
  read only = true
[root@jfscl arch]# rsync --daemon
```

O comando a seguir sincroniza o destino do log de arquivamento do novo servidor com o recurso `rsync waffle.arch` no servidor original. O `t` argumento em `rsync -ptg` faz com que a lista de arquivos seja comparada com base no timestamp, e apenas novos arquivos são copiados. Este processo fornece uma atualização incremental do novo servidor. Esse comando também pode ser programado no cron para ser executado regularmente.


```

[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

```

Depois que os logs tiverem sido recebidos, eles devem ser reproduzidos novamente. Exemplos anteriores mostram o uso do sqlplus para executar manualmente `recover database until cancel`, um processo que pode ser facilmente automatizado. O exemplo mostrado aqui usa o script descrito em ["Reproduzir Registros na base de dados"](#). Os scripts aceitam um argumento que especifica o banco de dados que requer uma operação de repetição. Isso permite que o mesmo script seja usado em um esforço de migração multibanco de dados.

```
[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
```

Redução

Quando você estiver pronto para cortar para o novo ambiente, você deve executar uma sincronização final que inclua Registros de arquivamento e os logs de refazer. Se a localização original do registro de refazer ainda não for conhecida, pode ser identificada da seguinte forma:

```
SQL> select member from v$logfile;
MEMBER
-----
-----
/logs/WAFFLE/redo/redo01.log
/logs/WAFFLE/redo/redo02.log
/logs/WAFFLE/redo/redo03.log
```

1. Encerre o banco de dados de origem.
2. Execute uma sincronização final dos logs de arquivo no novo servidor com o método desejado.
3. Os logs de refazer de origem devem ser copiados para o novo servidor. Neste exemplo, os logs de refazer foram relocados para um novo diretório em `/redo`.

```
[oracle@jpsc3 logs]$ scp jpsc1:/logs/WAFFLE/redo/* /redo/
oracle@jpsc1'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. Neste estágio, o novo ambiente de banco de dados contém todos os arquivos necessários para trazê-lo para o mesmo estado exato da origem. Os registros de arquivo têm de ser reproduzidos uma última vez.

```

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. Uma vez concluído, os logs de refazer devem ser reproduzidos novamente. Se a mensagem `Media recovery complete` for retornada, o processo será bem-sucedido e os bancos de dados serão sincronizados e poderão ser abertos.

```

SQL> recover database;
Media recovery complete.
SQL> alter database open;
Database altered.

```

Registo de envio - ASM para o sistema de ficheiros

Este exemplo demonstra o uso do Oracle RMAN para migrar um banco de dados. É muito semelhante ao exemplo anterior de sistema de arquivos para o envio de log do sistema de arquivos, mas os arquivos no ASM não são visíveis para o host. As únicas opções de migração de dados localizados em dispositivos ASM são a realocação do ASM LUN ou o Oracle RMAN para executar as operações de cópia.

Embora o RMAN seja um requisito para copiar arquivos do Oracle ASM, o uso do RMAN não se limita ao ASM. O RMAN pode ser usado para migrar de qualquer tipo de armazenamento para qualquer outro tipo.

Este exemplo mostra a realocação de um banco de dados chamado PANCAKE do armazenamento ASM para um sistema de arquivos regular localizado em um servidor diferente em caminhos `/oradata` e `/logs`.

Criar cópia de segurança da base de dados

O primeiro passo é criar um backup do banco de dados para ser migrado para um servidor alternativo. Como a fonte usa o Oracle ASM, o RMAN deve ser usado. Um simples backup RMAN pode ser executado da seguinte forma. Este método cria um backup marcado que pode ser facilmente identificado pelo RMAN mais tarde no procedimento.

O primeiro comando define o tipo de destino para o backup e o local a ser usado. O segundo inicia o backup dos arquivos de dados somente.

```

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/lgr6c161_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/lhr6c164_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

```

Ficheiro de controlo de cópia de segurança

Um ficheiro de controlo de cópia de segurança é necessário mais tarde no procedimento para a duplicate database operação.

```
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
```

Ficheiro de parâmetros de cópia de segurança

Um arquivo de parâmetro também é necessário no novo ambiente. O método mais simples é criar um pfile a partir do spfile ou pfile atual. Neste exemplo, o banco de dados de origem usa um spfile.

```
RMAN> create pfile='/rman/pancake/pfile' from spfile;
Statement processed
```

Script de renomeação do arquivo ASM

Vários locais de arquivo atualmente definidos nos controlfiles mudam quando o banco de dados é movido. O script a seguir cria um script RMAN para facilitar o processo. Este exemplo mostra um banco de dados com um número muito pequeno de arquivos de dados, mas normalmente os bancos de dados contêm centenas ou até mesmo milhares de arquivos de dados.

Este script pode ser encontrado em ["ASM para conversão de nome de sistema de arquivos"](#) e faz duas coisas.

Primeiro, ele cria um parâmetro para redefinir os locais de log refazer chamados `log_file_name_convert`. É essencialmente uma lista de campos alternados. O primeiro campo é a localização de um log de refazer atual e o segundo campo é a localização no novo servidor. O padrão é então repetido.

A segunda função é fornecer um modelo para renomeação de arquivos de dados. O script percorre os arquivos de dados, puxa as informações de nome e número do arquivo e formata-as como um script RMAN. Em seguida, ele faz o mesmo com os arquivos temporários. O resultado é um script rman simples que pode ser editado como desejado para garantir que os arquivos sejam restaurados para o local desejado.

```

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.

```

Capture a saída desta tela. O `log_file_name_convert` parâmetro é colocado no arquivo pfile como descrito abaixo. O nome do arquivo de dados RMAN e o script duplicado devem ser editados de acordo para colocar os arquivos de dados nos locais desejados. Neste exemplo, todos eles são colocados `/oradata/pancake` em .

```

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';
}

```

Prepare a estrutura do diretório

Os scripts estão quase prontos para serem executados, mas primeiro a estrutura de diretórios deve estar no lugar. Se os diretórios necessários ainda não estiverem presentes, eles devem ser criados ou o procedimento de inicialização do banco de dados falha. O exemplo abaixo reflete os requisitos mínimos.

```

[oracle@jpsc2 ~]$ mkdir /oradata/pancake
[oracle@jpsc2 ~]$ mkdir /logs/pancake
[oracle@jpsc2 ~]$ cd /orabin/admin
[oracle@jpsc2 admin]$ mkdir PANCAKE
[oracle@jpsc2 admin]$ cd PANCAKE
[oracle@jpsc2 PANCAKE]$ mkdir adump dpdump pfile scripts xdb_wallet

```

Crie uma entrada oratab

O comando a seguir é necessário para que utilitários como oraenv funcionem corretamente.

```
PANCAKE:/orabin/product/12.1.0/dbhome_1:N
```

Atualizações de parâmetros

O arquivo pfile salvo deve ser atualizado para refletir quaisquer alterações de caminho no novo servidor. As alterações no caminho do arquivo de dados são alteradas pelo script de duplicação RMAN, e quase todos os bancos de dados exigem alterações nos `control_files` parâmetros e `log_archive_dest`. Também pode haver locais de arquivo de auditoria que devem ser alterados e parâmetros como `db_create_file_dest` podem não ser relevantes fora do ASM. Um DBA experiente deve analisar cuidadosamente as alterações propostas antes de prosseguir.

Neste exemplo, as alterações de chave são as localizações do arquivo de controle, o destino do arquivo de log e a adição do `log_file_name_convert` parâmetro.


```

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'

```

Depois que os novos parâmetros são confirmados, os parâmetros devem ser colocados em vigor. Existem várias opções, mas a maioria dos clientes cria um spfile baseado no pfile de texto.

```
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.
```

Nomunt de arranque

A etapa final antes de replicar o banco de dados é abrir os processos do banco de dados, mas não montar os arquivos. Nesta etapa, os problemas com o spfile podem se tornar evidentes. Se o `startup nomount` comando falhar por causa de um erro de parâmetro, é simples desligar, corrigir o modelo pfile, recarregá-lo como um spfile e tentar novamente.

```
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
```

Duplique o banco de dados

Restaurar o backup RMAN anterior para o novo local consome mais tempo do que outras etapas deste processo. O banco de dados deve ser duplicado sem uma alteração no ID do banco de dados (DBID) ou redefinir os logs. Isso impede que os logs sejam aplicados, que é uma etapa necessária para sincronizar totalmente as cópias.

Conecte-se ao banco de dados com RMAN como aux e emita o comando duplicar banco de dados usando o script criado em uma etapa anterior.

```
[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

```

Replicação inicial do log

Agora você deve enviar as alterações do banco de dados de origem para um novo local. Fazer isso pode exigir uma combinação de etapas. O método mais simples seria fazer com que o RMAN no banco de dados de origem escrevesse logs de arquivo em uma conexão de rede compartilhada. Se um local compartilhado não estiver disponível, um método alternativo é usar o RMAN para gravar em um sistema de arquivos local e, em seguida, usar RCP ou rsync para copiar os arquivos.

Neste exemplo, o `/rman` diretório é um compartilhamento NFS disponível para o banco de dados original e migrado.

Uma questão importante aqui é a `disk format` cláusula. O formato de disco do backup é `%h_%e_%a.dbf`, o que significa que você deve usar o formato de número de thread, número de sequência e ID de ativação para o banco de dados. Embora as letras sejam diferentes, isso corresponde ao `log_archive_format='%t_%s_%r.dbf` parâmetro no `pfile`. Este parâmetro também especifica Registros de arquivo no formato de número de thread, número de sequência e ID de ativação. O resultado final é que os backups dos arquivos de log na fonte usam uma convenção de nomenclatura esperada pelo banco de dados. Isso torna operações como `recover database` muito mais simples porque `sqlplus` antecipa corretamente os nomes dos logs do arquivo a serem reproduzidos.

```

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

```

Reprodução inicial do registro

Depois que os arquivos estão no local do log de arquivamento, eles podem ser reproduzidos emitindo o comando `recover database until cancel` seguido da resposta `AUTO` para reproduzir automaticamente todos os logs disponíveis. O arquivo de parâmetros está direcionando Registros de arquivo para `/logs/archive`, mas isso não corresponde ao local onde o RMAN foi usado para salvar logs. O local pode ser temporariamente redirecionado da seguinte forma antes de recuperar o banco de dados.

```

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

```

A resposta final do log do arquivo relata um erro, mas isso é normal. O erro indica que sqlplus estava procurando um arquivo de log específico e não o encontrou. A razão é mais provável que o arquivo log ainda não existe.

Se o banco de dados de origem puder ser desligado antes de copiar logs de arquivo, esta etapa deve ser executada apenas uma vez. Os logs de arquivo são copiados e reproduzidos e, em seguida, o processo pode continuar diretamente para o processo de transição que replica os logs críticos de refazer.

Replicação e repetição de registos incrementais

Na maioria dos casos, a migração não é realizada imediatamente. Pode ser dias ou mesmo semanas antes que o processo de migração seja concluído, o que significa que os logs devem ser continuamente enviados para o banco de dados de réplica e reproduzidos. Isso garante que os dados mínimos devem ser transferidos e reproduzidos quando a transição chegar.

Este processo pode ser facilmente programado. Por exemplo, o comando a seguir pode ser agendado no banco de dados original para garantir que o local usado para o envio de logs seja atualizado continuamente.

```
[oracle@jfscl 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@jfscl 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.
```

Depois que os logs tiverem sido recebidos, eles devem ser reproduzidos novamente. Exemplos anteriores mostraram o uso do sqlplus para executar manualmente `recover database until cancel`, que pode ser facilmente automatizado. O exemplo mostrado aqui usa o script descrito em ["Reproduzir Registos na base de dados em espera"](#). O script aceita um argumento que especifica o banco de dados que requer uma operação de repetição. Esse processo permite que o mesmo script seja usado em um esforço de migração de multibanco de dados.

```
[root@jpsc2 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
```

Redução

Quando estiver pronto para ser cortado para o novo ambiente, você deve executar uma sincronização final. Ao trabalhar com sistemas de arquivos regulares, é fácil garantir que o banco de dados migrado seja 100% sincronizado com o original, pois os logs de refazer originais são copiados e reproduzidos. Não há uma boa maneira de fazer isso com ASM. Apenas os logs de arquivo podem ser facilmente retratados. Para se certificar de que nenhum dado é perdido, o encerramento final do banco de dados original deve ser realizado com cuidado.

1. Primeiro, o banco de dados deve ser silenciado, garantindo que nenhuma alteração esteja sendo feita. Essa quiescência pode incluir a desativação de operações agendadas, o desligamento de ouvintes e/ou o desligamento de aplicativos.
2. Depois que essa etapa é tomada, a maioria dos DBAs cria uma tabela fictícia para servir como um marcador do desligamento.
3. Forçar um arquivo de log para garantir que a criação da tabela fictícia seja gravada nos logs do arquivo. Para fazer isso, execute os seguintes comandos:

```
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. Para copiar o último dos registros de arquivo, execute os seguintes comandos. O banco de dados deve estar disponível, mas não aberto.

```
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. Para copiar os logs de arquivo, execute os seguintes comandos:

```

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. Por fim, reproduza os registros de arquivo restantes no novo servidor.

```

[root@jpsc2 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. Nesse estágio, replique todos os dados. O banco de dados está pronto para ser convertido de um banco de dados de reserva para um banco de dados operacional ativo e, em seguida, aberto.

```

SQL> alter database activate standby database;
Database altered.
SQL> alter database open;
Database altered.

```

8. Confirmar a presença da tabela fictícia e, em seguida, soltá-la.

```

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.

```

REDO log migração sem interrupções

Há momentos em que um banco de dados é organizado corretamente em geral, com exceção dos logs de refazer. Isso pode acontecer por muitos motivos, sendo que o mais comum está relacionado a instantâneos. Produtos como SnapManager para Oracle, SnapCenter e a estrutura de gerenciamento de storage NetApp Snap Creator permitem a recuperação quase instantânea de um banco de dados, mas somente se você reverter o estado dos volumes de arquivos de dados. Se os logs de refazer compartilham espaço com os arquivos de dados, a reversão não pode ser executada com segurança porque isso resultaria na destruição dos logs de refazer, provavelmente significando perda de dados. Portanto, os logs de refazer devem ser relocados.

Esse procedimento é simples e pode ser realizado sem interrupções.

Configuração atual do registo de reprocessamento

1. Identifique o número de grupos de registo de refazer e os respectivos números de grupo.

```

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. Introduza o tamanho dos registos de refazer.

```

SQL> select group#||' '||bytes from v$log;
GROUP#||' '||BYTES
-----
-----
1 524288000
2 524288000
3 524288000

```

Crie novos logs

1. Para cada log de refazer, crie um novo grupo com um tamanho e número correspondentes de membros.

```

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. Verifique a nova configuração.

```

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.

```

Soltar registros antigos

1. Solte os logs antigos (grupos 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 encontrar um erro que o impeça de largar um registro ativo, force um interruptor para o registro seguinte para libertar o bloqueio e force um ponto de verificação global. Veja o exemplo a seguir deste processo. A tentativa de soltar o grupo de arquivos de log 2, que estava localizado no local antigo, foi negada porque ainda havia dados ativos neste arquivo de log.

```

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. Um arquivo de log seguido por um ponto de verificação permite que você solte o arquivo de 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. Em seguida, elimine os registros do sistema de ficheiros. Você deve realizar este processo com extremo cuidado.

Cópia de dados do host

Assim como na migração no nível do banco de dados, a migração na camada de host fornece uma abordagem independente de fornecedor de storage.

Em outras palavras, em algum momento "basta copiar os arquivos" é a melhor opção.

Embora essa abordagem de baixa tecnologia possa parecer muito básica, ela oferece benefícios significativos porque nenhum software especial é necessário e os dados originais permanecem intactos com segurança durante o processo. A limitação principal é o fato de que uma migração de dados de cópia de arquivo é um processo disruptivo, porque o banco de dados deve ser desligado antes que a operação de cópia comece. Não há uma boa maneira de sincronizar as alterações dentro de um arquivo, então os arquivos devem ser completamente quiesced antes que a cópia comece.

Se o desligamento exigido por uma operação de cópia não for desejável, a próxima melhor opção baseada em host é aproveitar um gerenciador de volume lógico (LVM). Existem muitas opções de LVM, incluindo Oracle ASM, todas com capacidades semelhantes, mas também com algumas limitações que devem ser levadas em conta. Na maioria dos casos, a migração pode ser realizada sem tempo de inatividade e interrupção.

Filesystem para cópia do sistema de arquivos

A utilidade de uma operação de cópia simples não deve ser subestimada. Essa operação requer tempo de inatividade durante o processo de cópia, mas é um processo altamente confiável e não requer experiência especial com sistemas operacionais, bancos de dados ou sistemas de storage. Além disso, é muito seguro porque não afeta os dados originais. Normalmente, um administrador de sistema altera os sistemas de arquivos de origem para serem montados como somente leitura e, em seguida, reinicia um servidor para garantir que nada pode danificar os dados atuais. O processo de cópia pode ser programado para garantir que ele seja executado o mais rápido possível sem risco de erro do usuário. Como o tipo de e/S é uma transferência sequencial simples de dados, ele é altamente eficiente em largura de banda.

O exemplo a seguir demonstra uma opção para uma migração segura e rápida.

Ambiente

O ambiente a ser migrado é o seguinte:

- Sistemas de arquivos atuais

```

ontap-nfs1:/host1_oradata      52428800  16196928  36231872  31%
/oradata
ontap-nfs1:/host1_logs        49807360   548032   49259328  2% /logs

```

- Novos sistemas de arquivos

```

ontap-nfs1:/host1_logs_new    49807360      128  49807232  1%
/new/logs
ontap-nfs1:/host1_oradata_new 49807360      128  49807232  1%
/new/oradata

```

Visão geral

O banco de dados pode ser migrado por um DBA simplesmente desligando o banco de dados e copiando os arquivos, mas o processo é facilmente programado se muitos bancos de dados precisam ser migrados ou minimizar o tempo de inatividade é fundamental. O uso de scripts também reduz a chance de erro do usuário.

Os scripts de exemplo mostrados automatizam as seguintes operações:

- Encerrar o banco de dados
- Convertendo os sistemas de arquivos existentes em um estado somente leitura
- Copiar todos os dados da origem para os sistemas de arquivos de destino, o que preserva todas as permissões de arquivo
- Desmontar os sistemas de ficheiros antigos e novos
- Remontar os novos sistemas de arquivos nos mesmos caminhos que os sistemas de arquivos anteriores

Procedimento

1. Encerre o banco de dados.

```

[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. Converta os sistemas de arquivos para somente leitura. Isso pode ser feito mais rapidamente usando um script, como mostrado em "[Converter sistema de arquivos para somente leitura](#)".

```

[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. Confirme se os sistemas de arquivos agora são somente leitura.

```

ontap-nfs1:/host1_oradata on /oradata type nfs
(ro,bg,vers=3,rsize=65536,wsiz=65536,addr=172.20.101.10)
ontap-nfs1:/host1_logs on /logs type nfs
(ro,bg,vers=3,rsize=65536,wsiz=65536,addr=172.20.101.10)

```

4. Sincronize o conteúdo do sistema de arquivos com o `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. Desmonte os sistemas de arquivos antigos e reposicione os dados copiados. Isso pode ser feito mais rapidamente usando um script, como mostrado em ["Substitua o sistema de arquivos"](#).

```
[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. Confirme se os novos sistemas de ficheiros estão na posição correta.

```
ontap-nfs1:/host1_logs_new on /logs type nfs
(rw,bg,vers=3,rsz=65536,wsz=65536,addr=172.20.101.10)
ontap-nfs1:/host1_oradata_new on /oradata type nfs
(rw,bg,vers=3,rsz=65536,wsz=65536,addr=172.20.101.10)
```

7. Inicie o banco de dados.

```
[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
```

Redução totalmente automatizada

Este script de exemplo aceita argumentos do SID do banco de dados seguido por pares de sistemas de arquivos delimitados por comum. Para o exemplo mostrado acima, o comando é emitido da seguinte forma:

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs  
/oradata,/new/oradata
```

Quando executado, o script de exemplo tenta executar a seguinte sequência. Ele termina se encontrar um erro em qualquer etapa:

1. Encerre o banco de dados.
2. Converta os sistemas de arquivos atuais para o status somente leitura.
3. Use cada par delimitado por vírgulas de argumentos do sistema de arquivos e sincronize o primeiro sistema de arquivos para o segundo.
4. Desmonte os sistemas de ficheiros anteriores.
5. Atualize o `/etc/fstab` ficheiro da seguinte forma:
 - a. Crie uma cópia de segurança em `/etc/fstab.bak`.
 - b. Comente as entradas anteriores para os sistemas de ficheiros anteriores e novos.
 - c. Crie uma nova entrada para o novo sistema de arquivos que usa o ponto de montagem antigo.
6. Monte os sistemas de ficheiros.
7. Inicie o banco de dados.

O texto a seguir fornece um exemplo de execução para este 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
Sucesfully 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]#

```

Migração Oracle ASM spfile e passwd

Uma dificuldade em concluir a migração envolvendo ASM é o arquivo spfile específico ASM e o arquivo de senha. Por padrão, esses arquivos de metadados críticos são criados no primeiro grupo de discos ASM definido. Se um determinado grupo de discos ASM tiver de ser evacuado e removido, o ficheiro spfile e password que regem essa instância ASM deve ser realocado.

Outro caso de uso no qual esses arquivos podem precisar ser relocados é durante a implantação de software de gerenciamento de banco de dados, como o SnapManager para Oracle ou o plug-in SnapCenter Oracle. Um dos recursos desses produtos é restaurar rapidamente um banco de dados revertendo o estado dos LUNs ASM que hospedam os arquivos de dados. Fazer isso requer que o grupo de discos ASM fique offline antes de executar uma restauração. Este não é um problema, desde que os arquivos de dados de um determinado banco de dados sejam isolados em um grupo de discos ASM dedicado.

Quando esse grupo de discos também contém o arquivo ASM spfile/passwd, a única maneira que o grupo de discos pode ser colocado offline é desligar toda a instância ASM. Este é um processo disruptivo, o que significa que o arquivo spfile/passwd precisaria ser relocado.

Ambiente

1. Base de dados SID: TOAST
2. Ficheiros de dados atuais ligados +DATA
3. Ficheiros de registo e ficheiros de controlo atuais ligados +LOGS
4. Novos grupos de discos ASM estabelecidos como +NEWDATA e. +NEWLOGS

Localizações de ficheiros ASM spfile/passwd

A realocação desses arquivos pode ser feita sem interrupções. No entanto, por motivos de segurança, a NetApp recomenda desligar o ambiente de banco de dados para que você possa ter certeza de que os arquivos foram realocados e a configuração foi atualizada corretamente. Este procedimento deve ser repetido se várias instâncias ASM estiverem presentes em um servidor.

Identificar instâncias ASM

Identifique as instâncias ASM com base nos dados gravados no oratab arquivo. As instâncias ASM são denotadas por um símbolo.

```
-bash-4.1$ cat /etc/oratab | grep '^+'  
+ASM:/orabin/grid:N          # line added by Agent
```

Há uma instância ASM chamada ASM neste servidor.

Certifique-se de que todos os bancos de dados estão desligados

O único processo smon visível deve ser o Smon para a instância ASM em uso. A presença de outro processo Smon indica que um banco de dados ainda está em execução.

```
-bash-4.1$ ps -ef | grep smon  
oracle      857      1  0 18:26 ?          00:00:00 asm_smon_+ASM
```

O único processo smon é a própria instância ASM. Isso significa que nenhum outro banco de dados está sendo executado e é seguro prosseguir sem o risco de interromper as operações do banco de dados.

Localize arquivos

Identifique a localização atual do arquivo ASM spfile e senha usando os spget comandos e. pwget

```
bash-4.1$ asmcmd  
ASMCMD> spget  
+DATA/spfile.ora
```

```
ASMCMDB> pwget --asm  
+DATA/orapwasm
```

Os arquivos estão localizados na base do +DATA grupo de discos.

Copiar ficheiros

Copie os ficheiros para o novo grupo de discos ASM com os `spcopy` comandos e `pwcopy`. Se o novo grupo de discos tiver sido criado recentemente e estiver vazio, poderá ser necessário montar primeiro.

```
ASMCMDB> mount NEWDATA
```

```
ASMCMDB> spcopy +DATA/spfile.ora +NEWDATA/spfile.ora  
copying +DATA/spfile.ora -> +NEWDATA/spfilea.ora
```

```
ASMCMDB> pwcopy +DATA/orapwasm +NEWDATA/orapwasm  
copying +DATA/orapwasm -> +NEWDATA/orapwasm
```

Os ficheiros foram agora copiados de +DATA para +NEWDATA.

Atualizar instância ASM

A instância ASM agora deve ser atualizada para refletir a alteração no local. Os `spset` comandos e `pwset` atualizam os metadados ASM necessários para iniciar o grupo de discos ASM.

```
ASMCMDB> spset +NEWDATA/spfile.ora  
ASMCMDB> pwset --asm +NEWDATA/orapwasm
```

Ative ASM usando arquivos atualizados

Neste ponto, a instância ASM ainda usa os locais anteriores desses arquivos. A instância deve ser reiniciada para forçar uma releitura dos arquivos de seus novos locais e liberar bloqueios nos arquivos anteriores.

```
-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
```

Remove arquivos spfile e senhas antigos

Se o procedimento tiver sido executado com êxito, os ficheiros anteriores já não estão bloqueados e podem ser removidos.

```
-bash-4.1$ asmcmd
ASMCMD> rm +DATA/spfile.ora
ASMCMD> rm +DATA/orapwasm
```

Cópia Oracle ASM para ASM

O Oracle ASM é essencialmente um gerenciador de volumes e um sistema de arquivos combinados leves. Como o sistema de arquivos não é facilmente visível, o RMAN deve ser usado para executar operações de cópia. Embora um processo de migração baseado em cópia seja seguro e simples, isso resulta em algumas interrupções. A interrupção pode ser minimizada, mas não totalmente eliminada.

Se você quiser migração sem interrupções de um banco de dados baseado em ASM, a melhor opção é utilizar a funcionalidade do ASM para rebalancear as extensões ASM para novos LUNs e deixar cair os LUNs antigos. Isso geralmente é seguro e sem interrupções para operações, mas não oferece caminho de back-out. Se forem encontrados problemas funcionais ou de desempenho, a única opção é migrar os dados de volta para a origem.

Esse risco pode ser evitado copiando o banco de dados para o novo local em vez de mover dados, para que os dados originais fiquem intactos. O banco de dados pode ser totalmente testado em seu novo local antes de entrar em funcionamento, e o banco de dados original está disponível como uma opção de retorno se problemas forem encontrados.

Este procedimento é uma das muitas opções envolvendo RMAN. Ele foi projetado para permitir um processo de duas etapas no qual o backup inicial é criado e, em seguida, sincronizado mais tarde através da repetição de log. Esse processo é desejável para minimizar o tempo de inatividade, pois permite que o banco de dados permaneça operacional e forneça dados durante a cópia inicial da linha de base.

Copiar base de dados

O Oracle RMAN cria uma cópia de nível 0 (completa) do banco de dados de origem atualmente localizado no grupo de discos ASM +DATA para o novo local no +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

```

Forçar o interruptor de registro de arquivo

Você deve forçar um switch de log de arquivamento para garantir que os logs de arquivamento contenham todos os dados necessários para tornar a cópia totalmente consistente. Sem este comando, os dados de chave ainda podem estar presentes nos logs de refazer.

```

RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current

```

Desligue o banco de dados de origem

A interrupção começa nesta etapa porque o banco de dados é desligado e colocado em um modo de acesso limitado, somente leitura. Para encerrar o banco de dados de origem, execute os seguintes comandos:

```

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 do Controlfile

Você deve fazer backup do controlfile caso precise abortar a migração e reverter para o local de armazenamento original. Uma cópia do ficheiro de controlo de cópia de segurança não é 100% necessária, mas facilita o processo de reposição das localizações dos ficheiros de base de dados para a localização original.

```

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

```

Atualizações de parâmetros

O spfile atual contém referências aos controlfiles em seus locais atuais dentro do grupo de discos ASM antigo. Ele deve ser editado, o que é facilmente feito editando uma versão intermediária pfile.

```

RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed

```

Atualize o pfile

Atualize quaisquer parâmetros referentes a grupos de discos ASM antigos para refletir os novos nomes de grupos de discos ASM. Em seguida, salve o arquivo pfile atualizado. Certifique-se de que os db_create

parâmetros estão presentes.

No exemplo abaixo, as referências a +DATA que foram alteradas +NEWDATA são realçadas em amarelo. Dois parâmetros-chave são os db_create parâmetros que criam quaisquer novos arquivos no local correto.

```
*.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'
```

Atualize o arquivo init.ora

A maioria dos bancos de dados baseados em ASM usa um init.ora arquivo localizado no \$ORACLE_HOME/dbs diretório, que é um ponto para o spfile no grupo de discos ASM. Esse arquivo deve ser redirecionado para um local no novo grupo de discos ASM.

```
-bash-4.1$ cd $ORACLE_HOME/dbs  
-bash-4.1$ cat initTOAST.ora  
SPFILE='+DATA/TOAST/spfileTOAST.ora'
```

Altere este ficheiro da seguinte forma:

```
SPFILE='+NEWLOGS/TOAST/spfileTOAST.ora
```

Recriação do arquivo de parâmetros

O arquivo spfile agora está pronto para ser preenchido pelos dados no arquivo pfile editado.

```
RMAN> create spfile from pfile='/tmp/pfile';  
Statement processed
```

Inicie o banco de dados para começar a usar o novo spfile

Inicie o banco de dados para se certificar de que ele agora usa o arquivo spfile recém-criado e que quaisquer outras alterações aos parâmetros do sistema são registradas corretamente.

```

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

```

Restaure o ficheiro de controlo

O arquivo de controle de backup criado pelo RMAN também pode ser restaurado pelo RMAN diretamente para o local especificado no novo 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

```

Monte o banco de dados e verifique o uso do novo 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

```

Registo de reprodução

O banco de dados usa atualmente os arquivos de dados no local antigo. Antes que a cópia possa ser usada, ela deve ser sincronizada. O tempo passou durante o processo de cópia inicial e as alterações foram registradas principalmente nos logs de arquivo. Essas alterações são replicadas da seguinte forma:

1. Execute uma cópia de segurança incremental RMAN, que contém os registos de arquivo.

```
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. Repetir o registo.


```

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

```

Ativação

O arquivo de controle que foi restaurado ainda faz referência aos arquivos de dados no local original e também contém as informações de caminho para os arquivos de dados copiados.

1. Para alterar os arquivos de dados ativos, execute o `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"

```

Os arquivos de dados ativos agora são os arquivos de dados copiados, mas ainda podem haver alterações contidas nos logs de refazer finais.

2. Para reproduzir todos os logs restantes, execute o `recover database` comando. Se a mensagem `media recovery complete for` exibida, o processo foi bem-sucedido.

```

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

```

Este processo só alterou a localização dos ficheiros de dados normais. Os arquivos de dados temporários devem ser renomeados, mas não precisam ser copiados porque são apenas temporários. O banco de dados está inativo no momento, portanto não há dados ativos nos arquivos de dados temporários.

3. Para realocar os arquivos de dados temporários, primeiro identifique sua localização.

```

RMAN> select file#||' '||name from v$tempfile;
FILE#||' '||NAME
-----
-----
1 +DATA/TOAST/TEMPFILE/temp.263.897683145

```

4. Reposicione arquivos de dados temporários usando um comando RMAN que define o novo nome para cada arquivo de dados. Com o Oracle Managed Files (OMF), o nome completo não é necessário; o grupo de discos ASM é suficiente. Quando o banco de dados é aberto, o OMF vincula ao local apropriado no grupo de discos ASM. Para realocar arquivos, execute os seguintes comandos:

```

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

```

Refazer a migração de log

O processo de migração está quase concluído, mas os logs de refazer ainda estão localizados no grupo de discos ASM original. Os registos de refazer não podem ser transferidos diretamente. Em vez disso, um novo conjunto de logs de refazer é criado e adicionado à configuração, seguido de uma gota dos logs antigos.

1. Identifique o número de grupos de registo de refazer e os respetivos números de grupo.

```

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. Introduza o tamanho dos registos de refazer.

```

RMAN> select group#||' '||bytes from v$log;
GROUP#||' '||BYTES
-----
-----
1 52428800
2 52428800
3 52428800

```

3. Para cada log refazer, crie um novo grupo com uma configuração correspondente. Se você não estiver usando OMF, você deve especificar o caminho completo. Este também é um exemplo que usa os `db_create_online_log` parâmetros. Como foi mostrado anteriormente, este parâmetro foi definido para `-NEWLOGS`. Esta configuração permite que você use os seguintes comandos para criar novos logs on-line sem a necessidade de especificar um local de arquivo ou mesmo um grupo de discos ASM específico.

```

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. Abra o banco de dados.

```

SQL> alter database open;
Database altered.

```

5. Solte os logs antigos.

```

RMAN> alter database drop logfile group 1;
Statement processed

```

6. Se encontrar um erro que o impeça de largar um registo ativo, force um interruptor para o registo seguinte para libertar o bloqueio e force um ponto de verificação global. Um exemplo é mostrado abaixo. A tentativa de soltar o grupo de arquivos de log 3, que estava localizado no local antigo, foi negada porque ainda havia dados ativos neste arquivo de log. Um arquivo de log após um ponto de verificação permite que você exclua o arquivo de log.

```

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. Revise o ambiente para garantir que todos os parâmetros baseados em localização sejam atualizados.

```

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. O script a seguir demonstra como simplificar esse 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 os grupos de discos ASM foram completamente evacuados, eles agora podem ser desmontados com `asmcmd`. No entanto, em muitos casos, os arquivos pertencentes a outros bancos de dados ou o arquivo ASM `spfile/passwd` ainda podem estar presentes.

```

-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMDB> umount DATA
ASMCMDB>

```

Oracle ASM para cópia do sistema de arquivos

O procedimento de cópia do Oracle ASM para sistema de arquivos é muito semelhante ao procedimento de cópia ASM para ASM, com benefícios e restrições semelhantes. A principal diferença é a sintaxe dos vários comandos e parâmetros de configuração ao usar um sistema de arquivos visível em vez de um grupo de discos ASM.

Copiar base de dados

O Oracle RMAN é usado para criar uma cópia de nível 0 (completa) do banco de dados de origem atualmente localizado no grupo de discos ASM `+DATA` para o novo local no `/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

```

Forçar o interruptor de registo de arquivo

É necessário forçar o comutador de registo de arquivo para garantir que os registos de arquivo contêm todos os dados necessários para tornar a cópia totalmente consistente. Sem este comando, os dados de chave ainda podem estar presentes nos logs de refazer. Para forçar um switch de log de arquivo, execute o seguinte comando:

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

Desligue o banco de dados de origem

A interrupção começa nesta etapa porque o banco de dados é desligado e colocado em um modo somente leitura de acesso limitado. Para encerrar o banco de dados de origem, execute os seguintes comandos:

```
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 do Controlfile

Faça backup de arquivos de controle no caso de você precisar abortar a migração e reverter para o local de armazenamento original. Uma cópia do ficheiro de controlo de cópia de segurança não é 100% necessária, mas facilita o processo de reposição das localizações dos ficheiros de base de dados para a localização original.

```
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
```

Atualizações de parâmetros

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

Atualize o pfile

Quaisquer parâmetros referentes a grupos de discos ASM antigos devem ser atualizados e, em alguns casos, excluídos quando não forem mais relevantes. Atualize-os para refletir os novos caminhos do sistema de arquivos e salvar o arquivo pfile atualizado. Certifique-se de que o caminho de destino completo está listado. Para atualizar esses parâmetros, execute os seguintes comandos:

```
*.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'
```

Desative o arquivo init.ora original

Esse arquivo está localizado no \$ORACLE_HOME/dbs diretório e geralmente está em um arquivo pfile que serve como um ponteiro para o arquivo spfile no grupo de discos ASM. Para se certificar de que o arquivo spfile original não é mais usado, renomeie-o. Não o exclua, no entanto, porque esse arquivo é necessário se a migração tiver de ser abortada.

```
[oracle@jfscl ~]$ cd $ORACLE_HOME/dbs
[oracle@jfscl dbs]$ cat initTOAST.ora
SPFILE='+ASM0/TOAST/spfileTOAST.ora'
[oracle@jfscl dbs]$ mv initTOAST.ora initTOAST.ora.prev
[oracle@jfscl dbs]$
```


Recriação do arquivo de parâmetros

Esta é a etapa final na realocação do arquivo spfile. O arquivo spfile original não é mais usado e o banco de dados está atualmente iniciado (mas não montado) usando o arquivo intermediário. O conteúdo deste arquivo pode ser escrito para o novo local spfile da seguinte forma:

```
RMAN> create spfile from pfile='/tmp/pfile';  
Statement processed
```

Inicie o banco de dados para começar a usar o novo spfile

Você deve iniciar o banco de dados para liberar os bloqueios no arquivo intermediário e iniciar o banco de dados usando apenas o novo arquivo spfile. Iniciar o banco de dados também prova que o novo local spfile está correto e seus dados são válidos.

```
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
```

Restaure o ficheiro de controlo

Um arquivo de controle de backup foi criado no caminho /tmp/TOAST.ctrl anteriormente no procedimento. O novo spfile define os locais do controlfile como /logfs/TOAST/ctrl/ctrlfile1.ctrl e /logfs/TOAST/redo/ctrlfile2.ctrl. No entanto, esses arquivos ainda não existem.

1. Este comando restaura os dados do controlfile para os caminhos definidos no 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.ctrl  
output file name=/logs/TOAST/redo/control02.ctrl  
Finished restore at 13-MAY-16
```

2. Emita o comando mount para que os controlfiles sejam descobertos corretamente e contenham dados válidos.

```
RMAN> alter database mount;  
Statement processed  
released channel: ORA_DISK_1
```

Para validar o `control_files` parâmetro, execute o seguinte comando:

```
SQL> show parameter control_files;  
NAME                                TYPE                                VALUE  
-----                                -  
control_files                        string  
/logs/TOAST/arch/control01.ctl  
  
/logs/TOAST/redo/control02.c  
  
t1
```

Registro de reprodução

O banco de dados está usando os arquivos de dados no local antigo. Antes que a cópia possa ser usada, os arquivos de dados devem ser sincronizados. O tempo passou durante o processo de cópia inicial, e as alterações foram registradas principalmente nos Registros de arquivamento. Essas alterações são replicadas nas duas etapas a seguir.

1. Execute uma cópia de segurança incremental RMAN, que contém os registros de arquivo.

```

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. Repetir os registros.

```

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

```

Ativação

O arquivo de controle que foi restaurado ainda faz referência aos arquivos de dados no local original e também contém as informações de caminho para os arquivos de dados copiados.

1. Para alterar os arquivos de dados ativos, execute o `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. Embora os arquivos de dados devam ser totalmente consistentes, uma etapa final é necessária para reproduzir as alterações restantes registradas nos logs de refazer on-line. Use o `recover database` comando para reproduzir essas alterações e tornar a cópia 100% idêntica ao original. No entanto, a cópia ainda não está aberta.

```

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

```

Realocar arquivos de dados temporários

1. Identificar a localização dos ficheiros de dados temporários ainda em utilização no grupo de discos original.

```

RMAN> select file#||' '||name from v$tempfile;
FILE#||' '||NAME
-----
1 +ASM0/TOAST/temp01.dbf

```

2. Para realocar os arquivos de dados, execute os seguintes comandos. Se houver muitos tempfiles, use um editor de texto para criar o comando RMAN e, em seguida, corte e cole-o.

```

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

```

Refazer a migração de log

O processo de migração está quase concluído, mas os logs de refazer ainda estão localizados no grupo de discos ASM original. Os registos de refazer não podem ser transferidos diretamente. Em vez disso, um novo conjunto de logs de refazer é criado e adicionado à configuração, seguindo-se uma gota dos logs antigos.

1. Identifique o número de grupos de registo de refazer e os respetivos números de grupo.

```

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. Introduza o tamanho dos registos de refazer.

```

RMAN> select group#||' '||bytes from v$log;
GROUP#||' '||BYTES
-----
-----
1 52428800
2 52428800
3 52428800

```

3. Para cada log de refazer, crie um novo grupo usando o mesmo tamanho que o grupo de log de refazer atual usando o novo local do sistema de arquivos.

```

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. Remova os grupos de arquivos de log antigos que ainda estão localizados no armazenamento anterior.

```

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 for encontrado um erro que bloqueia a queda de um log ativo, force um switch para o próximo log para liberar o bloqueio e forçar um ponto de verificação global. Um exemplo é mostrado abaixo. A tentativa de soltar o grupo de arquivos de log 3, que estava localizado no local antigo, foi negada porque ainda havia

dados ativos neste arquivo de log. Um arquivo de log seguido por um ponto de verificação permite a exclusão de arquivos de 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. Revise o ambiente para garantir que todos os parâmetros baseados em localização sejam atualizados.

```

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. O script a seguir demonstra como tornar esse processo mais fácil.

```

[root@jfscl 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 os grupos de discos ASM foram completamente evacuados, eles agora podem ser desmontados com `asmcmd`. Em muitos casos, os arquivos pertencentes a outros bancos de dados ou o arquivo ASM `spfile/passwd` ainda podem estar presentes.

```

-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>

```

Procedimento de limpeza do ficheiro de dados

O processo de migração pode resultar em arquivos de dados com sintaxe longa ou críptica, dependendo de como o Oracle RMAN foi usado. No exemplo mostrado aqui, o backup foi realizado com o formato de arquivo `/oradata/TOAST/%U do . %U`. Indica que o RMAN deve criar um nome exclusivo padrão para cada arquivo de dados. O resultado é semelhante ao que é mostrado no texto a seguir. Os nomes tradicionais para os arquivos de dados são incorporados nos nomes. Isso pode ser limpo usando a abordagem roteirizada mostrada em "[Limpeza de migração ASM](#)".


```
[root@jfscl 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;
```

Rebalancear o Oracle ASM

Como discutido anteriormente, um grupo de discos Oracle ASM pode ser migrado de forma transparente para um novo sistema de storage usando o processo de rebalanceamento. Em resumo, o processo de rebalanceamento requer a adição de LUNs de tamanho igual ao grupo de LUNs existente, seguido de uma operação de queda do LUN anterior. O Oracle ASM realocaliza automaticamente os dados subjacentes para o novo storage em um layout ideal e, em seguida, libera os LUNs antigos quando concluído.

O processo de migração usa e/S sequenciais eficientes e geralmente não causa interrupções no desempenho, mas a taxa de migração pode ser controlada quando necessário.

Identifique os dados a serem migrados

```
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
```

Criar novos LUNs

Crie novos LUNs do mesmo tamanho e defina a associação de usuário e grupo conforme necessário. Os LUNs devem aparecer como CANDIDATE discos.

```
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
```

Adicione novos LUNS

Embora as operações de adição e exclusão possam ser executadas em conjunto, geralmente é mais fácil adicionar novos LUNs em duas etapas. Primeiro, adicione os novos LUNs ao grupo de discos. Essa etapa faz com que metade das extensões sejam migradas dos LUNs ASM atuais para os novos LUNs.

A energia de reequilíbrio indica a taxa à qual os dados estão sendo transferidos. Quanto maior o número, maior o paralelismo da transferência de dados. A migração é realizada com operações de e/S sequenciais eficientes que provavelmente causarão problemas de performance. No entanto, se desejado, o poder de reequilíbrio de uma migração contínua pode ser ajustado com o `alter diskgroup [name] rebalance power [level]` comando. Migrações típicas usam um valor de 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.
```

Monitorização da operação

Uma operação de rebalanceamento pode ser monitorada e gerenciada de várias maneiras. Usamos o seguinte comando para este exemplo.

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
-----
1 REBAL RUN
1 REBAL WAIT
```

Quando a migração estiver concluída, nenhuma operação de rebalanceamento será relatada.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

Soltar LUNs antigos

A migração está agora a meio caminho. Pode ser desejável realizar alguns testes básicos de desempenho para garantir que o ambiente esteja saudável. Após a confirmação, os dados restantes podem ser relocados deixando cair os LUNs antigos. Observe que isso não resulta no lançamento imediato dos LUNs. A operação de queda sinaliza ao Oracle ASM para realocar as extensões primeiro e, em seguida, liberar o 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.
```

Monitorização da operação

A operação de rebalanceamento pode ser monitorada e gerenciada de várias maneiras. Usamos o seguinte comando para este exemplo:

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
-----
1 REBAL RUN
1 REBAL WAIT
```

Quando a migração estiver concluída, nenhuma operação de rebalanceamento será relatada.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

Remova LUNs antigos

Antes de remover os LUNs antigos do grupo de discos, deve efetuar uma verificação final sobre o estado do cabeçalho. Depois que um LUN é liberado do ASM, ele não tem mais um nome listado e o status do cabeçalho é listado como FORMER. Isso indica que esses LUNs podem ser removidos com segurança do 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.
```

Migração para LVM

O procedimento apresentado aqui mostra os princípios de uma migração baseada em LVM de um grupo de volumes `datavg` chamado . Os exemplos são extraídos do LVM Linux, mas os princípios se aplicam igualmente a AIX, HP-UX e VxVM. Os comandos precisos podem variar.

1. Identificar os LUNs atualmente no `datavg` grupo de volumes.

```
[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. Crie novos LUNs do mesmo tamanho físico ou ligeiramente maior e defina-os como volumes físicos.

```
[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. Adicione os novos volumes ao grupo de volumes.

```
[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. Emita o `pvmove` comando para realocar as extensões de cada LUN atual para o novo LUN. O `-i [seconds]` argumento monitora o progresso da operação.

```

[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. Quando esse processo estiver concluído, solte os LUNs antigos do grupo de volumes usando o `vgreduce` comando. Se for bem-sucedido, o LUN pode agora ser removido de forma segura do sistema.

```
[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"
```

Importação LUN estrangeiro

Planejamento

Os procedimentos para migrar recursos SAN usando FLI estão documentados no NetApp ["Documentação de importação de LUN estrangeiro da ONTAP"](#) .

Do ponto de vista de um banco de dados e host, não são necessárias etapas especiais. Depois que as zonas FC forem atualizadas e os LUNs ficarem disponíveis no ONTAP, o LVM poderá ler os metadados do LVM dos LUNs. Além disso, os grupos de volume estão prontos para uso sem etapas de configuração adicionais. Em casos raros, os ambientes podem incluir arquivos de configuração que foram codificados com referências à matriz de armazenamento anterior. Por exemplo, um sistema Linux que incluía `/etc/multipath.conf` regras que referiam uma WWN de um determinado dispositivo deve ser atualizado para refletir as alterações introduzidas pela FLI.



Consulte a Matriz de compatibilidade do NetApp para obter informações sobre as configurações suportadas. Se o seu ambiente não estiver incluído, contacte o seu representante da NetApp para obter assistência.

Este exemplo mostra a migração de LUNs ASM e LVM hospedados em um servidor Linux. O FLI é suportado em outros sistemas operacionais e, embora os comandos do lado do host possam diferir, os princípios são os mesmos e os procedimentos do ONTAP são idênticos.

Identificar LUNs LVM

O primeiro passo em preparação é identificar os LUNs a serem migrados. No exemplo mostrado aqui, dois sistemas de arquivos baseados em SAN são montados `/orabin` em 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
```

O nome do grupo de volume pode ser extraído do nome do dispositivo, que usa o formato (nome do grupo de volume)-(nome do volume lógico). Neste caso, o grupo de volume é `sanvg` chamado .

O `pvdisplay` comando pode ser usado da seguinte forma para identificar os LUNs que suportam este grupo de volumes. Nesse caso, existem 10 LUNs que compõem o `sanvg` grupo de volumes.

```
[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
```

Identificar LUNs ASM

LUNs ASM também devem ser migrados. Para obter o número de LUNs e caminhos LUN do `sqlplus` como usuário do `sysasm`, execute o seguinte comando:


```

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>

```

Alterações na rede FC

O ambiente atual contém 20 LUNs a serem migrados. Atualize a SAN atual para que o ONTAP possa acessar os LUNs atuais. Os dados ainda não foram migrados, mas o ONTAP deve ler informações de configuração dos LUNs atuais para criar a nova casa para esses dados.

No mínimo, pelo menos uma porta HBA no sistema AFF/FAS deve ser configurada como uma porta do iniciador. Além disso, as zonas FC precisam ser atualizadas para que o ONTAP possa acessar os LUNs no storage array estrangeiro. Alguns storages de armazenamento têm o mascaramento LUN configurado, o que limita quais WWNs podem acessar um determinado LUN. Nesses casos, o mascaramento de LUN também deve ser atualizado para conceder acesso às WWNs do ONTAP.

Depois que esta etapa for concluída, o ONTAP deve ser capaz de visualizar a matriz de armazenamento estrangeira com o `storage array show` comando. O campo chave que retorna é o prefixo usado para identificar o LUN estranho no sistema. No exemplo abaixo, os LUNs na matriz estrangeira `FOREIGN_1` aparecem no ONTAP usando o prefixo do `FOR-1`.

Identificar matriz estrangeira

```

Cluster01::> storage array show -fields name,prefix
name          prefix
-----
FOREIGN_1     FOR-1
Cluster01::>

```

Identificar LUNs estranhos

Os LUNs podem ser listados passando o `array-name` para o `storage disk show` comando. Os dados retornados são referenciados várias vezes durante o procedimento de migração.

```

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$HuVWBW
FOR-1.4   800DT$HuVWBX
FOR-1.5   800DT$HuVWB/
FOR-1.6   800DT$HuVWBa
FOR-1.7   800DT$HuVWBd
FOR-1.8   800DT$HuVWBb
FOR-1.9   800DT$HuVWBc
FOR-1.10  800DT$HuVWBc
FOR-1.11  800DT$HuVWBf
FOR-1.12  800DT$HuVWBg
FOR-1.13  800DT$HuVWBh
FOR-1.14  800DT$HuVWBh
FOR-1.15  800DT$HuVWBj
FOR-1.16  800DT$HuVWBk
FOR-1.17  800DT$HuVWBm
FOR-1.18  800DT$HuVWBn
FOR-1.19  800DT$HuVWBn
FOR-1.20  800DT$HuVWBn
20 entries were displayed.
Cluster01::>

```

Registre LUNs de matriz estrangeira como candidatos à importação

Os LUNs estrangeiros são inicialmente classificados como qualquer tipo de LUN específico. Antes que os dados possam ser importados, os LUNs devem ser marcados como estrangeiros e, portanto, um candidato para o processo de importação. Esta etapa é concluída passando o número de série para `storage disk modify` o comando, como mostrado no exemplo a seguir. Observe que esse processo marca somente o LUN como estranho dentro do ONTAP. Nenhum dado é gravado no próprio LUN estrangeiro.

```

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$HuVWBn} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign true
Cluster01::*>

```

Criar volumes para hospedar LUNs migrados

É necessário um volume para hospedar os LUNs migrados. A configuração exata do volume depende do plano geral para aproveitar os recursos do ONTAP. Neste exemplo, os LUNs ASM são colocados em um volume e os LUNs LVM são colocados em um segundo volume. Com isso, você pode gerenciar os LUNs como grupos independentes para fins como disposição em camadas, criação de snapshots ou configuração de controles de QoS.

Defina o `snapshot-policy` para `none`. O processo de migração pode incluir uma grande quantidade de rotatividade de dados. Portanto, pode haver um grande aumento no consumo de espaço se os snapshots forem criados acidentalmente porque os dados indesejados são capturados nos snapshots.

```
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::>
```

Criar LUNs ONTAP

Após a criação dos volumes, é necessário criar os novos LUNs. Normalmente, a criação de um LUN requer que o usuário especifique tais informações como o tamanho do LUN, mas neste caso o argumento de disco externo é passado para o comando. Como resultado, o ONTAP replica os dados de configuração de LUN atuais a partir do número de série especificado. Ele também usa a geometria LUN e os dados da tabela de partição para ajustar o alinhamento LUN e estabelecer o desempenho ideal.

Nesta etapa, os números de série devem ser cruzados em relação à matriz estrangeira para garantir que o LUN estranho correto seja correspondido ao novo LUN correto.

```
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)
```

Crie relações de importação

Os LUNs agora foram criados, mas não estão configurados como um destino de replicação. Antes que essa etapa possa ser executada, os LUNs devem primeiro ser colocados off-line. Esta etapa extra foi projetada para proteger dados contra erros do usuário. Se o ONTAP permitisse que uma migração fosse executada em um LUN on-line, isso criaria o risco de que um erro tipográfico pudesse resultar na substituição de dados ativos. A etapa adicional de forçar o usuário a primeiro colocar um LUN off-line ajuda a verificar se o LUN de destino correto é usado como um destino de migração.

```
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
```

Depois que os LUNs estiverem offline, você pode estabelecer a relação de importação passando o número de série LUN estrangeiro para `lun import create` o 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::*>
```

Depois que todas as relações de importação forem estabelecidas, os LUNs podem ser colocados online novamente.

```
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::*>
```

Criar grupo de iniciadores

Um grupo de iniciadores (igroup) faz parte da arquitetura de mascaramento de LUN do ONTAP. Um LUN recém-criado não é acessível a menos que um host tenha acesso concedido pela primeira vez. Isso é feito criando um grupo que lista os nomes dos iniciadores FC WWNs ou iSCSI que devem ser concedidos acesso. Na época em que esse relatório foi escrito, a FLI era compatível apenas com LUNs FC. No entanto, a conversão para iSCSI pós-migração é uma tarefa simples, como mostrado na ["Conversão de protocolo"](#).

Neste exemplo, um grupo é criado que contém duas WWNs que correspondem às duas portas disponíveis no HBA do 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
```

Mapear novos LUNs para o host

Após a criação do grupo, os LUNs são então mapeados para o grupo definido. Esses LUNs estão disponíveis apenas para as WWNs incluídas neste grupo. O NetApp assume nesta fase do processo de migração que o host não foi zoneado para o ONTAP. Isso é importante porque, se o host for simultaneamente zoneado para o array estrangeiro e o novo sistema ONTAP, existe o risco de que LUNs com o mesmo número de série possam ser descobertos em cada array. Essa situação pode levar a falhas de multipath ou danos aos dados.

```
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::*>
```

Redução

Alguma interrupção durante uma importação LUN estrangeira é inevitável devido à necessidade de alterar a configuração da rede FC. No entanto, a interrupção não precisa durar muito mais do que o tempo necessário para reiniciar o ambiente de banco de

dados e atualizar o zoneamento FC para alternar a conectividade FC do host do LUN externo para o ONTAP.

Este processo pode ser resumido da seguinte forma:

1. Quiesce toda a atividade LUN nos LUNs externos.
2. Redirecione as conexões FC do host para o novo sistema ONTAP.
3. Acione o processo de importação.
4. Redescubra os LUNs.
5. Reinicie o banco de dados.

Não é necessário esperar que o processo de migração seja concluído. Assim que a migração para um determinado LUN começar, ele estará disponível no ONTAP e poderá servir dados enquanto o processo de cópia de dados continuar. Todas as leituras são passadas para o LUN estrangeiro, e todas as gravações são escritas de forma síncrona em ambos os arrays. A operação de cópia é muito rápida e a sobrecarga de redirecionar o tráfego FC é mínima, portanto, qualquer impacto no desempenho deve ser transitório e mínimo. Se houver problema, você pode atrasar a reinicialização do ambiente até que o processo de migração seja concluído e as relações de importação tenham sido excluídas.

Encerre a base de dados

O primeiro passo para silenciar o ambiente neste exemplo é desligar o banco de dados.

```
[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>
```

Encerre os serviços da grade

Um dos sistemas de arquivos baseados em SAN que está sendo migrado também inclui os serviços Oracle ASM. A supressão dos LUNs subjacentes requer a desmontagem dos sistemas de arquivos, o que, por sua vez, significa parar todos os processos com arquivos abertos neste sistema de arquivos.

```
[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]$
```

Desmontar sistemas de arquivos

Se todos os processos forem desligados, a operação `umount` será bem-sucedida. Se a permissão for negada, deve haver um processo com um bloqueio no sistema de arquivos. O `fuser` comando pode ajudar a identificar esses processos.

```
[root@host1 ~]# umount /orabin
[root@host1 ~]# umount /backups
```

Desativar grupos de volume

Depois de todos os sistemas de arquivos em um determinado grupo de volume serem desmontados, o grupo de volume pode ser desativado.

```
[root@host1 ~]# vgchange --activate n sanvg
  0 logical volume(s) in volume group "sanvg" now active
[root@host1 ~]#
```

Alterações na rede FC

As zonas FC agora podem ser atualizadas para remover todo o acesso do host ao array externo e estabelecer acesso ao ONTAP.

Inicie o processo de importação

Para iniciar os processos de importação de LUN, execute o `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*>

```

Monitorar o progresso da importação

A operação de importação pode ser monitorada com o `lun import show` comando. Como mostrado abaixo, a importação de todos os LUNs 20 está em andamento, o que significa que os dados agora estão acessíveis por meio do ONTAP, mesmo que a operação de cópia de dados ainda progrida.

```

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$HuVWBX /vol/new_asm/LUN1 6
vserver1   800DT$HuVWBZ /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$HuVWBh /vol/new_lvm/LUN2 4
vserver1   800DT$HuVWBh /vol/new_lvm/LUN3 3
vserver1   800DT$HuVWBj /vol/new_lvm/LUN4 3
vserver1   800DT$HuVWBk /vol/new_lvm/LUN5 3
vserver1   800DT$HuVWBk /vol/new_lvm/LUN6 4
vserver1   800DT$HuVWBm /vol/new_lvm/LUN7 3
vserver1   800DT$HuVWBn /vol/new_lvm/LUN8 2
vserver1   800DT$HuVWBn /vol/new_lvm/LUN9 2
20 entries were displayed.

```

Se você precisar de um processo off-line, atrasar a redescoberta ou a reinicialização dos serviços até que o `lun import show` comando indique que toda a migração foi bem-sucedida e concluída. Em seguida, você pode concluir o processo de migração conforme descrito em ["Importação LUN estrangeiro - conclusão"](#).

Se precisar de uma migração online, continue a redescobrir os LUNs na sua nova casa e a abrir os serviços.

Verifique se há alterações no dispositivo SCSI

Na maioria dos casos, a opção mais simples para redescobrir novos LUNs é reiniciar o host. Isso remove automaticamente dispositivos obsoletos antigos, descobre corretamente todos os novos LUNs e cria dispositivos associados, como dispositivos multipathing. O exemplo aqui mostra um processo totalmente online para fins de demonstração.

Cuidado: Antes de reiniciar um host, certifique-se de que todas as entradas `/etc/fstab` nessa referência migradas dos recursos SAN sejam comentadas. Se isso não for feito e houver problemas com o acesso LUN, o sistema operacional pode não inicializar. Esta situação não danifica os dados. No entanto, pode ser muito inconveniente inicializar no modo de recuperação ou em um modo semelhante e corrigir o `/etc/fstab` para que o sistema operacional possa ser inicializado para habilitar a solução de problemas.

Os LUNs na versão do Linux usada neste exemplo podem ser reconfigurados com o `rescan-scsi-bus.sh` comando. Se o comando for bem-sucedido, cada caminho LUN deverá aparecer na saída. A saída pode ser difícil de interpretar, mas, se o zoneamento e a configuração do igrop estavam corretos, muitos LUNs devem aparecer que incluem uma `NETAPP` string de fornecedor.

```

[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.

```

Verifique se existem dispositivos multipath

O processo de descoberta LUN também aciona a recriação de dispositivos multipath, mas o driver de multipathing Linux é conhecido por ter problemas ocasionais. A saída de `multipath -ll` deve ser verificada para verificar se a saída parece como esperado. Por exemplo, a saída abaixo mostra os dispositivos multipath associados a uma NETAPP cadeia de caracteres de fornecedor. Cada dispositivo tem quatro caminhos, com dois em uma prioridade de 50 e dois em uma prioridade de 10. Embora a saída exata possa

variar com diferentes versões do Linux, essa saída parece como esperado.



Consulte a documentação dos utilitários do host para a versão do Linux que você usa para verificar se as `/etc/multipath.conf` configurações estão corretas.

```
[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 sdbx 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
```

Reative o grupo de volumes LVM

Se os LUNs LVM tiverem sido detetados corretamente, o `vgchange --activate y` comando deverá ser bem-sucedido. Este é um bom exemplo do valor de um gerenciador de volume lógico. Uma alteração na WWN de um LUN ou mesmo de um número de série não é importante porque os metadados do grupo de volume são gravados no próprio LUN.

O sistema operacional digitalizou os LUNs e descobriu uma pequena quantidade de dados gravados no LUN que os identifica como um volume físico pertencente ao `sanvg` volumegroup. Em seguida, ele construiu todos os dispositivos necessários. Tudo o que é necessário é reativar o grupo de volume.

```
[root@host1 /]# vgchange --activate y sanvg
  Found duplicate PV fpCzdLTuKfy2xDZjailNliJh3TjLUBiT: 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
```

Remontagem dos sistemas de arquivos

Depois que o grupo de volume é reativado, os sistemas de arquivos podem ser montados com todos os dados originais intactos. Como discutido anteriormente, os sistemas de arquivos estão totalmente operacionais, mesmo que a replicação de dados ainda esteja ativa no grupo de volta.

```

[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					

Redigitalização para dispositivos ASM

Os dispositivos ASMLib devem ter sido redescobertos quando os dispositivos SCSI foram reconfigurados. A redescoberta pode ser verificada on-line reiniciando o ASMLib e, em seguida, digitalizando os discos.



Esta etapa só é relevante para configurações ASM onde ASMLib é usado.

Atenção: Onde o ASMLib não é usado, os `/dev/mapper` dispositivos devem ter sido recriados automaticamente. No entanto, as permissões podem não estar corretas. Você deve definir permissões especiais nos dispositivos subjacentes para ASM na ausência de ASMLib. Isso geralmente é feito através de entradas especiais nas `/etc/multipath.conf` regras ou `udev`, ou possivelmente em ambos os conjuntos de regras. Esses arquivos podem precisar ser atualizados para refletir alterações no ambiente em termos de WWNs ou números de série para garantir que os dispositivos ASM ainda tenham as permissões corretas.

Neste exemplo, reiniciar o ASMLib e procurar discos mostra os mesmos LUNs ASM 10 do ambiente original.

```
[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"
```

Reinicie os serviços de grade

Agora que os dispositivos LVM e ASM estão online e disponíveis, os serviços de grade podem ser reiniciados.

```
[root@host1 ~]# cd /orabin/product/12.1.0/grid/bin
[root@host1 bin]# ./crsctl start has
```

Reinicie a base de dados

Depois que os serviços de grade tiverem sido reiniciados, o banco de dados pode ser criado. Pode ser necessário esperar alguns minutos para que os serviços ASM fiquem totalmente disponíveis antes de tentar iniciar o banco de dados.

```
[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>
```

Conclusão

Do ponto de vista do host, a migração está concluída, mas a e/S ainda é atendida do array estrangeiro até que as relações de importação sejam excluídas.

Antes de excluir os relacionamentos, confirme se o processo de migração está concluído para todos os LUNs.

```

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$HuVWBX /vol/new_asm/LUN1 completed
vserver1 800DT$HuVWBZ /vol/new_asm/LUN2 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$HuVWB e /vol/new_asm/LUN9 completed
vserver1 800DT$HuVWBf /vol/new_lvm/LUN0 completed
vserver1 800DT$HuVWBg /vol/new_lvm/LUN1 completed
vserver1 800DT$HuVWBh /vol/new_lvm/LUN2 completed
vserver1 800DT$HuVWB i /vol/new_lvm/LUN3 completed
vserver1 800DT$HuVWBj /vol/new_lvm/LUN4 completed
vserver1 800DT$HuVWBk /vol/new_lvm/LUN5 completed
vserver1 800DT$HuVWB l /vol/new_lvm/LUN6 completed
vserver1 800DT$HuVWBm /vol/new_lvm/LUN7 completed
vserver1 800DT$HuVWBn /vol/new_lvm/LUN8 completed
vserver1 800DT$HuVWB o /vol/new_lvm/LUN9 completed
20 entries were displayed.

```

Eliminar relações de importação

Quando o processo de migração estiver concluído, exclua a relação de migração. Depois de fazer isso, e/S é servido exclusivamente a partir das unidades no 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

```

Anular o registo de LUNs estranhos

Finalmente, modifique o disco para remover a `is-foreign` designação.


```

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::*>

```

Conversão de protocolo

Alterar o protocolo usado para acessar um LUN é um requisito comum.

Em alguns casos, faz parte de uma estratégia geral para migrar dados para a nuvem. O TCP/IP é o protocolo da nuvem e a mudança de FC para iSCSI permite uma migração mais fácil para vários ambientes de nuvem. Em outros casos, o iSCSI pode ser desejável para aproveitar os custos reduzidos de uma SAN IP. Ocasionalmente, uma migração pode usar um protocolo diferente como medida temporária. Por exemplo, se um array estrangeiro e LUNs baseados em ONTAP não contiverem nos mesmos HBAs, você poderá usar LUNs iSCSI o suficiente para copiar dados do array antigo. Em seguida, você pode converter de volta para FC depois que os LUNs antigos forem removidos do sistema.

O procedimento a seguir demonstra a conversão de FC para iSCSI, mas os princípios gerais se aplicam a uma conversão de iSCSI para FC reversa.

Instale o iniciador iSCSI

A maioria dos sistemas operacionais inclui um iniciador iSCSI de software por padrão, mas se um não estiver incluído, ele pode ser facilmente instalado.

```

[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.e17 will be
updated
--> Processing Dependency: iscsi-initiator-utils = 6.2.0.873-32.e17 for
package: iscsi-initiator-utils-iscsiuio-6.2.0.873-32.e17.x86_64
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.e17 will be
an update
--> Running transaction check
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.e17 will
be updated
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.e17

```

```

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 o17_latest 416
k
Updating for dependencies:
  iscsi-initiator-utils-iscsiuio x86_64 6.2.0.873-32.0.2.el7 o17_latest  84
k
Transaction Summary
=====
===
Upgrade 1 Package (+1 Dependent package)
Total download size: 501 k
Downloading packages:
No Presto metadata available for o17_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.e17.x86_64
3/4
Verifying   : iscsi-initiator-utils-6.2.0.873-32.e17.x86_64
4/4
Updated:
  iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.e17
Dependency Updated:
  iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.e17
Complete!
[root@host1 /]#
```

Identificar o nome do iniciador iSCSI

Um nome exclusivo do iniciador iSCSI é gerado durante o processo de instalação. No Linux, ele está localizado no `/etc/iscsi/initiatorname.iscsi` arquivo. Esse nome é usado para identificar o host na SAN IP.

```
[root@host1 /]# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1992-05.com.redhat:497bd66ca0
```

Crie um novo grupo de iniciadores

Um grupo de iniciadores (igroup) faz parte da arquitetura de mascaramento de LUN do ONTAP. Um LUN recém-criado não é acessível a menos que um host tenha acesso concedido pela primeira vez. Esta etapa é realizada criando um grupo que lista os nomes de iniciador iSCSI ou WWNs FC que exigem acesso.

Neste exemplo, um grupo é criado que contém o iniciador iSCSI do host Linux.

```
Cluster01::*> igroup create -igroup linuxiscsi -protocol iscsi -ostype
linux -initiator iqn.1994-05.com.redhat:497bd66ca0
```

Encerrar o ambiente

Antes de alterar o protocolo LUN, os LUNs devem estar totalmente quietos. Qualquer banco de dados em um dos LUNs sendo convertidos deve ser desligado, sistemas de arquivos devem ser desmontados e grupos de volumes devem ser desativados. Quando o ASM for usado, certifique-se de que o grupo de discos ASM esteja desmontado e desligue todos os serviços de grade.

Desmapear LUNs da rede FC

Depois que os LUNs estiverem totalmente quietos, remova os mapeamentos do grupo FC original.

```
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
```

Remapear LUNs para a rede IP

Conceda acesso a cada LUN ao novo grupo de iniciadores baseado em 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::*>
```

Descubra iSCSI Targets

Existem duas fases para a descoberta iSCSI. O primeiro é descobrir os alvos, o que não é o mesmo que descobrir um LUN. O `iscsiadm` comando mostrado abaixo sonda o grupo do portal especificado pelo `-p` argument e armazena uma lista de todos os endereços IP e portas que oferecem serviços iSCSI. Neste caso, existem quatro endereços IP que têm serviços iSCSI na porta padrão 3260.



Este comando pode levar vários minutos para ser concluído se algum dos endereços IP de destino não puder ser alcançado.

```
[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
```

Descubra iSCSI LUNs

Depois que os iSCSI Targets forem descobertos, reinicie o serviço iSCSI para descobrir os iSCSI LUNs disponíveis e criar dispositivos associados, como os dispositivos multipath ou ASMLib.

```
[root@host1 ~]# service iscsi restart
Redirecting to /bin/systemctl restart iscsi.service
```

Reinicie o ambiente

Reinicie o ambiente reativando grupos de volumes, remontando sistemas de arquivos, reiniciando serviços RAC e assim por diante. Como precaução, o NetApp recomenda que você reinicie o servidor após o processo de conversão estar concluído para ter certeza de que todos os arquivos de configuração estão corretos e todos os dispositivos obsoletos são removidos.

Cuidado: Antes de reiniciar um host, certifique-se de que todas as entradas `/etc/fstab` nessa referência migradas dos recursos SAN sejam comentadas. Se esta etapa não for tomada e houver problemas com o acesso LUN, o resultado pode ser um sistema operacional que não inicializa. Este problema não danifica os dados. No entanto, pode ser muito inconveniente inicializar no modo de recuperação ou em um modo semelhante e corrigir `/etc/fstab` para que o sistema operacional possa ser inicializado para permitir que os esforços de solução de problemas comecem.

Exemplos de scripts

Os scripts apresentados são fornecidos como exemplos de como fazer scripts de várias tarefas do sistema operacional e do banco de dados. Eles são fornecidos como estão. Se for necessário suporte para um procedimento específico, entre em Contato com a NetApp ou um revendedor da NetApp.

Encerramento da base de dados

O seguinte script Perl toma um único argumento do Oracle SID e desliga um banco de dados. Ele pode ser executado como o usuário Oracle ou como 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;}

```

Inicialização do banco de dados

O seguinte script Perl toma um único argumento do Oracle SID e desliga um banco de dados. Ele pode ser executado como o usuário Oracle ou como 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;}

```

Converter sistema de arquivos para somente leitura

O script a seguir toma um argumento file-system e tenta desmontá-lo e remontá-lo como somente leitura. Isso é útil durante os processos de migração nos quais um sistema de arquivos deve ser mantido disponível para replicar dados e, no entanto, deve ser protegido contra danos acidentais.

```

#!/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;

```

Substitua o sistema de arquivos

O exemplo de script a seguir é usado para substituir um sistema de arquivos por outro. Como edita o arquivo `/etc/fstab`, ele deve ser executado como root. Ele aceita um único argumento delimitado por vírgulas dos sistemas de arquivos antigos e novos.

1. Para substituir o sistema de arquivos, execute o seguinte script:

```

#!/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";}

```



```

elseif ($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;

```

Como exemplo do uso deste script, suponha que os dados em /oradata são migrados para /neworadata e /logs são migrados para `newlogs`o . Um dos métodos mais simples para executar esta tarefa é usando uma operação de cópia de arquivo simples para realocar o novo dispositivo de volta para o ponto de montagem original.

2. Suponha que os sistemas de arquivos antigos e novos estão presentes no `/etc/fstab` arquivo da seguinte forma:

```
cluster01:/vol_oradata /oradata nfs rw,bg,vers=3,rsize=65536,wsiz=65536
0 0
cluster01:/vol_logs /logs nfs rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
cluster01:/vol_newlogs /newlogs nfs rw,bg,vers=3,rsize=65536,wsiz=65536
0 0
```

3. Quando executado, esse script desmonta o sistema de arquivos atual e o substitui pelo novo:

```
[root@jpsc3 scripts]# ./swap.fs.pl /oradata,/neworadata
/neworadata unmounted
/oradata unmounted
Mounted updated /oradata
[root@jpsc3 scripts]# ./swap.fs.pl /logs,/newlogs
/newlogs unmounted
/logs unmounted
Mounted updated /logs
```

4. O script também atualiza o `/etc/fstab` arquivo de acordo. No exemplo mostrado aqui, ele inclui as seguintes alterações:

```
#Removed by swap script cluster01:/vol_oradata /oradata nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
#Removed by swap script cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
cluster01:/vol_neworadata /oradata nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
#Removed by swap script cluster01:/vol_logs /logs nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
#Removed by swap script cluster01:/vol_newlogs /newlogs nfs
rw,bg,vers=3,rsize=65536,wsiz=65536 0 0
cluster01:/vol_newlogs /logs nfs rw,bg,vers=3,rsize=65536,wsiz=65536 0
0
```

Migração automatizada de banco de dados

Este exemplo demonstra o uso de scripts de desligamento, inicialização e substituição do sistema de arquivos para automatizar totalmente uma migração.

```

#!/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;
```

Apresentar localizações dos ficheiros

Este script coleta uma série de parâmetros críticos do banco de dados e os imprime em um formato fácil de ler. Este script pode ser útil ao revisar layouts de dados. Além disso, o script pode ser modificado para outros usos.

```
#!/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";

```

Limpeza da migração do 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";
```

ASM para conversão de nome de sistema de arquivos

```

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;
/

```


Repetir registros na base de dados

Este script aceita um único argumento de um SID Oracle para um banco de dados que está no modo de montagem e tenta reproduzir todos os logs de arquivo disponíveis atualmente.

```
#!/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;
```

Repetir registros na base de dados em espera

Esse script é idêntico ao script anterior, exceto que ele é projetado para um banco de dados em espera.

```

#! /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;

```

Notas adicionais

Otimização de desempenho e benchmarking

O teste preciso do desempenho de armazenamento de banco de dados é um assunto extremamente complicado. Requer uma compreensão dos seguintes problemas:

- IOPS e taxa de transferência
- A diferença entre as operações de e/S de primeiro plano e segundo plano
- O efeito da latência sobre o banco de dados
- Várias configurações de sistema operacional e rede que também afetam o desempenho do armazenamento

Além disso, há tarefas de bancos de dados que não são de storage a serem consideradas. Há um ponto em que a otimização do desempenho de storage não produz benefícios úteis porque a performance do storage não é mais um fator limitante para o desempenho.

A maioria dos clientes de bancos de dados agora seleciona all-flash arrays, o que cria algumas considerações adicionais. Por exemplo, considere o teste de desempenho em um sistema AFF A900 de dois nós:

- Com uma taxa de leitura/gravação de 80/20, dois nós de A900 TB podem fornecer mais de 1M IOPS de banco de dados aleatórios antes mesmo de a latência ultrapassar a marca de 150µs. Isso está muito além das demandas atuais de desempenho da maioria dos bancos de dados que é difícil prever a melhoria esperada. O storage seria, em grande parte, apagado como um gargalo.
- A largura de banda da rede é uma fonte cada vez mais comum de limitações de desempenho. Por exemplo, as soluções de disco giratório costumam ser gargalos para a performance do banco de dados, pois a latência de e/S é muito alta. Quando as limitações de latência são removidas por um array all-flash, a barreira frequentemente muda para a rede. Isso é especialmente notável com ambientes virtualizados e sistemas blade, onde a verdadeira conectividade de rede é difícil de visualizar. Isso pode complicar o teste de desempenho se o próprio sistema de armazenamento não puder ser totalmente utilizado devido a limitações de largura de banda.
- Em geral, não é possível comparar o desempenho de um array all-flash com um array que contém discos giratórios devido à latência drasticamente aprimorada dos all-flash arrays. Os resultados dos testes normalmente não são significativos.
- Comparar o desempenho máximo de IOPS com um array all-flash geralmente não é um teste útil, pois os bancos de dados não são limitados pela e/S do storage. Por exemplo, suponha que um array possa suportar 500K IOPS aleatórios, enquanto outro pode sustentar 300K. A diferença é irrelevante no mundo real se um banco de dados está gastando 99% do seu tempo em processamento de CPU. Os workloads nunca utilizam todas as funcionalidades do storage array. Em contraste, os recursos de IOPS de pico podem ser críticos em uma plataforma de consolidação na qual o storage array deve ser carregado para seus recursos de pico.
- Considere sempre a latência e o IOPS em qualquer teste de storage. Muitos storage arrays no mercado reivindicam níveis extremos de IOPS, mas a latência torna esses IOPS inúteis em tais níveis. O destino típico com all-flash arrays é a marca 1ms. Uma abordagem melhor para testar não é medir o máximo de IOPS possível, mas determinar quantos IOPS um storage array pode sustentar antes que a latência média seja maior que 1ms.

Oracle Automatic Workload Repository e benchmarking

O padrão-ouro para comparações de desempenho Oracle é um relatório do Oracle Automatic Workload Repository (AWR).

Existem vários tipos de relatórios AWR. Do ponto de vista de armazenamento, um relatório gerado pela execução do `awrrpt.sql` comando é o mais abrangente e valioso, porque tem como alvo uma instância de banco de dados específica e inclui alguns histogramas detalhados que quebram eventos de e/S de armazenamento com base na latência.

Comparar dois arrays de desempenho idealmente envolve executar a mesma carga de trabalho em cada array e produzir um relatório AWR que segmente precisamente a carga de trabalho. No caso de uma carga de trabalho muito longa, um único relatório de AWR com um tempo decorrido que engloba o tempo de início e de parada pode ser usado, mas é preferível dividir os dados de AWR como vários relatórios. Por exemplo, se um trabalho em lote foi executado da meia-noite às 6 da manhã, crie uma série de relatórios AWR de uma hora das meia-noite às 1 da manhã, das 1 às 2 da manhã, e assim por diante.

Em outros casos, uma consulta muito curta deve ser otimizada. A melhor opção é um relatório AWR baseado em um instantâneo AWR criado quando a consulta começa e um segundo instantâneo AWR criado quando a consulta termina. O servidor de banco de dados deve ser silencioso para minimizar a atividade em segundo plano que obscureceria a atividade da consulta em análise.



Onde os relatórios AWR não estão disponíveis, os relatórios do Oracle statspack são uma boa alternativa. Eles contêm a maioria das mesmas estatísticas de e/S que um relatório AWR.

Oracle AWR e solução de problemas

Um relatório AWR também é a ferramenta mais importante para analisar um problema de desempenho.

Assim como no benchmarking, a solução de problemas de desempenho exige que você meça com precisão uma carga de trabalho específica. Quando possível, forneça dados de AWR ao relatar um problema de desempenho ao centro de suporte da NetApp ou ao trabalhar com um NetApp ou equipe de conta de parceiro sobre uma nova solução.

Ao fornecer dados AWR, considere os seguintes requisitos:

- Execute o `awrrpt.sql` comando para gerar o relatório. A saída pode ser texto ou HTML.
- Se os Oracle Real Application clusters (RACs) forem usados, gere relatórios AWR para cada instância no cluster.
- Segmente a hora específica em que o problema existia. O tempo decorrido máximo aceitável de um relatório AWR é geralmente de uma hora. Se um problema persistir por várias horas ou envolver uma operação de várias horas, como um trabalho em lote, forneça vários relatórios AWR de uma hora que cobrem todo o período a ser analisado.
- Se possível, ajuste o intervalo de instantâneos AWR para 15 minutos. Esta definição permite efetuar uma análise mais detalhada. Isso também requer execuções adicionais de `awrrpt.sql` para fornecer um relatório para cada intervalo de 15 minutos.
- Se o problema for uma consulta em execução muito curta, forneça um relatório AWR com base em um instantâneo AWR criado quando a operação começar e um segundo instantâneo AWR criado quando a operação terminar. O servidor de banco de dados deve ser silencioso para minimizar a atividade em segundo plano que obscureceria a atividade da operação em análise.
- Se um problema de desempenho for relatado em determinados momentos, mas não em outros, forneça dados AWR adicionais que demonstrem bom desempenho para comparação.

calibrar_io

O `calibrate_io` comando nunca deve ser usado para testar, comparar ou comparar sistemas de storage. Conforme indicado na documentação da Oracle, este procedimento calibra os recursos de e/S do armazenamento.

A calibração não é a mesma que o benchmarking. O objetivo deste comando é emitir e/S para ajudar a calibrar as operações do banco de dados e melhorar sua eficiência otimizando o nível de e/S emitido para o host. Como o tipo de I/O realizado pela `calibrate_io` operação não representa a e/S real do usuário do banco de dados, os resultados não são previsíveis e frequentemente nem reproduzíveis.

SLOB2

SLOB2, o silly Little Oracle Benchmark, tornou-se a ferramenta preferida para avaliar o desempenho do banco de dados. Foi desenvolvido por Kevin Closson e está disponível em "<https://kevinclosson.net/slob/>". Leva minutos para instalar e configurar, e ele usa um banco de dados Oracle real para gerar padrões de e/S em um espaço de tabela definido pelo usuário. É uma das poucas opções de teste disponíveis que pode saturar um array all-flash com e/S. Também é útil para gerar níveis muito mais baixos de e/S para simular workloads de storage com IOPS baixo, mas sensíveis à latência.

Banco oscilante

O Swingbench pode ser útil para testar o desempenho do banco de dados, mas é extremamente difícil usar o Swingbench de uma forma que estressa o armazenamento. NetApp não viu nenhum teste do Swingbench que rendeu e/S suficiente para ser uma carga significativa em qualquer array AFF. Em casos limitados, o teste de entrada de pedidos (OET) pode ser usado para avaliar o armazenamento de um ponto de vista de latência. Isso pode ser útil em situações em que um banco de dados tem uma dependência de latência conhecida para consultas específicas. É preciso ter cuidado para garantir que o host e a rede estejam configurados adequadamente para realizar os potenciais de latência de um array all-flash.

HammerDB

HammerDB é uma ferramenta de teste de banco de dados que simula benchmarks TPC-C e TPC-H, entre outros. Pode levar muito tempo para construir um conjunto de dados suficientemente grande para executar corretamente um teste, mas pode ser uma ferramenta eficaz para avaliar o desempenho para aplicativos OLTP e data warehouse.

Orion

A ferramenta Oracle Orion foi comumente usada com o Oracle 9, mas não foi mantida para garantir a compatibilidade com alterações em vários sistemas operacionais de host. Raramente é usado com Oracle 10i ou Oracle 11i devido a incompatibilidades com o SO e configuração de armazenamento.

A Oracle reescreveu a ferramenta, e ela é instalada por padrão com o Oracle 12c. Embora este produto tenha sido melhorado e use muitas das mesmas chamadas que um banco de dados Oracle real usa, ele não usa exatamente o mesmo caminho de código ou comportamento de e/S usado pela Oracle. Por exemplo, a maioria das e/S Oracle são executadas de forma síncrona, o que significa que o banco de dados pára até que a e/S esteja concluída à medida que a operação de e/S for concluída em primeiro plano. Simplesmente inundar um sistema de armazenamento com e/S aleatórias não é uma reprodução de e/S Oracle real e não oferece um método direto de comparar matrizes de armazenamento ou medir o efeito das alterações de configuração.

Dito isto, existem alguns casos de uso para Orion, como a medição geral do desempenho máximo possível de uma configuração particular de armazenamento de rede-host, ou para medir a integridade de um sistema de armazenamento. Com testes cuidadosos, testes Orion utilizáveis podem ser desenvolvidos para comparar matrizes de armazenamento ou avaliar o efeito de uma alteração de configuração, desde que os parâmetros incluam consideração de IOPS, taxa de transferência e latência e tentativa de replicar fielmente uma carga de trabalho realista.

Stale NFSv3 fechaduras

Se um servidor de banco de dados Oracle falhar, ele pode ter problemas com bloqueios NFS obsoletos ao reiniciar. Este problema é evitável prestando atenção cuidadosa à configuração da resolução de nomes no servidor.

Este problema surge porque criar um bloqueio e limpar um bloqueio usam dois métodos ligeiramente diferentes de resolução de nomes. Dois processos estão envolvidos, o Network Lock Manager (NLM) e o cliente NFS. O NLM usa `uname -n` para determinar o nome do host, enquanto o `rpc.statd` processo usa `gethostbyname()`. Esses nomes de host devem corresponder para que o sistema operacional limpe corretamente os bloqueios obsoletos. Por exemplo, o host pode estar procurando bloqueios de propriedade ``dbserver5 do`, mas os bloqueios foram registrados pelo host como `dbserver5.mydomain.org`. Se `gethostbyname()` não retornar o mesmo valor que `uname -a`, o processo de liberação do bloqueio não foi bem-sucedido.

O script de exemplo a seguir verifica se a resolução do nome é totalmente consistente:

```
#!/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` não corresponder `uname`, é provável que os bloqueios obsoletos. Por exemplo, este resultado revela um problema potencial:

```
uname -n yields: dbserver5
gethostbyname yields: dbserver5.mydomain.org
```

A solução é geralmente encontrada alterando a ordem em que os hosts aparecem em `/etc/hosts`. por exemplo, suponha que o arquivo `hosts` inclua esta entrada:

```
10.156.110.201 dbserver5.mydomain.org dbserver5 loghost
```

Para resolver esse problema, altere a ordem em que o nome de domínio totalmente qualificado e o nome de host curto aparecem:

```
10.156.110.201 dbserver5 dbserver5.mydomain.org loghost
```

`gethostbyname()` agora retorna o nome de host curto `dbserver5`, que corresponde à saída `uname` do . Assim, os bloqueios são apagados automaticamente após uma falha do servidor.

Verificação do alinhamento do WAFL

O alinhamento correto do WAFL é essencial para um bom desempenho. Embora o ONTAP gere blocos em 4KB unidades, esse fato não significa que o ONTAP realize todas as operações em 4KB unidades. Na verdade, o ONTAP suporta operações de blocos de diferentes tamanhos, mas a contabilidade subjacente é gerenciada pela WAFL em 4KB unidades.

O termo "alinhamento" refere-se a como Oracle I/O corresponde a essas 4KB unidades. O desempenho ideal requer um bloco Oracle 8KBi para residir em dois blocos físicos de 4KB WAFL em uma unidade. Se um bloco é compensado por 2KB, este bloco reside em metade de um bloco 4KB, um bloco 4KB completo separado e, em seguida, metade de um terceiro bloco 4KB. Este arranjo causa degradação do desempenho.

O alinhamento não é um problema com sistemas de arquivos nas. Os arquivos de dados Oracle estão alinhados ao início do arquivo com base no tamanho do bloco Oracle. Portanto, os tamanhos de bloco de 8KB, 16KB e 32KB estão sempre alinhados. Todas as operações de bloco são compensadas desde o início do arquivo em unidades de 4 kilobytes.

Os LUNs, em contraste, geralmente contêm algum tipo de cabeçalho de driver ou metadados do sistema de arquivos no início que cria um deslocamento. O alinhamento raramente é um problema em sistemas operacionais modernos porque esses sistemas operacionais são projetados para unidades físicas que podem usar um setor 4KB nativo, que também requer que e/S sejam alinhados aos limites 4KB para um desempenho ideal.

Há, no entanto, algumas exceções. Um banco de dados pode ter sido migrado de um sistema operacional antigo que não foi otimizado para e/S 4KB, ou erro de usuário durante a criação da partição pode ter levado a um deslocamento que não está em unidades de tamanho 4KB.

Os exemplos a seguir são específicos do Linux, mas o procedimento pode ser adaptado para qualquer sistema operacional.

Alinhado

O exemplo a seguir mostra uma verificação de alinhamento em um único LUN com uma única partição.

Primeiro, crie a partição que usa todas as partições disponíveis na unidade.

```
[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]#
```

O alinhamento pode ser verificado matematicamente com o seguinte 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
```

A saída mostra que as unidades são 512 bytes e o início da partição é 32 unidades. Este é um total de 32 x 512 de 16.834 bytes, que é um múltiplo inteiro de 4KB blocos WAFL. Esta partição está corretamente alinhada.

Para verificar o alinhamento correto, execute as seguintes etapas:

1. Identifique o identificador universal único (UUID) do LUN.

```
FAS8040SAP::> lun show -v /vol/jfs_luns/lun0
      Vserver Name: jfs
      LUN UUID: ed95d953-1560-4f74-9006-85b352f58fcd
      Mapped: mapped`
```

2. Insira o shell do nó no controlador 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. Inicie coleções estatísticas no UUID alvo identificado no primeiro passo.

```
FAS8040SAP-02*> stats start lun:ed95d953-1560-4f74-9006-85b352f58fcd
Stats identifier name is 'Ind0xffffffff08b9536188'
FAS8040SAP-02*>
```

4. Execute algumas I/O.. É importante usar o `iflag` argumento para garantir que a e/S seja síncrona e não armazenada em buffer.



Tenha muito cuidado com este comando. Reverter os `if` argumentos e `of` destrói os dados.

```
[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. Pare as estatísticas e veja o histograma de alinhamento. Toda a e/S deve estar no `.0` balde, o que indica e/S que está alinhada a um limite de bloco 4KB.

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffffff08b9536188
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%
```

Desalinhado

O exemplo a seguir mostra e/S desalinhadas:

1. Crie uma partição que não esteja alinhada a um limite 4KB. Este não é um comportamento padrão em sistemas operacionais modernos.

```

[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. A partição foi criada com um deslocamento de 33 setores em vez do padrão 32. Repita o procedimento descrito em "Alinhado". O histograma aparece da seguinte forma:

```

FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffffff0468242e78
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%

```

O desalinhamento é claro. A e/S cai principalmente no bucket**.1, que corresponde ao deslocamento esperado. Quando a partição foi criada, ela foi movida 512 bytes mais para o dispositivo do que o padrão otimizado, o que significa que o histograma é deslocado em 512 bytes.

Além disso, a `read_partial_blocks` estatística é diferente de zero, o que significa que I/O foi realizado que não preencheu um bloco 4KB inteiro.

Refazer o registro

Os procedimentos aqui explicados são aplicáveis aos datafiles. Os logs do Oracle refazer e os logs de arquivamento têm padrões de e/S diferentes. Por exemplo, refazer o Registro é uma substituição circular de um único arquivo. Se o tamanho padrão de bloco de 512 bytes for usado, as estatísticas de gravação são semelhantes a isso:

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffffff0468242e78
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%
```

A e/S seria distribuída em todos os intervalos de histograma, mas isso não é uma preocupação de desempenho. No entanto, as taxas de refazer o log extremamente altas podem se beneficiar do uso de um tamanho de bloco de 4KBMB. Nesse caso, é desejável garantir que os LUNs de refazer o log estejam alinhados corretamente. No entanto, isso não é tão crítico para um bom desempenho como o alinhamento de arquivos de dados.

PostgreSQL

Visão geral

PostgreSQL vem com variantes que incluem PostgreSQL, PostgreSQL Plus e EDB Postgres Advanced Server (EPAS). O PostgreSQL é normalmente implantado como banco de dados back-end para aplicativos de várias camadas. Ele é suportado por pacotes de middleware comuns (como PHP, Java, Python, Tcl/Tk, ODBC e JDBC) e tem sido historicamente uma escolha popular para sistemas de gerenciamento de banco de dados de código aberto. O ONTAP é uma excelente escolha para executar bancos de dados PostgreSQL de acordo com sua confiabilidade, alto desempenho e recursos eficientes de gerenciamento de dados.



Esta documentação sobre o ONTAP e o banco de dados PostgreSQL substitui o banco de dados *TR-4770: PostgreSQL publicado anteriormente sobre as melhores práticas do ONTAP*.

À medida que os dados crescem exponencialmente, o gerenciamento de dados se torna mais complexo para as empresas. Essa complexidade aumenta os custos de licenciamento, operação, suporte e manutenção. Para reduzir o TCO geral, considere mudar de bancos de dados comerciais para de código aberto com armazenamento de back-end confiável e de alto desempenho.

O ONTAP é uma plataforma ideal porque o ONTAP é literalmente projetado para bancos de dados. Vários recursos, como otimizações aleatórias de latência de e/S para qualidade avançada de serviço (QoS) e funcionalidade básica do FlexClone, foram criados especificamente para atender às necessidades dos workloads de banco de dados.

Recursos adicionais, como atualizações sem interrupções, (incluindo substituição de storage) garantem que seus bancos de dados essenciais permaneçam disponíveis. Você também pode ter recuperação instantânea de desastres para ambientes grandes por meio do MetroCluster ou selecionar bancos de dados usando o SnapMirror active Sync.

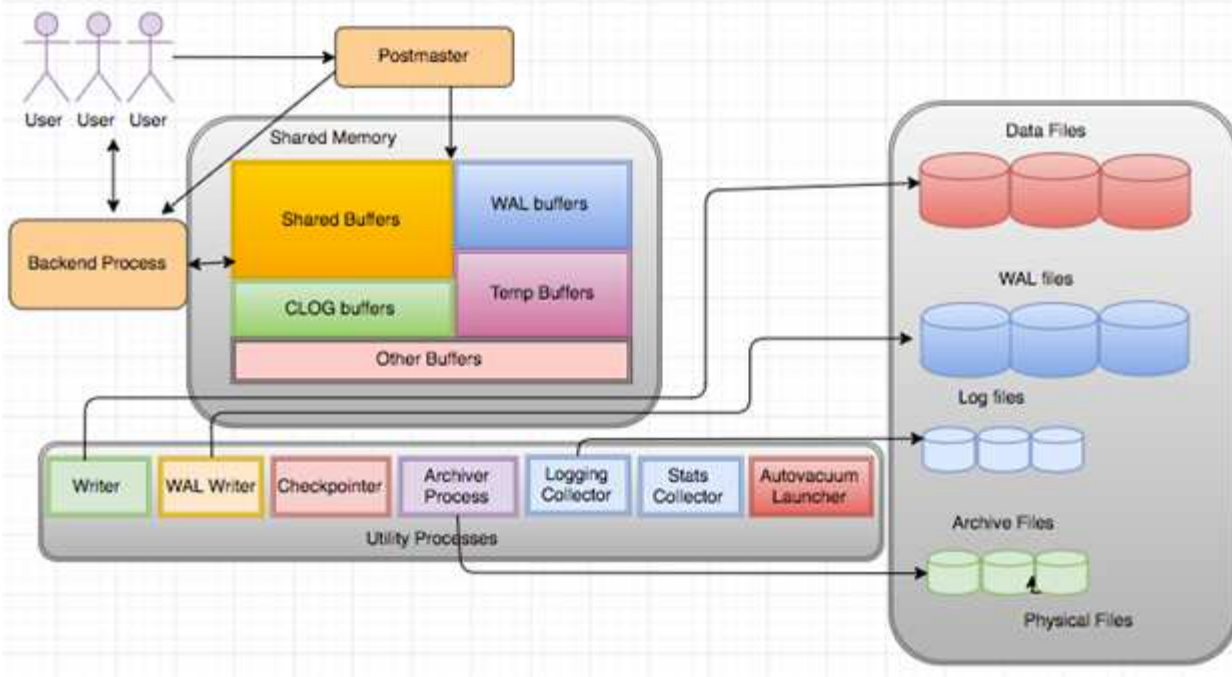
Mais importante ainda, o ONTAP oferece desempenho incomparável com a capacidade de dimensionar a solução para suas necessidades exclusivas. Nossos sistemas high-end podem fornecer mais de 1M IOPS com latências medidas em microssegundos. No entanto, se você só precisa de 100K IOPS, pode redimensionar sua solução de storage com uma controladora menor que ainda executa o mesmo sistema operacional de storage.

Configuração do banco de dados

Arquitetura

PostgreSQL é um RDBMS baseado na arquitetura cliente e servidor. Uma instância do PostgreSQL é conhecida como um cluster de banco de dados, que é uma coleção de bancos de dados em vez de uma coleção de servidores.

PostgreSQL Basic Architecture



Há três elementos principais em um banco de dados PostgreSQL: O postmaster, o front-end (cliente) e o back-end. O cliente envia solicitações para o postmaster com informações como protocolo IP e a que banco de dados se conectar. O postmaster autentica a conexão e a passa para o processo de back-end para comunicação adicional. O processo back-end executa a consulta e envia resultados diretamente para o front-end (cliente).

Uma instância do PostgreSQL é baseada em um modelo multiprocessado em vez de um modelo multithread. Ele gera vários processos para diferentes trabalhos, e cada processo tem sua própria funcionalidade. Os principais processos incluem o processo do cliente, o processo do escritor DE WAL, o processo do escritor de fundo, e o processo do checkpointer:

- Quando um processo de cliente (primeiro plano) envia solicitações de leitura ou gravação para a instância do PostgreSQL, ele não lê ou grava dados diretamente no disco. Primeiro, ele armazena os dados em buffers compartilhados e buffers de log write-ahead (WAL).
- Um processo de escritor WAL manipula o conteúdo dos buffers compartilhados e buffers WAL para escrever nos logs DO WAL. Os logs WAL são normalmente Registros de transações do PostgreSQL e são sequencialmente escritos. Portanto, para melhorar o tempo de resposta do banco de dados, o PostgreSQL primeiro grava nos logs de transação e reconhece o cliente.
- Para colocar o banco de dados em um estado consistente, o processo de gravação em segundo plano verifica o buffer compartilhado periodicamente para páginas sujas. Em seguida, ele limpa os dados para os arquivos de dados armazenados em volumes NetApp ou LUNs.
- O processo de checkpointer também é executado periodicamente (com menos frequência do que o processo em segundo plano) e impede qualquer modificação nos buffers. Ele sinaliza para o processo de gravação WAL para escrever e lavar o Registro do ponto de verificação para o final dos logs DO WAL que são armazenados no disco NetApp. Ele também sinaliza o processo de gravação em segundo plano para escrever e lavar todas as páginas sujas para o disco.

Parâmetros de inicialização

Você cria um novo cluster de banco de dados usando o `initdb` programa. `initdb`Um`

script cria os arquivos de dados, tabelas do sistema e bancos de dados de modelos (template0 e template1) que definem o cluster.

A base de dados de modelos representa uma base de dados de stock. Ele contém definições para tabelas de sistema, exibições padrão, funções e tipos de dados. `pgdata` atua como um argumento para o `initdb` script que especifica a localização do cluster do banco de dados.

Todos os objetos de banco de dados no PostgreSQL são gerenciados internamente pelos respectivos OIDs. Tabelas e índices também são gerenciados por OIDs individuais. As relações entre objetos de banco de dados e seus respectivos OIDs são armazenadas em tabelas de catálogo de sistema apropriadas, dependendo do tipo de objeto. Por exemplo, OIDs de bancos de dados e tabelas de heap são armazenados em `pg_database` e `'pg_class`, respectivamente. Você pode determinar os OIDs emitindo consultas no cliente PostgreSQL.

Cada banco de dados tem suas próprias tabelas individuais e arquivos de índice que são restritos a 1GB. Cada tabela tem dois arquivos associados, sufixos respectivamente com `_fsm` e `_vm`. Eles são referidos como o mapa de espaço livre e o mapa de visibilidade. Esses arquivos armazenam as informações sobre a capacidade de espaço livre e têm visibilidade em cada página no arquivo de tabela. Os índices têm apenas mapas de espaço livre individuais e não têm mapas de visibilidade.

O `pg_xlog/pg_wal` diretório contém os registros de escrita antecipada. Os logs write-ahead são usados para melhorar a confiabilidade e o desempenho do banco de dados. Sempre que você atualizar uma linha em uma tabela, o PostgreSQL primeiro grava a alteração no log write-ahead e, mais tarde, grava as modificações nas páginas de dados reais em um disco. O `pg_xlog` diretório geralmente contém vários arquivos, mas `initdb` cria apenas o primeiro. Arquivos extras são adicionados conforme necessário. Cada arquivo xlog tem 16MBcm de comprimento.

Definições

Existem várias configurações de ajuste do PostgreSQL que podem melhorar o desempenho.

Os parâmetros mais utilizados são os seguintes:

- `max_connections` = <num>: O número máximo de conexões de banco de dados a ter ao mesmo tempo. Use este parâmetro para restringir a troca para o disco e eliminar o desempenho. Dependendo do requisito do aplicativo, você também pode ajustar esse parâmetro para as configurações do pool de conexões.
- `shared_buffers` = <num>: O método mais simples para melhorar o desempenho do seu servidor de banco de dados. O padrão é baixo para a maioria dos hardwares modernos. Ele é definido durante a implantação para aproximadamente 25% da RAM disponível no sistema. Essa configuração de parâmetro varia dependendo de como ele funciona com instâncias de banco de dados específicas; você pode ter que aumentar e diminuir os valores por tentativa e erro. No entanto, a configuração alta pode degradar o desempenho.
- `effective_cache_size` = <num>: Este valor diz ao otimizador do PostgreSQL quanta memória o PostgreSQL tem disponível para armazenar dados em cache e ajuda a determinar se deve usar um índice. Um valor maior aumenta a probabilidade de usar um índice. Este parâmetro deve ser definido para a quantidade de memória alocada para `shared_buffers` mais a quantidade de cache de SO disponível. Muitas vezes, esse valor é mais de 50% da memória total do sistema.
- `work_mem` = <num>: Este parâmetro controla a quantidade de memória a ser usada em operações de classificação e tabelas de hash. Se você fizer uma classificação pesada em seu aplicativo, talvez seja necessário aumentar a quantidade de memória, mas tenha cuidado. Não é um parâmetro amplo do

sistema, mas um parâmetro por operação. Se uma consulta complexa tiver várias operações de ordenação nela, ela usará várias unidades de memória `work_mem`, e vários back-ends poderiam estar fazendo isso simultaneamente. Essa consulta pode muitas vezes levar o servidor de banco de dados a trocar se o valor for muito grande. Essa opção foi anteriormente chamada `sort_mem` em em versões mais antigas do PostgreSQL.

- `fsync = <boolean> (on or off)`: Este parâmetro determina se todas as suas páginas WAL devem ser sincronizadas com o disco usando `fsync()` antes que uma transação seja confirmada. Desligá-lo às vezes pode melhorar o desempenho de gravação e ativá-lo aumenta a proteção contra o risco de corrupção quando o sistema trava.
- `checkpoint_timeout`: O processo de checkpoint elimina dados comprometidos no disco. Isso envolve muitas operações de leitura/gravação no disco. O valor é definido em segundos e valores mais baixos diminuem o tempo de recuperação de falhas e o aumento de valores pode reduzir a carga nos recursos do sistema reduzindo as chamadas de ponto de verificação. Dependendo da criticidade do aplicativo, uso, disponibilidade do banco de dados, defina o valor de `checkpoint_timeout`.
- `commit_delay = <num>` E `commit_siblings = <num>`: essas opções são usadas em conjunto para ajudar a melhorar o desempenho, escrevendo várias transações que estão cometendo de uma só vez. Se houver vários objetos `commit_siblings` ativos no momento em que a transação está sendo feita, o servidor aguarda por `commit_delay` microssegundos para tentar cometer várias transações de uma só vez.
- `max_worker_processes / max_parallel_workers`: Configurar o número ideal de trabalhadores para processos. `Max_Parallel_workers` corresponde ao número de CPUs disponíveis. Dependendo do design do aplicativo, as consultas podem exigir um número menor de trabalhadores para operações paralelas. É melhor manter o valor para ambos os parâmetros o mesmo, mas ajustar o valor após o teste.
- `random_page_cost = <num>`: Este valor controla a forma como o PostgreSQL visualiza leituras de discos não sequenciais. Um valor maior significa que o PostgreSQL é mais provável de usar uma varredura sequencial em vez de uma varredura de índice, indicando que seu servidor tem discos rápidos. Modificar esta configuração depois de avaliar outras opções como otimização baseada em plano, aspiração, indexação para alterar consultas ou esquema.
- `effective_io_concurrency = <num>`: Este parâmetro define o número de operações de e/S de disco simultâneas que o PostgreSQL tenta executar simultaneamente. Aumentar esse valor aumenta o número de operações de e/S que qualquer sessão individual do PostgreSQL tenta iniciar em paralelo. O intervalo permitido é de 1 a 1.000, ou zero para desativar a emissão de solicitações de e/S assíncronas. Atualmente, essa configuração afeta somente digitalizações de heap bitmap. As unidades de estado sólido (SSDs) e outro storage baseado em memória (NVMe) costumam processar muitas solicitações simultâneas. Assim, o melhor valor pode ser nas centenas.

Consulte a documentação do PostgreSQL para obter uma lista completa dos parâmetros de configuração do PostgreSQL.

BRINDE

TOAST representa a técnica de armazenamento de atributos oversized. PostgreSQL usa um tamanho de página fixo (geralmente 8KB) e não permite que tuplas abranjam várias páginas. Portanto, não é possível armazenar grandes valores de campo diretamente. Quando você tenta armazenar uma linha que excede esse tamanho, O TOAST divide os dados de colunas grandes em "pedaços" menores e os armazena em uma MESA DE BRINDE.

Os grandes valores dos atributos tostados são retirados (se selecionados) apenas no momento em que o conjunto de resultados é enviado ao cliente. A tabela em si é muito menor e pode caber mais linhas no cache de buffer compartilhado do que poderia sem qualquer armazenamento fora de linha (TOAST).

VÁCUO

Na operação normal do PostgreSQL, tuplas que são excluídas ou tornadas obsoletas por uma atualização não são removidas fisicamente de sua tabela; elas permanecem presentes até que O VÁCUO seja executado. Portanto, você deve executar O VÁCUO periodicamente, especialmente em tabelas atualizadas com frequência. O espaço que ocupa deve ser recuperado para reutilização por novas linhas, para evitar a interrupção do espaço em disco. No entanto, ele não retorna o espaço para o sistema operacional.

O espaço livre dentro de uma página não é fragmentado. VACUUM reescreve todo o bloco, empacotando eficientemente as linhas restantes e deixando um único bloco contíguo de espaço livre em uma página.

Em contraste, O VACUUM FULL compacta ativamente as tabelas escrevendo uma versão completamente nova do arquivo de tabela sem espaço morto. Esta ação minimiza o tamanho da mesa, mas pode levar muito tempo. Ele também requer espaço em disco extra para a nova cópia da tabela até que a operação seja concluída. O objetivo do VÁCUO de rotina é evitar a atividade TOTAL DO VÁCUO. Este processo não só mantém as tabelas em seu tamanho mínimo, mas também mantém o uso em estado estável do espaço em disco.

Tablespaces

Dois espaços de tablespaces são criados automaticamente quando o cluster de banco de dados é inicializado.

O `pg_global` espaço de tabela é usado para catálogos de sistemas compartilhados. O `pg_default` espaço de tabela é o espaço de tabela padrão dos bancos de dados `template1` e `template0`. Se a partição ou o volume em que o cluster foi inicializado ficar sem espaço e não puder ser estendido, um espaço de tabela pode ser criado em uma partição diferente e usado até que o sistema possa ser reconfigurado.

Um índice que é muito usado pode ser colocado em um disco rápido e altamente disponível, como um dispositivo de estado sólido. Além disso, uma tabela que armazena dados arquivados que raramente são usados ou não críticos ao desempenho pode ser armazenada em um sistema de disco mais lento e menos caro, como unidades SAS ou SATA.

As tablespaces são uma parte do cluster do banco de dados e não podem ser tratadas como uma coleção autônoma de arquivos de dados. Eles dependem dos metadados contidos no diretório de dados principal e, portanto, não podem ser anexados a um cluster de banco de dados diferente ou fazer backup individualmente. Da mesma forma, se você perder um espaço de tabela (por meio da exclusão de arquivos, falha de disco, etc.), o cluster do banco de dados pode se tornar ilegível ou incapaz de iniciar. Colocar um espaço de tabela em um sistema de arquivos temporário como um disco RAM corre o risco de a confiabilidade de todo o cluster.

Depois que ele é criado, um espaço de tabela pode ser usado a partir de qualquer banco de dados se o usuário solicitante tiver Privileges suficiente. PostgreSQL usa links simbólicos para simplificar a implementação de espaços de tabela. O PostgreSQL adiciona uma linha à `pg_tablespace` tabela (uma tabela abrangente) e atribui um novo identificador de objeto (OID) a essa linha. Finalmente, o servidor usa o OID para criar um link simbólico entre o cluster e o diretório fornecido. O diretório `$PGDATA/pg_tblspc` contém links simbólicos que apontam para cada um dos espaços de tablespaces não incorporados definidos no cluster.

Configuração de armazenamento

NFS

Os bancos de dados PostgreSQL podem ser hospedados em sistemas de arquivos

NFSv3 ou NFSv4. A melhor opção depende de fatores fora do banco de dados.

Por exemplo, o comportamento de bloqueio NFSv4D pode ser preferível em determinados ambientes em cluster. ("aqui"Consulte para obter detalhes adicionais)

A funcionalidade do banco de dados deve estar próxima da mesma, incluindo o desempenho. O único requisito é o uso da `hard` opção de montagem. Isso é necessário para garantir que os tempos limite flexíveis não produzam erros de e/S irrecuperáveis.

Se NFSv4 for escolhido como um protocolo, a NetApp recomenda o uso de NFSv4,1. Há algumas melhorias funcionais no protocolo NFSv4 em NFSv4,1 que aumentam a resiliência em relação a NFSv4,0.

Use as seguintes opções de montagem para cargas de trabalho gerais de banco de dados:

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=65536,wsiz=65536
```

Se for esperado uma e/S sequencial grande, os tamanhos de transferência NFS podem ser aumentados conforme descrito na seção a seguir.

Tamanhos de transferência NFS

Por padrão, o ONTAP limita os tamanhos de e/S de NFS a 64K.

A e/S aleatória com a maioria dos aplicativos e bancos de dados usa um tamanho de bloco muito menor que está bem abaixo do máximo 64K. A e/S de bloco grande geralmente é paralelizada, portanto o máximo de 64K GB também não é uma limitação para obter largura de banda máxima.

Existem algumas cargas de trabalho em que o máximo 64K cria uma limitação. Em particular, operações de um único processo, como operação de backup ou recuperação ou uma verificação de tabela completa de banco de dados, são executadas de forma mais rápida e eficiente se o banco de dados puder executar menos, mas maiores I/os. O tamanho ideal de manuseio de e/S para ONTAP é 256K.

O tamanho máximo de transferência para um determinado SVM do ONTAP pode ser alterado da seguinte forma:

```
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::*>
```

Cuidado

Nunca diminua o tamanho máximo de transferência permitido no ONTAP abaixo do valor de rsize/wsize dos sistemas de arquivos NFS atualmente instalados. Isso pode criar pendências ou até mesmo corrupção de dados com alguns sistemas operacionais. Por exemplo, se os clientes NFS estiverem atualmente definidos em um rsize/wsize de 65536, o tamanho máximo de transferência do ONTAP poderá ser ajustado entre 65536 e 1048576 sem efeito porque os próprios clientes são limitados. Reduzir o tamanho máximo de transferência abaixo de 65536 pode danificar a disponibilidade ou os dados.

Uma vez que o tamanho da transferência é aumentado no nível ONTAP, as seguintes opções de montagem seriam usadas:

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=262144,wsiz=262144
```

NFSv3 Tabelas de slot TCP

Se NFSv3 for usado com Linux, é fundamental definir corretamente as tabelas de slot TCP.

As tabelas de slot TCP são equivalentes a NFSv3 mm de profundidade de fila do adaptador de barramento do host (HBA). Essas tabelas controlam o número de operações NFS que podem ficar pendentes de uma só vez. O valor padrão é geralmente 16, o que é muito baixo para um desempenho ideal. O problema oposto ocorre em kernels Linux mais recentes, que podem aumentar automaticamente o limite da tabela de slots TCP para um nível que satura o servidor NFS com solicitações.

Para um desempenho ideal e para evitar problemas de desempenho, ajuste os parâmetros do kernel que controlam as tabelas de slots TCP.

Executar o `sysctl -a | grep tcp.*.slot_table` comando e respeitar os seguintes parâmetros:

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Todos os sistemas Linux devem incluir `sunrpc.tcp_slot_table_entries`, mas apenas alguns incluem `sunrpc.tcp_max_slot_table_entries`. Ambos devem ser definidos para 128.

Cuidado

A falha em definir esses parâmetros pode ter efeitos significativos no desempenho. Em alguns casos, o desempenho é limitado porque o sistema operacional linux não está emitindo e/S suficiente. Em outros casos, as latências de e/S aumentam à medida que o sistema operacional linux tenta emitir mais e/S do que pode ser reparado.

SAN

Os bancos de dados PostgreSQL com SAN geralmente são hospedados em sistemas de arquivos xfs, mas outros podem ser usados se suportados pelo fornecedor do sistema operacional.

Embora um único LUN possa geralmente suportar até 100K IOPS, os bancos de dados com uso intenso de e/S geralmente exigem o uso de LVM com striping.

LVM Striping

Antes da era das unidades flash, a distribuição foi usada para ajudar a superar as limitações de desempenho das unidades giratórias. Por exemplo, se um sistema operacional precisar executar uma operação de leitura 1MB, ler que 1MB TB de dados de uma única unidade exigiria muita busca e leitura de cabeça de unidade, pois o 1MB é transferido lentamente. Se esse 1MB TB de dados fosse distribuído em 8 LUNs, o sistema operacional poderia emitir oito operações de leitura 128K em paralelo e reduzir o tempo necessário para concluir a transferência 1MB.

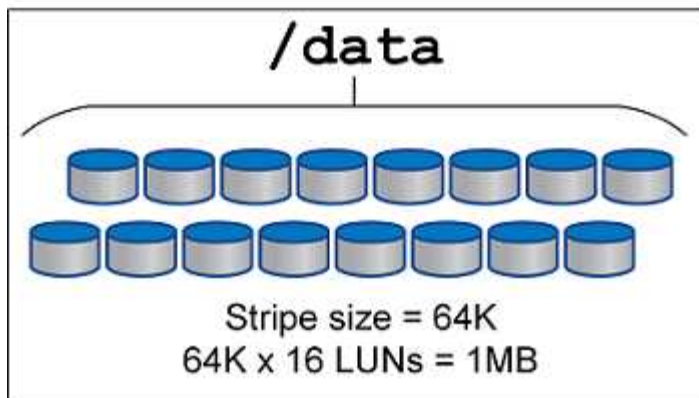
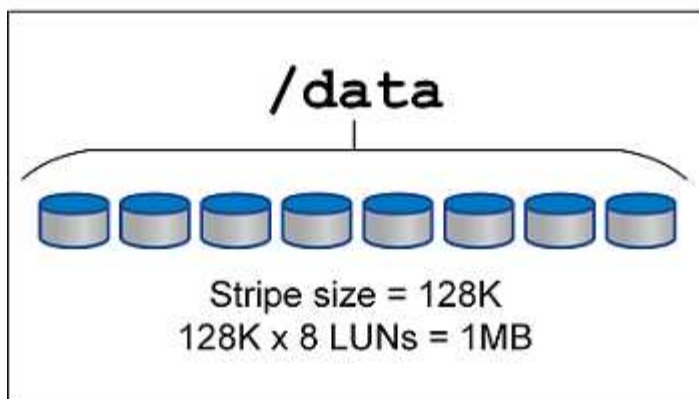
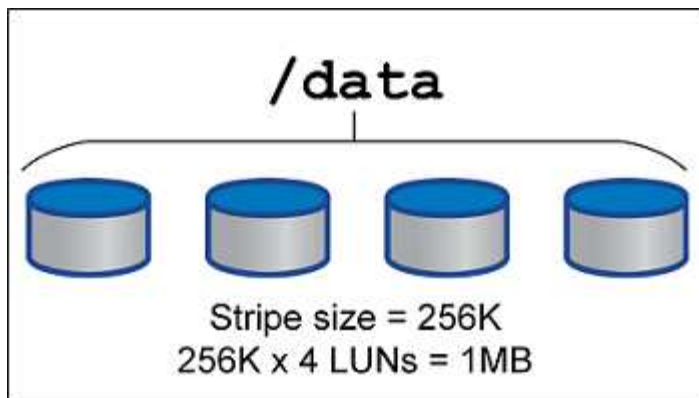
Riscar com unidades giratórias foi mais difícil porque o padrão de e/S tinha que ser conhecido com antecedência. Se o striping não foi ajustado corretamente para os padrões de e/S verdadeiros, as configurações listradas podem danificar o desempenho. Com os bancos de dados Oracle e, especialmente, com configurações all-flash, a distribuição é muito mais fácil de configurar e, comprovadamente, aumentou significativamente o desempenho.

Gerenciadores de volume lógicos, como o Oracle ASM stripe por padrão, mas o LVM de SO nativo não. Alguns deles unem vários LUNs como um dispositivo concatenado, o que resulta em datafiles que existem em um e apenas um dispositivo LUN. Isso causa pontos quentes. Outras implementações de LVM são padrão para extensões distribuídas. Isso é semelhante ao striping, mas é mais grosseiro. Os LUNs no grupo de volumes são cortados em pedaços grandes, chamados de extensões e normalmente medidos em muitos megabytes, e os volumes lógicos são então distribuídos por essas extensões. O resultado é a e/S aleatória contra um arquivo deve ser bem distribuída entre LUNs, mas as operações de e/S sequenciais não são tão eficientes quanto poderiam ser.

A e/S de aplicativos intensivos quase sempre é (a) em unidades do tamanho básico do bloco ou (b) um megabyte.

O principal objetivo de uma configuração de distribuição é garantir que a e/S de arquivo único possa ser executada como uma única unidade, e as e/S de vários blocos, que devem ter 1MB MB de tamanho, possam ser paralelizadas uniformemente em todos os LUNs no volume distribuído. Isso significa que o tamanho do stripe não deve ser menor que o tamanho do bloco do banco de dados e o tamanho do stripe multiplicado pelo número de LUNs deve ser 1MB.

A figura a seguir mostra três opções possíveis para o tamanho da faixa e ajuste de largura. O número de LUNs é selecionado para atender aos requisitos de performance conforme descrito acima, mas, em todos os casos, o total de dados em um único stripe é de 1MB.



Proteção de dados

Proteção Ddta nativa

Um dos principais aspetos do design de armazenamento é habilitar a proteção para volumes PostgreSQL. Os clientes podem proteger seus bancos de dados PostgreSQL usando a abordagem dump ou usando backups do sistema de arquivos. Esta seção explica as diferentes abordagens de backup de bancos de dados individuais ou de todo o cluster.

Existem três abordagens para fazer backup de dados PostgreSQL:

- Despejo do SQL Server
- Backup no nível do sistema de arquivo

- Arquivamento contínuo

A ideia por trás do método de despejo do SQL Server é gerar um arquivo com comandos do SQL Server que, quando retornados ao servidor, pode recriar o banco de dados como era no momento do despejo.

PostgreSQL fornece os programas utilitários `pg_dump` e `pg_dump_all` para criar backup individual e em nível de cluster. Esses despejos são lógicos e não contêm informações suficientes para serem usados pelo replay WAL.

Uma estratégia alternativa de backup é usar o backup em nível de sistema de arquivos, no qual os administradores copiam diretamente os arquivos que o PostgreSQL usa para armazenar os dados no banco de dados. Este método é feito no modo offline: O banco de dados ou cluster deve ser desligado. Outra alternativa é usar `pg_basebackup` para executar o backup de streaming quente do banco de dados PostgreSQL.

Instantâneos

Os backups baseados em snapshot com PostgreSQL exigem a configuração de snapshots para datafiles, arquivos WAL e ARQUIVOS WAL arquivados para fornecer recuperação completa ou pontual.

Para bancos de dados PostgreSQL, o tempo médio de backup com snapshots está no intervalo de alguns segundos a alguns minutos. Essa velocidade de backup é 60 a 100 vezes mais rápida do que `pg_basebackup` outras abordagens de backup baseadas no sistema de arquivos.

Os snapshots no storage do NetApp podem ser consistentes com falhas e com aplicações. Um snapshot consistente com falhas é criado no storage sem silenciar o banco de dados, enquanto um snapshot consistente com aplicativos é criado enquanto o banco de dados está no modo de backup. O NetApp também garante que os snapshots subsequentes sejam backups incrementais para sempre, a fim de promover a economia de storage e a eficiência da rede.

Como os snapshots são rápidos e não afetam o desempenho do sistema, você pode agendar vários snapshots diariamente em vez de criar um único backup diário, como acontece com outras tecnologias de backup de streaming. Quando uma operação de restauração e recuperação é necessária, o tempo de inatividade do sistema é reduzido por dois recursos principais:

- A tecnologia de recuperação de dados NetApp SnapRestore significa que a operação de restauração é executada em segundos.
- Objetivos agressivos do ponto de restauração (RPOs) significam que menos logs do banco de dados devem ser aplicados e a recuperação avançada também é acelerada.

Para fazer backup do PostgreSQL, você deve garantir que os volumes de dados sejam protegidos simultaneamente com O WAL (Consistency-group) e os logs arquivados. Enquanto você estiver usando a tecnologia Snapshot para copiar ARQUIVOS WAL, certifique-se de executar `pg_stop` para liberar todas as entradas WAL que devem ser arquivadas. Se você lavar as ENTRADAS WAL durante a restauração, precisará apenas parar o banco de dados, desmontar ou excluir o diretório de dados existente e executar uma operação SnapRestore no armazenamento. Depois que a restauração for concluída, você pode montar o sistema e trazê-lo de volta ao seu estado atual. Para recuperação pontual, você também pode restaurar logs DO WAL e do arquivo; em seguida, o PostgreSQL decide o ponto mais consistente e recupera-o automaticamente.

Os grupos de consistência são um recurso no ONTAP e são recomendados quando há vários volumes montados em uma única instância ou em um banco de dados com vários espaços de tabela. Um snapshot de grupo de consistência garante que todos os volumes sejam agrupados e protegidos. Um grupo de

consistência pode ser gerenciado de forma eficiente a partir do Gerenciador de sistemas do ONTAP e até mesmo cloná-lo para criar uma cópia de instância de um banco de dados para fins de teste ou desenvolvimento.

Software de proteção de dados

O plugin NetApp SnapCenter para banco de dados PostgreSQL, combinado com as tecnologias Snapshot e NetApp FlexClone, oferece benefícios como:

- Backup e restauração rápidos.
- Clones com uso eficiente de espaço.
- A capacidade de construir um sistema de recuperação de desastres rápido e eficaz.

Você pode preferir escolher os parceiros premium de backup da NetApp, como o Veeam Software e a CommVault, nas seguintes circunstâncias:



- Gerenciamento de workloads em um ambiente heterogêneo
- Armazenamento de backups na nuvem ou em fita para retenção de longo prazo
- Suporte para uma ampla gama de versões e tipos de SO

O plugin SnapCenter para PostgreSQL é um plugin compatível com a comunidade e a configuração e documentação estão disponíveis na loja de automação NetApp. Com o SnapCenter, o usuário pode fazer backup do banco de dados, clonar e restaurar dados remotamente.

VMware

VMware vSphere com ONTAP

VMware vSphere com ONTAP

O ONTAP é uma solução de storage líder para ambientes VMware vSphere há quase duas décadas e continua adicionando recursos inovadores para simplificar o gerenciamento e reduzir custos. Este documento apresenta a solução ONTAP para vSphere, incluindo as informações mais recentes sobre produtos e as melhores práticas, para simplificar a implantação, reduzir riscos e simplificar o gerenciamento.



Esta documentação substitui os relatórios técnicos publicados anteriormente *TR-4597: VMware vSphere for ONTAP*

As práticas recomendadas complementam outros documentos, como guias e listas de compatibilidade. Eles são desenvolvidos com base em testes de laboratório e extensa experiência de campo por engenheiros e clientes da NetApp. Elas podem não ser as únicas práticas suportadas que funcionam em todos os ambientes, mas são geralmente as soluções mais simples que atendem às necessidades da maioria dos clientes.

Este documento se concentra em recursos nas versões recentes do ONTAP (9.x) executadas no vSphere 7,0 ou posterior. Consulte "[Ferramenta de Matriz de interoperabilidade do NetApp](#)" e "[Guia de compatibilidade da VMware](#)" para obter detalhes relacionados a lançamentos específicos.

Por que o ONTAP para vSphere?

Há muitos motivos pelos quais dezenas de milhares de clientes escolheram o ONTAP como sua solução de storage para o vSphere, como um sistema de storage unificado compatível com protocolos SAN e nas, recursos robustos de proteção de dados usando snapshots com uso eficiente de espaço e diversas ferramentas para ajudar você a gerenciar dados de aplicações. O uso de um sistema de storage separado do hipervisor permite descarregar várias funções e maximizar seu investimento nos sistemas de host vSphere. Essa abordagem não só garante que os recursos de host estejam focados nas cargas de trabalho dos aplicativos, mas também evita efeitos aleatórios de desempenho nos aplicativos das operações de storage.

O uso do ONTAP em conjunto com o vSphere é uma ótima combinação que permite reduzir as despesas de hardware do host e software VMware. Você também pode proteger seus dados a um custo menor com alta performance consistente. Como as cargas de trabalho virtualizadas são móveis, você pode explorar diferentes abordagens usando o Storage vMotion para mover VMs entre armazenamentos de dados VMFS, NFS ou vVols, tudo no mesmo sistema de storage.

Aqui estão os principais fatores que os clientes valorizam hoje:

- **Armazenamento unificado.** Os sistemas que executam o ONTAP são unificados de várias maneiras significativas. Originalmente, essa abordagem se referia aos protocolos nas e SAN, e a ONTAP continua a ser uma plataforma líder para SAN, juntamente com sua força original no nas. No mundo vSphere, essa abordagem também pode significar um sistema unificado para infraestrutura de desktop virtual (VDI), juntamente com a infraestrutura de servidor virtual (VSI). Os sistemas que executam o ONTAP geralmente são mais baratos para VSI do que os arrays empresariais tradicionais e ainda têm recursos avançados de eficiência de storage para lidar com a VDI no mesmo sistema. O ONTAP também unifica uma variedade de Mídia de armazenamento, de SSDs a SATA, e pode estender isso facilmente para a nuvem. Não é necessário comprar um array flash para obter desempenho, um array SATA para arquivos e sistemas

separados para a nuvem. ONTAP une todos eles.

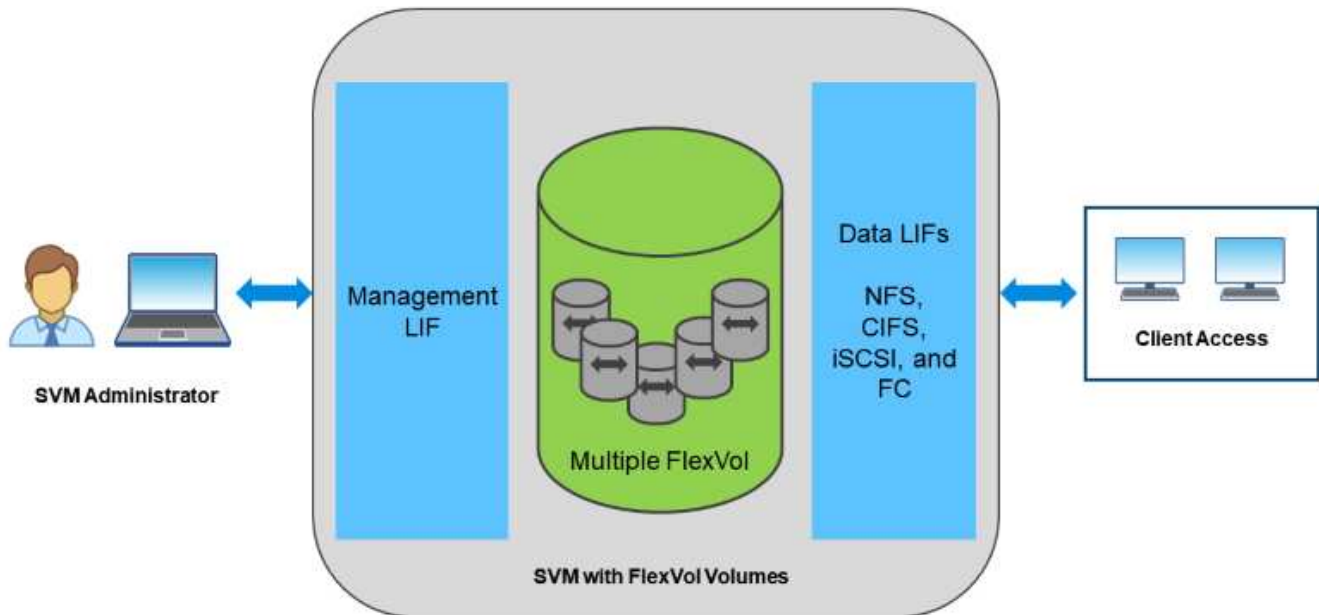
- **Gerenciamento baseado em políticas de armazenamento e volumes virtuais.** O NetApp foi um dos primeiros parceiros de design da VMware no desenvolvimento do vSphere Virtual volumes (vVols), fornecendo informações arquitetônicas e suporte antecipado para vVols e VMware vSphere APIs for Storage Awareness (VASA). Essa abordagem não só trouxe gerenciamento granular de storage de VM para o VMFS, mas também suportou a automação do provisionamento de storage por meio do gerenciamento baseado em políticas de storage. Com essa abordagem, os arquitetos de storage podem projetar pools de storage com recursos diferentes que podem ser facilmente consumidos pelos administradores de VM. A ONTAP lidera o setor de storage na escala da evolução, dando suporte a centenas de milhares de vols em um único cluster, enquanto os fornecedores de array empresarial e flash arrays menores oferecem suporte a apenas milhares de vols por array. O NetApp também está impulsionando a evolução do gerenciamento granular de VM com futuros recursos para suporte ao vVols 3,0.
- **Eficiência de armazenamento.** Embora a NetApp tenha sido a primeira a fornecer deduplicação para cargas de trabalho de produção, essa inovação não foi a primeira ou última nessa área. Ele começou com snapshots, um mecanismo de proteção de dados com uso eficiente de espaço sem efeito de performance e a tecnologia FlexClone para fazer cópias de leitura/gravação instantaneamente das VMs para uso em produção e backup. A NetApp passou a fornecer recursos inline, incluindo deduplicação, compactação e deduplicação de bloco zero, para aproveitar a maior parte do storage com SSDs caros. Mais recentemente, o ONTAP adicionou a capacidade de embalar operações e/S menores em um bloco de disco usando a compactação. A combinação desses recursos resultou em que os clientes viram economias de até 5:1 dólares para VSI e até 30:1 dólares para VDI.
- **Nuvem híbrida.** Sejam usadas para nuvem privada no local, infraestrutura de nuvem pública ou uma nuvem híbrida que combina o melhor dos dois, as soluções da ONTAP ajudam você a criar seu Data Fabric para otimizar e otimizar o gerenciamento de dados. Comece com sistemas all-flash de alta performance e combine-os com sistemas de storage em disco ou em nuvem para proteção de dados e computação em nuvem. Aproveite o suporte avançado para as tecnologias OpenStack e de contêiner, conforme necessário. O NetApp também oferece backup baseado em nuvem (nuvem SnapMirror, Cloud Backup Service e Cloud Sync) e ferramentas de arquivamento e disposição em camadas de storage (FabricPool) para ONTAP para ajudar a reduzir as despesas operacionais e aproveitar o amplo alcance da nuvem.
- **E muito mais.** Aproveite a performance extrema dos arrays NetApp AFF A-Series para acelerar sua infraestrutura virtualizada e gerenciar custos. Aproveite operações totalmente ininterruptas, desde a manutenção até os upgrades até a substituição completa do seu sistema de storage, usando clusters ONTAP com escalabilidade horizontal. Proteger dados em repouso com os recursos de criptografia do NetApp sem custo adicional. Certifique-se de que o desempenho atenda aos níveis de serviço de negócios por meio de recursos de qualidade de serviço refinados. Todos eles fazem parte da ampla variedade de recursos fornecidos com o ONTAP, o software de gerenciamento de dados empresariais líder do setor.

Storage unificado

O ONTAP unifica o storage por meio de uma abordagem simplificada definida por software para gerenciamento seguro e eficiente, performance aprimorada e escalabilidade otimizada. Essa abordagem aprimora a proteção de dados e permite o uso eficaz de recursos de nuvem.

Originalmente, essa abordagem unificada se referia ao suporte aos protocolos nas e SAN em um sistema de storage, e a ONTAP continua sendo uma plataforma líder de SAN, juntamente com sua força original em nas. O ONTAP agora também fornece suporte ao protocolo de objetos S3. Embora o S3 não seja usado para datastores, você pode usá-lo para aplicações in-Guest. Pode obter mais informações sobre o suporte ao protocolo S3 no ONTAP no ["Visão geral da configuração do S3"](#).

Uma máquina virtual de storage (SVM) é a unidade de alocação segura a vários clientes no ONTAP. É uma construção lógica que permite o acesso do cliente a sistemas que executam o ONTAP. Os SVMs podem servir dados simultaneamente por meio de vários protocolos de acesso a dados por meio de interfaces lógicas (LIFs). As SVMs fornecem acesso a dados no nível do arquivo por meio de protocolos nas, como CIFS e NFS, e acesso a dados em nível de bloco por meio de protocolos SAN, como iSCSI, FC/FCoE e NVMe. Os SVMs podem fornecer dados a clientes SAN e nas de forma independente ao mesmo tempo, bem como ao S3.



No mundo vSphere, essa abordagem também pode significar um sistema unificado para infraestrutura de desktop virtual (VDI), juntamente com a infraestrutura de servidor virtual (VSI). Os sistemas que executam o ONTAP geralmente são mais baratos para VSI do que os arrays empresariais tradicionais e ainda têm recursos avançados de eficiência de storage para lidar com a VDI no mesmo sistema. O ONTAP também unifica uma variedade de Mídia de armazenamento, de SSDs a SATA, e pode estender isso facilmente para a nuvem. Não é necessário comprar um array flash para obter desempenho, um array SATA para arquivos e sistemas separados para a nuvem. ONTAP une todos eles.

OBSERVAÇÃO: para obter mais informações sobre SVMs, armazenamento unificado e acesso ao cliente, consulte "[Virtualização de storage](#)" no Centro de Documentação do ONTAP 9.

Ferramentas de virtualização para ONTAP

O NetApp oferece várias ferramentas de software independentes que podem ser usadas em conjunto com o ONTAP e o vSphere para gerenciar seu ambiente virtualizado.

As ferramentas a seguir estão incluídas na licença ONTAP sem nenhum custo adicional. Consulte a Figura 1 para ver uma descrição de como essas ferramentas funcionam juntas no seu ambiente vSphere.

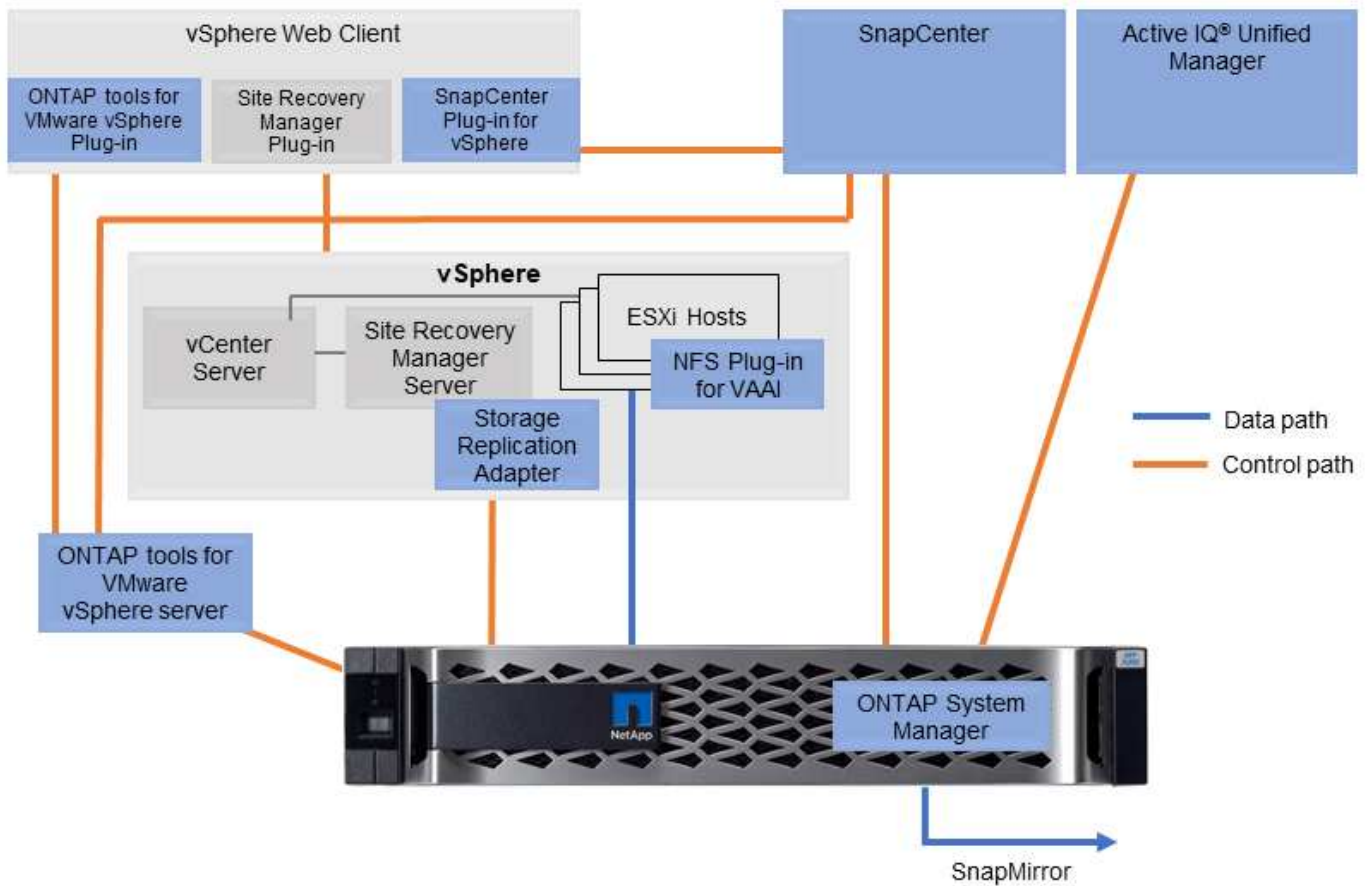
Ferramentas do ONTAP para VMware vSphere

As ferramentas do ONTAP para VMware vSphere são um conjunto de ferramentas para usar o armazenamento do ONTAP junto com o vSphere. O plug-in do vCenter, anteriormente conhecido como Virtual

Storage Console (VSC), simplifica o gerenciamento de storage e os recursos de eficiência, aprimora a disponibilidade e reduz os custos de storage e a sobrecarga operacional, independentemente de você estar usando SAN ou nas. Ele usa as práticas recomendadas para provisionar armazenamentos de dados e otimiza as configurações de host ESXi para ambientes de storage de bloco e NFS. Para todos esses benefícios, a NetApp recomenda o uso dessas ferramentas do ONTAP como uma prática recomendada ao usar o vSphere com sistemas que executam o ONTAP. Ele inclui um dispositivo de servidor, extensões de interface de usuário para vCenter, provedor VASA e adaptador de replicação de armazenamento. Quase tudo nas ferramentas do ONTAP pode ser automatizado com o uso de APIs REST simples, consumíveis pela maioria das ferramentas de automação modernas.

- *** Extensões de IU do vCenter.*** As extensões de IU das ferramentas do ONTAP simplificam o trabalho das equipes de operações e administradores do vCenter ao incorporar menus sensíveis ao contexto e fáceis de usar para gerenciar hosts e storage, portlets informativos e recursos de alerta nativos diretamente na IU do vCenter para fluxos de trabalho simplificados.
- **Fornecedor VASA para ONTAP.** O provedor VASA para ONTAP oferece suporte à estrutura VMware vStorage APIs for Storage Awareness (VASA). Ele é fornecido como parte das ferramentas do ONTAP para o VMware vSphere como um único dispositivo virtual para facilitar a implantação. O provedor VASA conecta o vCenter Server com o ONTAP para auxiliar no provisionamento e monitoramento do armazenamento de VM. Ele permite o suporte do VMware Virtual volumes (vVols), o gerenciamento de perfis de funcionalidades de storage e o desempenho individual de vVols de VM, além de alarmes para monitorar a capacidade e a conformidade com os perfis.
- **Adaptador de replicação de armazenamento.** O SRA é usado em conjunto com o VMware Site Recovery Manager (SRM) para gerenciar a replicação de dados entre locais de produção e recuperação de desastres e testar as réplicas de DR sem interrupções. Ele ajuda a automatizar as tarefas de descoberta, recuperação e reprotção. Inclui um dispositivo de servidor SRA e adaptadores SRA para o servidor SRM do Windows e o dispositivo SRM.

A figura a seguir mostra as ferramentas do ONTAP para vSphere.



Plug-in NFS para VMware VAAI

O plug-in NFS do NetApp para VMware VAAI é um plug-in para hosts ESXi que permite que eles usem recursos do VAAI com datastores NFS no ONTAP. Ele dá suporte a descarga de cópia para operações de clone, reserva de espaço para arquivos de disco virtual com espessura e descarga de snapshot. Descarregar operações de cópia para armazenamento não é necessariamente mais rápido de ser concluído, mas reduz os requisitos de largura de banda da rede e descarrega recursos do host, como ciclos de CPU, buffers e filas. Você pode usar as ferramentas do ONTAP para o VMware vSphere para instalar o plug-in em hosts ESXi ou, onde for compatível, o vLCM (vSphere Lifecycle Manager).

Volumes virtuais (vVols) e gerenciamento baseado em políticas de storage (SPBM)

O NetApp foi um dos primeiros parceiros de design da VMware no desenvolvimento do vSphere Virtual volumes (vVols), fornecendo informações arquitetônicas e suporte antecipado para vVols e VMware vSphere APIs for Storage Awareness (VASA). Essa abordagem não só trouxe gerenciamento de armazenamento granular de VM para o VMFS, mas também suportou a automação do provisionamento de armazenamento por meio do Storage Policy Based Management (SPBM).

O SPBM fornece uma estrutura que serve como uma camada de abstração entre os serviços de armazenamento disponíveis para o seu ambiente de virtualização e os elementos de armazenamento provisionados por meio de políticas. Com essa abordagem, os arquitetos de storage podem projetar pools de storage com recursos diferentes que podem ser facilmente consumidos pelos administradores de VM. Os administradores podem então combinar os requisitos de carga de trabalho da máquina virtual com os pools de armazenamento provisionados, permitindo o controle granular de várias configurações em um nível por VM ou disco virtual.

A ONTAP é líder do setor de storage em escala vVols, oferecendo suporte a centenas de milhares de vVols em um único cluster. Já os fornecedores de array empresarial e flash array menores oferecem suporte a apenas milhares de vols por array. O NetApp também está impulsionando a evolução do gerenciamento granular de VM com futuros recursos para suporte ao vVols 3,0.



Para obter mais informações sobre o VMware vSphere Virtual volumes, SPBM e ONTAP, "[TR-4400: VMware vSphere Virtual volumes com ONTAP](#)" consulte .

Armazenamentos de dados e protocolos

Visão geral dos recursos do vSphere datastore e do protocolo

Seis protocolos são usados para conectar o VMware vSphere a datastores em um sistema que executa o ONTAP:

- FCP
- NVMe/FC
- NVMe/TCP
- iSCSI
- NFS v3
- NFS v4.1

FCP, NVMe/FC, NVMe/TCP e iSCSI são protocolos de bloco que usam o vSphere Virtual Machine File System (VMFS) para armazenar VMs dentro de LUNs ONTAP ou namespaces NVMe contidos em um ONTAP FlexVol volume. O NFS é um protocolo de arquivos que coloca as VMs em armazenamentos de dados (que são simplesmente volumes ONTAP) sem a necessidade de VMFS. SMB (CIFS), iSCSI, NVMe/TCP ou NFS também podem ser usados diretamente de um sistema operacional convidado para o ONTAP.

As tabelas a seguir apresentam recursos tradicionais de datastore compatíveis com o vSphere com o ONTAP. Essas informações não se aplicam aos armazenamentos de dados vVols, mas geralmente se aplicam ao vSphere 6.x e versões posteriores usando versões compatíveis do ONTAP. Você também pode consultar "[Valores máximos de configuração da VMware](#)" versões específicas do vSphere para confirmar limites específicos.

Capacidade/função	FC	iSCSI	NVMe-of	NFS
Formato	VMFS ou mapeamento de dispositivo bruto (RDM)	VMFS ou RDM	VMFS	N/A.
Número máximo de armazenamentos de dados ou LUNs	1024 LUNs por host	1024 LUNs por servidor	256 Namespaces por servidor	O 256 monta o NFS padrão. MaxVolumes é 8. Use as ferramentas do ONTAP para o VMware vSphere para aumentar para 256.

Capacidade/função nalidade	FC	ISCSI	NVMe-of	NFS
Tamanho máximo do datastore	64 TB	64 TB	64 TB	100TB FlexVol volume ou superior com volume FlexGroup
Tamanho máximo do arquivo do datastore	62 TB	62 TB	62 TB	62TB com ONTAP 9.12.1P2 e posterior
Profundidade de fila ideal por LUN ou sistema de ficheiros	64-256	64-256	Negociação automática	Consulte NFS.MaxQueueDepth em " Host ESXi recomendado e outras configurações do ONTAP ".

A tabela a seguir lista as funcionalidades relacionadas ao armazenamento VMware suportadas.

Capacidade/recursos	FC	ISCSI	NVMe-of	NFS
VMotion	Sim	Sim	Sim	Sim
Storage vMotion	Sim	Sim	Sim	Sim
VMware HA	Sim	Sim	Sim	Sim
Storage Distributed Resource Scheduler (SDRS)	Sim	Sim	Sim	Sim
Software de backup habilitado para VMware vStorage APIs for Data Protection (VADP)	Sim	Sim	Sim	Sim
Microsoft Cluster Service (MSCS) ou cluster de failover em uma VM	Sim	Sim 1	Sim 1	Não suportado
Tolerância de falhas	Sim	Sim	Sim	Sim
Site Recovery Manager	Sim	Sim	2	V3 apenas 2
VMs com thin Provisioning (discos virtuais)	Sim	Sim	Sim	Sim esta configuração é o padrão para todas as VMs no NFS quando não estiver usando o VAAI.

Capacidade/recursos	FC	ISCSI	NVMe-of	NFS
Multipathing nativo da VMware	Sim	Sim	Sim, usando o novo HPP (High Performance Plugin)	O entroncamento de sessão NFS v4,1 requer ONTAP 9.14,1 e posterior

A tabela a seguir lista os recursos de gerenciamento de armazenamento ONTAP compatíveis.

Capacidade/função	FC	ISCSI	NVMe-of	NFS
Deduplicação de dados	Economia no array	Economia no array	Economia no array	Economia no datastore
Thin Provisioning	Datastore ou RDM	Datastore ou RDM	Armazenamento de dados	Armazenamento de dados
Redimensione o datastore	Cresça apenas	Cresça apenas	Cresça apenas	Crescer, crescer com crescimento automático e diminuir
Plug-ins do SnapCenter para Windows, aplicações Linux (no convidado)	Sim	Sim	Não	Sim
Monitoramento e configuração de host usando as ferramentas do ONTAP para VMware vSphere	Sim	Sim	Não	Sim
Provisionamento usando as ferramentas do ONTAP para VMware vSphere	Sim	Sim	Não	Sim

A tabela a seguir lista os recursos de backup suportados.

Capacidade/função	FC	ISCSI	NVMe-of	NFS
Snapshots ONTAP	Sim	Sim	Sim	Sim
SRM suportado por backups replicados	Sim	Sim	2	V3 apenas 2
Volume SnapMirror	Sim	Sim	Sim	Sim

Capacidade/função	FC	iSCSI	NVMe-of	NFS
Acesso à imagem VMDK	Software de backup habilitado para VADP	Software de backup habilitado para VADP	Software de backup habilitado para VADP	Software de backup habilitado para VADP, vSphere Client e vSphere Web Client datastore browser
Acesso ao nível do arquivo VMDK	Software de backup habilitado para VADP, somente Windows	Software de backup habilitado para VADP, somente Windows	Software de backup habilitado para VADP, somente Windows	Software de backup habilitado para VADP e aplicativos de terceiros
Granularidade NDMP	Armazenamento de dados	Armazenamento de dados	Armazenamento de dados	Datastore ou VM

1 *a NetApp recomenda o uso de iSCSI in-Guest para clusters da Microsoft em vez de VMDKs habilitados para vários gravadores em um armazenamento de dados VMFS. Essa abordagem é totalmente suportada pela Microsoft e VMware, oferece grande flexibilidade com o ONTAP (sistemas SnapMirror para ONTAP no local ou na nuvem), é fácil de configurar e automatizar e pode ser protegido com o SnapCenter. O vSphere 7 adiciona uma nova opção de VMDK em cluster. Isso é diferente dos VMDKs habilitados para vários gravadores, o que requer um armazenamento de dados apresentado por meio do protocolo FC que tem o suporte do VMDK em cluster habilitado. Aplicam-se outras restrições. Consulte a documentação da VMware ["Configuração para Cluster de failover do Windows Server"](#) para obter diretrizes de configuração.

2 os armazenamentos de dados usando NVMe-of e NFS v4,1 exigem replicação do vSphere. A replicação baseada em array não é suportada pelo SRM.

Selecionar um protocolo de armazenamento

Os sistemas que executam o ONTAP dão suporte a todos os principais protocolos de storage. Assim, os clientes podem escolher o que é melhor para o ambiente, dependendo das habilidades da equipe e da infraestrutura de rede existentes e planejadas. Os testes de NetApp geralmente mostraram pouca diferença entre os protocolos executados em velocidades de linha semelhantes, por isso é melhor se concentrar na infraestrutura de rede e nos recursos da equipe sobre o desempenho bruto do protocolo.

Os seguintes fatores podem ser úteis para considerar uma escolha de protocolo:

- * Ambiente atual do cliente.* Embora as equipes DE TI geralmente sejam qualificadas para gerenciar a infraestrutura Ethernet IP, nem todas elas são qualificadas para gerenciar uma malha FC SAN. No entanto, usar uma rede IP de uso geral que não foi projetada para o tráfego de armazenamento pode não funcionar bem. Considere a infraestrutura de rede que você tem em vigor, quaisquer melhorias planejadas e as habilidades e disponibilidade da equipe para gerenciá-los.
- * Facilidade de configuração.* Além da configuração inicial da malha FC (switches e cabeamento adicionais, zoneamento e verificação de interoperabilidade de HBA e firmware), os protocolos de bloco também exigem criação e mapeamento de LUNs e descoberta e formatação pelo SO convidado. Depois que os volumes NFS são criados e exportados, eles são montados pelo host ESXi e prontos para uso. O NFS não tem nenhuma qualificação especial de hardware ou firmware para gerenciar.
- * Facilidade de gestão.* Com os protocolos SAN, se mais espaço for necessário, várias etapas são necessárias, incluindo o crescimento de um LUN, a digitalização novamente para descobrir o novo tamanho e, em seguida, o crescimento do sistema de arquivos). Embora seja possível aumentar um LUN, reduzir o tamanho de um LUN não é, e recuperar espaço não utilizado pode exigir esforço adicional. O

NFS permite um dimensionamento fácil para cima ou para baixo, e esse redimensionamento pode ser automatizado pelo sistema de armazenamento. A SAN oferece recuperação de espaço através de comandos DE TRIM/UNMAP do sistema operacional convidado, permitindo que o espaço dos arquivos excluídos seja retornado ao array. Esse tipo de recuperação de espaço é mais difícil com armazenamentos de dados NFS.

- **Transparência do espaço de armazenamento.** A utilização do storage geralmente é mais fácil de ver em ambientes NFS porque o thin Provisioning devolve economia imediatamente. Da mesma forma, a economia de deduplicação e clonagem ficam imediatamente disponíveis para outras VMs no mesmo armazenamento de dados ou para outros volumes do sistema de storage. Normalmente, a densidade da VM também é maior em um armazenamento de dados NFS, o que pode melhorar a economia de deduplicação e reduzir os custos de gerenciamento com menos armazenamentos de dados para gerenciar.

Layout do datastore

Os sistemas de storage ONTAP oferecem grande flexibilidade na criação de datastores para VMs e discos virtuais. Embora muitas práticas recomendadas do ONTAP sejam aplicadas ao usar o VSC para provisionar armazenamentos de dados para o vSphere (listadas na "[Host ESXi recomendado e outras configurações do ONTAP](#)" seção), aqui estão algumas diretrizes adicionais a serem consideradas:

- A implantação do vSphere com armazenamentos de dados NFS do ONTAP resulta em uma implementação de alta performance e fácil de gerenciar que fornece taxas de VM para armazenamento de dados que não podem ser obtidas com protocolos de storage baseados em bloco. Essa arquitetura pode resultar em um aumento de dez vezes na densidade do datastore com uma redução correlacionada no número de datastores. Embora um datastore maior possa beneficiar a eficiência de storage e fornecer benefícios operacionais, considere o uso de pelo menos quatro datastores (volumes FlexVol) para armazenar suas VMs em uma única controladora ONTAP para obter o máximo de performance com os recursos de hardware. Essa abordagem também permite que você estabeleça datastores com diferentes políticas de recuperação. Alguns podem ser copiados ou replicados com mais frequência do que outros com base nas necessidades empresariais. Vários armazenamentos de dados não são necessários com o FlexGroup volumes para obter performance, pois são dimensionados de acordo com o design.
- **A NetApp recomenda** o uso de volumes FlexVol para a maioria dos armazenamentos de dados NFS. A partir do ONTAP 9.8, os volumes FlexGroup também são compatíveis para uso como datastores, e geralmente são recomendados para certos casos de uso. Outros contentores de armazenamento do ONTAP, como qtrees, geralmente não são recomendados, porque eles não são atualmente suportados pelas ferramentas do ONTAP para VMware vSphere ou pelo plug-in do NetApp SnapCenter para VMware vSphere. Dito isso, a implantação de datastores como vários qtrees em um único volume pode ser útil para ambientes altamente automatizados que podem se beneficiar de cotas em nível de datastore ou clones de arquivos VM.
- Um bom tamanho para um datastore FlexVol volume é de cerca de 4TB a 8TB. Esse tamanho é um bom ponto de equilíbrio para performance, facilidade de gerenciamento e proteção de dados. Comece pequeno (digamos, 4TB) e cresça o datastore conforme necessário (até o máximo de 100TB). Armazenamentos de dados menores são mais rápidos para se recuperar do backup ou após um desastre e podem ser movidos rapidamente pelo cluster. Considere o uso do dimensionamento automático do ONTAP para aumentar e diminuir automaticamente o volume conforme o espaço usado muda. As ferramentas do ONTAP para o Assistente de provisionamento de datastore do VMware vSphere usam o dimensionamento automático por padrão para novos datastores. A personalização adicional dos limites de crescimento e redução e o tamanho máximo e mínimo podem ser feitos com o System Manager ou com a linha de comando.
- Como alternativa, armazenamentos de dados VMFS podem ser configurados com LUNs acessados pelo FC e iSCSI. O VMFS permite que os LUNs sejam acessados simultaneamente por cada servidor ESX em um cluster. Os armazenamentos de dados VMFS podem ter até 64TB TB de tamanho e consistem em até 32 2TB LUNs (VMFS 3) ou um único LUN 64TB (VMFS 5). O tamanho máximo de LUN do ONTAP é de 128TB GB em sistemas AFF, ASA e FAS. O NetApp geralmente recomenda o uso de um único LUN

grande para cada datastore. Assim como o NFS, considere o uso de vários armazenamentos de dados (volumes) para maximizar a performance em uma única controladora ONTAP.

- Os sistemas operacionais Guest (SO) mais antigos precisavam de alinhamento com o sistema de storage para obter o melhor desempenho e eficiência de storage. No entanto, os sistemas operacionais modernos suportados por fornecedores de distribuidores Microsoft e Linux, como a Red Hat, não precisam mais de ajustes para alinhar a partição do sistema de arquivos com os blocos do sistema de armazenamento subjacente em um ambiente virtual. Se você estiver usando um sistema operacional antigo que pode exigir alinhamento, procure na base de conhecimento de suporte da NetApp artigos usando "alinhamento de VM" ou solicite uma cópia do TR-3747 de um Contato de vendas ou parceiro da NetApp.
- Evite o uso de utilitários de desfragmentação no sistema operacional convidado, pois isso não oferece nenhum benefício de desempenho e afeta a eficiência de armazenamento e o uso de espaço instantâneo. Considere também desativar a indexação de pesquisa no SO convidado para desktops virtuais.
- A ONTAP liderou o setor com recursos de eficiência de storage inovadores, permitindo que você aproveite ao máximo seu espaço em disco utilizável. Os sistemas AFF levam essa eficiência ainda mais longe com a deduplicação e a compactação in-line padrão. Os dados são deduplicados em todos os volumes de um agregado. Portanto, você não precisa mais agrupar sistemas operacionais semelhantes e aplicativos semelhantes em um único datastore para maximizar a economia.
- Em alguns casos, talvez você nem precise de um datastore. Para obter o melhor desempenho e capacidade de gerenciamento, evite usar um datastore para aplicativos de e/S de alta qualidade, como bancos de dados e alguns aplicativos. Em vez disso, considere sistemas de arquivos de propriedade de convidados, como sistemas de arquivos NFS ou iSCSI gerenciados pelo convidado ou com RDMS. Para obter orientações específicas sobre aplicações, consulte relatórios técnicos da NetApp para a sua aplicação. Por exemplo, "[Bancos de dados Oracle no ONTAP](#)" tem uma seção sobre virtualização com detalhes úteis.
- Os discos de primeira classe (ou discos virtuais aprimorados) permitem discos gerenciados pelo vCenter, independentemente de uma VM com o vSphere 6,5 e posterior. Embora gerenciados principalmente pela API, eles podem ser úteis com o vVols, especialmente quando gerenciados por ferramentas OpenStack ou Kubernetes. Eles são suportados pelo ONTAP, bem como pelas ferramentas do ONTAP para VMware vSphere.

Migração de datastore e VM

Ao migrar VMs de um datastore existente em outro sistema de storage para o ONTAP, veja algumas práticas a serem lembradas:

- Use o Storage vMotion para mover o volume de suas máquinas virtuais para o ONTAP. Essa abordagem não só não causa interrupções às VMs em execução, como também permite que recursos de eficiência de storage da ONTAP, como deduplicação e compactação, processem os dados à medida que migram. Considere usar os recursos do vCenter para selecionar várias VMs da lista de inventário e, em seguida, agendar a migração (use a tecla Ctrl enquanto clica em ações) em um momento apropriado.
- Embora você possa Planejar cuidadosamente uma migração para armazenamentos de dados de destino apropriados, muitas vezes é mais simples migrar em massa e depois organizar mais tarde, conforme necessário. Talvez você queira usar essa abordagem para orientar sua migração para diferentes armazenamentos de dados se tiver necessidades específicas de proteção de dados, como diferentes programações do Snapshot.
- A maioria das VMs e seu storage podem ser migrados durante a execução (quente), mas a migração de storage conectado (não no datastore), como ISOs, LUNs ou volumes NFS de outro sistema de storage, pode exigir migração sem interrupção.
- As máquinas virtuais que precisam de uma migração mais cuidadosa incluem bancos de dados e aplicativos que usam armazenamento anexado. Em geral, considere o uso das ferramentas do aplicativo para gerenciar a migração. Para Oracle, considere usar ferramentas Oracle como RMAN ou ASM para

migrar os arquivos do banco de dados. Consulte "[Migração de bancos de dados Oracle para sistemas de storage ONTAP](#)" para obter mais informações. Da mesma forma, para o SQL Server, considere usar ferramentas do SQL Server Management Studio ou do NetApp, como o SnapManager para SQL Server ou SnapCenter.

Ferramentas do ONTAP para VMware vSphere

A prática recomendada mais importante ao usar o vSphere com sistemas executando o ONTAP é instalar e usar as ferramentas do ONTAP para o plug-in do VMware vSphere (anteriormente conhecido como console de armazenamento virtual). Esse plug-in do vCenter simplifica o gerenciamento de storage, aumenta a disponibilidade e reduz os custos de storage e a sobrecarga operacional, seja usando SAN ou nas. Ele usa as práticas recomendadas para provisionar armazenamentos de dados e otimiza as configurações do host ESXi para timeouts multipath e HBA (descritos no Apêndice B). Como é um plug-in do vCenter, ele está disponível para todos os clientes da Web vSphere que se conectam ao servidor vCenter.

O plug-in também ajuda a usar outras ferramentas do ONTAP em ambientes vSphere. Ele permite instalar o plug-in NFS para VMware VAAI, que permite descarga de cópia para o ONTAP para operações de clonagem de VM, reserva de espaço para arquivos de disco virtual espessos e descarga de snapshot ONTAP.

O plug-in também é a interface de gerenciamento para muitas funções do provedor VASA para ONTAP, oferecendo suporte ao gerenciamento baseado em políticas de storage com vVols. Depois que as ferramentas do ONTAP para VMware vSphere forem registradas, use-as para criar perfis de capacidade de storage, mapeá-las para o storage e garantir a conformidade do datastore com os perfis ao longo do tempo. O provedor VASA também fornece uma interface para criar e gerenciar datastores da VVol.

Em geral, **a NetApp recomenda** o uso das ferramentas do ONTAP para a interface do VMware vSphere no vCenter para provisionar armazenamentos de dados tradicionais e vVols para garantir que as práticas recomendadas sejam seguidas.

Rede geral

A configuração das configurações de rede ao usar o vSphere com sistemas executando o ONTAP é simples e semelhante a outras configurações de rede. Aqui estão algumas coisas a considerar:

- Separe o tráfego de rede de armazenamento de outras redes. Uma rede separada pode ser obtida usando uma VLAN dedicada ou switches separados para armazenamento. Se a rede de armazenamento partilhar caminhos físicos, como uplinks, poderá necessitar de portas de QoS ou uplink adicionais para garantir uma largura de banda suficiente. Não conecte os hosts diretamente ao storage; use os switches para ter caminhos redundantes e permitir que o VMware HA funcione sem intervenção. "[Ligação direta em rede](#)" Consulte para obter informações adicionais.
- Os frames grandes podem ser usados se desejado e suportados pela sua rede, especialmente ao usar iSCSI. Se forem usados, certifique-se de que estejam configurados de forma idêntica em todos os dispositivos de rede, VLANs e assim por diante no caminho entre o armazenamento e o host ESXi. Caso contrário, você pode ver problemas de desempenho ou conexão. A MTU também deve ser definida de forma idêntica no switch virtual ESXi, na porta VMkernel e também nas portas físicas ou grupos de interface de cada nó ONTAP.
- O NetApp recomenda apenas desativar o controle de fluxo de rede nas portas de rede do cluster dentro de um cluster ONTAP. O NetApp não faz outras recomendações sobre as práticas recomendadas para as portas de rede restantes usadas para tráfego de dados. Você deve ativar ou desativar conforme necessário. "[TR-4182](#)" Consulte para obter mais informações sobre o controle de fluxo.
- Quando os storages ESXi e ONTAP estão conectados a redes de armazenamento Ethernet, **a NetApp recomenda** configurar as portas Ethernet às quais esses sistemas se conectam como portas de borda de protocolo de árvore de expansão rápida (RSTP) ou usando o recurso Cisco PortFast. **A NetApp recomenda** ativar o recurso de tronco de porta de árvore de expansão rápida em ambientes que usam o

recurso Cisco PortFast e que têm entroncamento de VLAN 802,1Q habilitado para o servidor ESXi ou os storages ONTAP.

- **A NetApp recomenda** as seguintes práticas recomendadas para agregação de links:
 - Use switches que suportam agregação de links de portas em dois chassis de switch separados usando uma abordagem de grupo de agregação de links de vários gabinetes, como o Virtual PortChannel (VPC) da Cisco.
 - Desative o LACP para portas de switch conetadas ao ESXi a menos que você esteja usando dvSwitches 5,1 ou posterior com o LACP configurado.
 - Use o LACP para criar agregados de link para sistemas de storage ONTAP com grupos de interface multimodo dinâmico com hash de porta ou IP. ["Gerenciamento de rede"](#) Consulte para obter mais orientações.
 - Use uma política de agrupamento de hash IP no ESXi ao usar agregação de link estático (por exemplo, EtherChannel) e vSwitches padrão ou agregação de link baseada em LACP com switches distribuídos vSphere. Se a agregação de links não for usada, use "Rota baseada no ID de porta virtual de origem".

A tabela a seguir fornece um resumo dos itens de configuração de rede e indica onde as configurações são aplicadas.

Item	ESXi	Interrutor	Nó	SVM
Endereço IP	VMkernel	4	4	Sim
Agregação de links	Switch virtual	Sim	Sim	3
VLAN	Grupos de portas VMkernel e VM	Sim	Sim	3
Controle de fluxo	NIC	Sim	Sim	3
Spanning tree	Não	Sim	Não	Não
MTU (para quadros jumbo)	Switch virtual e porta VMkernel (9000)	Sim (definido para máx.)	Sim (9000)	3
Grupos de failover	Não	Não	Sim (criar)	Sim (selecione)

3 os LIFs SVM se conetam a portas, grupos de interfaces ou interfaces VLAN que têm VLAN, MTU e outras configurações. No entanto, as configurações não são gerenciadas no nível da SVM.

4 esses dispositivos têm endereços IP próprios para gerenciamento, mas esses endereços não são usados no contexto da rede de armazenamento ESXi.

SAN (FC, FCoE, NVMe/FC, iSCSI), RDM

O ONTAP fornece storage de bloco de classe empresarial para o VMware vSphere usando iSCSI, Fibre Channel Protocol (FCP ou FC em resumo) e NVMe over Fabrics (NVMe-of). Veja a seguir as práticas recomendadas para a implementação de protocolos de bloco para storage de VM com o vSphere e o ONTAP.

No vSphere, há três maneiras de usar LUNs de armazenamento em bloco:

- Com datastores VMFS

- Com mapeamento de dispositivo bruto (RDM)
- Como um LUN acessado e controlado por um iniciador de software de um sistema operacional convidado VM

O VMFS é um sistema de arquivos em cluster de alto desempenho que fornece datastores que são pools de armazenamento compartilhado. Armazenamentos de dados VMFS podem ser configurados com LUNs acessados usando FC, iSCSI, FCoE ou namespaces NVMe acessados usando os protocolos NVMe/FC ou NVMe/TCP. O VMFS permite que o armazenamento seja acessado simultaneamente por cada servidor ESX em um cluster. O tamanho máximo de LUN é geralmente 128TB, começando com ONTAP 9.12.1P2 (e anterior com sistemas ASA); portanto, um datastore VMFS 5 ou 6 de tamanho máximo de 64TB pode ser criado usando um único LUN.

O vSphere inclui suporte integrado para vários caminhos para dispositivos de armazenamento, conhecido como NMP (NMP). O NMP pode detectar o tipo de storage para sistemas de storage compatíveis e configurar automaticamente a pilha NMP para suportar os recursos do sistema de storage em uso.

Tanto o NMP quanto o ONTAP suportam o Acesso lógico assimétrico (ALUA) para negociar caminhos otimizados e não otimizados. No ONTAP, um caminho otimizado para ALUA segue um caminho de dados direto, usando uma porta de destino no nó que hospeda o LUN que está sendo acessado. O ALUA é ativado por padrão no vSphere e no ONTAP. O NMP reconhece o cluster ONTAP como ALUA e usa o plug-in tipo array de armazenamento ALUA (`VMW_SATP_ALUA`) e seleciona o plug-in de seleção de caminho round robin (`VMW_PSP_RR`).

O ESXi 6 suporta até 256 LUNs e até 1.024 caminhos totais para LUNs. O ESXi não vê LUNs ou caminhos além desses limites. Assumindo o número máximo de LUNs, o limite de caminho permite quatro caminhos por LUN. Em um cluster ONTAP maior, é possível alcançar o limite de caminho antes do limite de LUN. Para lidar com essa limitação, o ONTAP oferece suporte ao mapa de LUN seletivo (SLM) na versão 8,3 e posterior.

O SLM limita os nós que anunciam caminhos para um determinado LUN. É uma prática recomendada do NetApp ter pelo menos um LIF por nó por SVM e usar o SLM para limitar os caminhos anunciados para o nó que hospeda o LUN e seu parceiro de HA. Embora existam outros caminhos, eles não são anunciados por padrão. É possível modificar os caminhos anunciados com os argumentos adicionar e remover nó de relatório dentro do SLM. Observe que os LUNs criados em versões anteriores a 8,3 anunciam todos os caminhos e precisam ser modificados para anunciar apenas os caminhos para o par de HA de hospedagem. Para obter mais informações sobre o SLM, consulte a seção 5,9 do "[TR-4080](#)". O método anterior de portsets também pode ser usado para reduzir ainda mais os caminhos disponíveis para um LUN. Os Portsets ajudam reduzindo o número de caminhos visíveis através dos quais os iniciadores em um iggroup podem ver LUNs.

- O SLM está ativado por predefinição. A menos que você esteja usando portsets, nenhuma configuração adicional é necessária.
- Para LUNs criadas antes do Data ONTAP 8.3, aplique manualmente o SLM executando o `lun mapping remove-reporting-nodes` comando para remover os nós de relatórios de LUN e restringir o acesso LUN ao nó proprietário de LUN e ao seu parceiro de HA.

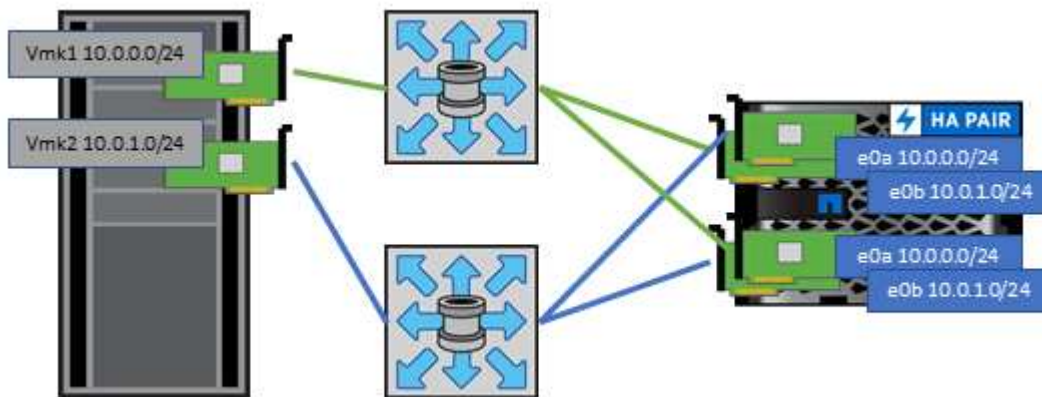
Os protocolos de bloqueio (iSCSI, FC e FCoE) acessam LUNs usando IDs de LUN e números de série, juntamente com nomes exclusivos. FC e FCoE usam nomes mundiais (WWNNs e WWPNS) e iSCSI usa nomes qualificados iSCSI (IQNs). O caminho para LUNs dentro do armazenamento não tem sentido para os protocolos de bloco e não é apresentado em nenhum lugar do protocolo. Portanto, um volume que contém apenas LUNs não precisa ser montado internamente, e um caminho de junção não é necessário para volumes que contêm LUNs usados em datastores. O subsistema NVMe no ONTAP funciona da mesma forma.

Outras práticas recomendadas a considerar:

- Certifique-se de que é criada uma interface lógica (LIF) para cada SVM em cada nó no cluster do ONTAP

para obter disponibilidade e mobilidade máximas. A prática recomendada de SAN ONTAP é usar duas portas físicas e LIFs por nó, uma para cada malha. O ALUA é usado para analisar caminhos e identificar caminhos otimizados ativos (diretos) versus caminhos não otimizados ativos. O ALUA é usado para FC, FCoE e iSCSI.

- Para redes iSCSI, use várias interfaces de rede VMkernel em sub-redes de rede diferentes com agrupamento NIC quando vários switches virtuais estiverem presentes. Você também pode usar várias NICs físicas conectadas a vários switches físicos para fornecer HA e maior taxa de transferência. A figura a seguir fornece um exemplo de conectividade multipath. No ONTAP, configure um grupo de interface de modo único para failover com dois ou mais links conectados a dois ou mais switches ou use LACP ou outra tecnologia de agregação de links com grupos de interface multimodo para fornecer HA e os benefícios da agregação de links.
- Se o CHAP (Challenge-Handshake Authentication Protocol) for usado no ESXi para autenticação de destino, ele também deve ser configurado no ONTAP usando a CLI (`vserver iscsi security create`) ou com o Gerenciador de sistema (edite a segurança do iniciador em armazenamento > SVMs > Configurações da SVM > Protocolos > iSCSI).
- Use as ferramentas do ONTAP para o VMware vSphere para criar e gerenciar LUNs e grupos de pessoas. O plug-in determina automaticamente as WWPNs de servidores e cria grupos apropriados. Ele também configura LUNs de acordo com as melhores práticas e os mapeia para os grupos corretos.
- Use RDMS com cuidado porque eles podem ser mais difíceis de gerenciar e também usam caminhos, que são limitados como descrito anteriormente. Os LUNs ONTAP suportam ambos "modo de compatibilidade física e virtual" os RDMS.
- Para saber mais sobre como usar o NVMe/FC com o vSphere 7,0, consulte este "Guia de configuração de host ONTAP NVMe/FC" e "TR-4684". a figura a seguir mostra a conectividade multipath de um host vSphere para um LUN ONTAP.



NFS

O ONTAP é, entre muitas outras coisas, um array nas com escalabilidade horizontal de classe empresarial. O ONTAP capacita o VMware vSphere com acesso simultâneo a datastores conectados a NFS de muitos hosts ESXi, excedendo muito os limites impostos aos sistemas de arquivos VMFS. O uso do NFS com o vSphere oferece alguns benefícios de visibilidade da eficiência de storage e facilidade de uso, como mencionado na "armazenamentos de dados" seção.

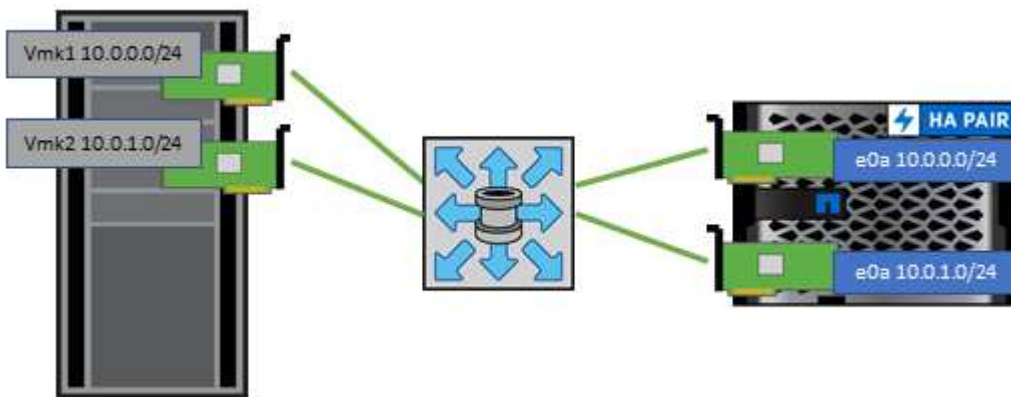
As práticas recomendadas a seguir são recomendadas ao usar o ONTAP NFS com vSphere:

- Use uma única interface lógica (LIF) para cada SVM em cada nó no cluster do ONTAP. As recomendações

anteriores de um LIF por datastore não são mais necessárias. Embora o acesso direto (LIF e datastore no mesmo nó) seja o melhor, não se preocupe com o acesso indireto porque o efeito de desempenho é geralmente mínimo (microsegundos).

- A VMware tem suporte para NFSv3 desde o VMware Infrastructure 3. O vSphere 6,0 adicionou suporte para NFSv4,1, o que permite alguns recursos avançados, como a segurança Kerberos. Onde o NFSv3 usa o bloqueio do lado do cliente, o NFSv4,1 usa o bloqueio do lado do servidor. Embora um volume ONTAP possa ser exportado através de ambos os protocolos, o ESXi só pode ser montado através de um protocolo. Esta montagem de protocolo único não impede que outros hosts ESXi montem o mesmo datastore através de uma versão diferente. Certifique-se de especificar a versão do protocolo a ser usada durante a montagem para que todos os hosts usem a mesma versão e, portanto, o mesmo estilo de bloqueio. Não misture versões de NFS entre hosts. Se possível, use perfis de host para verificar a conformidade.
 - Como não há conversão automática de datastore entre NFSv3 e NFSv4,1, crie um novo datastore NFSv4,1 e use o Storage vMotion para migrar VMs para o novo datastore.
 - Consulte as notas da tabela de interoperabilidade NFS v4,1 no "[Ferramenta Matriz de interoperabilidade NetApp](#)" para obter os níveis de patch específicos do ESXi necessários para suporte.
 - A VMware oferece suporte ao nconnect com o NFSv3 a partir do vSphere 8.0U2. Mais informações sobre o nconnect podem ser encontradas em "[NFSv3 nLigue o recurso ao NetApp e VMware](#)"
- As políticas de exportação de NFS são usadas para controlar o acesso pelos hosts vSphere. Você pode usar uma diretiva com vários volumes (datastores). Com o NFSv3, o ESXi usa o estilo de segurança sys (UNIX) e requer a opção de montagem raiz para executar VMs. No ONTAP, essa opção é chamada de superusuário e, quando a opção superusuário é usada, não é necessário especificar o ID de usuário anônimo. Observe que regras de política de exportação com valores diferentes `-anon` e `-allow-suid` podem causar problemas de descoberta de SVM com as ferramentas do ONTAP. Aqui está uma política de exemplo:
 - Protocolo de acesso: `nfs` (que inclui `nfs3` e `nfs4`)
 - Especificação correspondência Cliente: 192.168.42.21
 - Regra de Acesso RO: `Sys`
 - Regra de Acesso RW: `Sys`
 - UID anônimo
 - Superusuário: `Sys`
- Se o plug-in NFS NetApp for usado, o protocolo deve ser definido como `nfs` em vez de `nfs3` quando a regra de política de exportação for criada ou modificada. O recurso de descarga de cópia VAAI requer que o protocolo NFSv4 funcione, mesmo que o protocolo de dados seja NFSv3. Especificar o protocolo como `nfs` inclui as versões NFSv3 e NFSv4.
- Os volumes do armazenamento de dados NFS são juntados do volume raiz do SVM. Portanto, o ESXi também precisa ter acesso ao volume raiz para navegar e montar volumes do armazenamento de dados. A política de exportação para o volume raiz e para quaisquer outros volumes em que a junção do volume do datastore esteja aninhada, deve incluir uma regra ou regras para os servidores ESXi concedendo acesso somente leitura. Aqui está uma política de exemplo para o volume raiz, também usando o plug-in VAAI:
 - Protocolo de acesso: `nfs` (que inclui `nfs3` e `nfs4`)
 - Especificação correspondência Cliente: 192.168.42.21
 - Regra de Acesso RO: `Sys`
 - Regra de acesso RW: Nunca (melhor segurança para o volume raiz)

- UID anônimo
- Superusuário: Sys (também necessário para o volume raiz com VAAI)
- Use as ferramentas do ONTAP para VMware vSphere (a prática recomendada mais importante):
 - Use as ferramentas do ONTAP para o VMware vSphere para provisionar armazenamentos de dados porque simplifica o gerenciamento de políticas de exportação automaticamente.
 - Ao criar datastores para clusters VMware com o plug-in, selecione o cluster em vez de um único servidor ESX. Essa opção o aciona para montar automaticamente o datastore em todos os hosts do cluster.
 - Use a função de montagem plug-in para aplicar datastores existentes a novos servidores.
 - Quando não estiver usando as ferramentas do ONTAP para VMware vSphere, use uma única política de exportação para todos os servidores ou para cada cluster de servidores onde é necessário controle de acesso adicional.
- Embora o ONTAP ofereça uma estrutura de namespace de volume flexível para organizar volumes em uma árvore usando junções, essa abordagem não tem valor para o vSphere. Ele cria um diretório para cada VM na raiz do datastore, independentemente da hierarquia do namespace do storage. Assim, a prática recomendada é simplesmente montar o caminho de junção para volumes para vSphere no volume raiz do SVM, que é como as ferramentas do ONTAP para VMware vSphere provisionam datastores. Não ter caminhos de junção aninhados também significa que nenhum volume é dependente de qualquer volume que não seja o volume raiz e que tirar um volume off-line ou destruí-lo, mesmo intencionalmente, não afeta o caminho para outros volumes.
- Um tamanho de bloco de 4K é bom para partições NTFS em datastores NFS. A figura a seguir mostra a conectividade de um host vSphere para um datastore ONTAP NFS.



A tabela a seguir lista as versões de NFS e os recursos compatíveis.

Recursos do vSphere	NFSv3	NFSv4.1
VMotion e Storage vMotion	Sim	Sim
Alta disponibilidade	Sim	Sim
Tolerância de falhas	Sim	Sim
DRS	Sim	Sim
Perfis de host	Sim	Sim
Armazenamento DRS	Sim	Não
Controle de e/S de storage	Sim	Não

Recursos do vSphere	NFSv3	NFSv4.1
SRM	Sim	Não
Volumes virtuais	Sim	Não
Aceleração de hardware (VAAI)	Sim	Sim
Autenticação Kerberos	Não	Sim (aprimorado com o vSphere 6,5 e posterior para oferecer suporte a AES, krb5i)
Suporte multipathing	Não	Sim

Volumes FlexGroup

Use o ONTAP e o FlexGroup volumes com o VMware vSphere para armazenamentos de dados simples e dimensionáveis que aproveitam todo o poder de um cluster ONTAP inteiro.

O ONTAP 9.8, juntamente com as ferramentas do ONTAP para VMware vSphere 9,8 e o plug-in SnapCenter para versões do VMware 4,4, adicionou suporte para datastores com suporte em volume do FlexGroup no vSphere. O FlexGroup volumes simplifica a criação de grandes armazenamentos de dados e cria automaticamente os volumes constituintes distribuídos necessários no cluster do ONTAP para obter o máximo de desempenho a partir de um sistema ONTAP.

Use o FlexGroup volumes com vSphere se você precisar de um único armazenamento de dados dimensionável vSphere com o poder de um cluster completo do ONTAP ou se tiver workloads de clonagem muito grandes que podem se beneficiar do novo mecanismo de clonagem do FlexGroup.

Descarga de cópia

Além de testes extensivos do sistema com cargas de trabalho do vSphere, o ONTAP 9.8 adicionou um novo mecanismo de descarga de cópia para datastores FlexGroup. Este novo sistema usa um mecanismo de cópia melhorado para replicar arquivos entre constituintes em segundo plano, permitindo o acesso à origem e ao destino. Esse cache local é então usado para instanciar rapidamente clones de VM sob demanda.

Para ativar a descarga de cópia otimizada do FlexGroup, consulte ["Como configurar FlexGroups do ONTAP para permitir a descarga de cópia do VAAI"](#)

É possível que, se você usar a clonagem do VAAI, mas não clonar o suficiente para manter o cache aquecido, seus clones podem não ser mais rápidos do que uma cópia baseada em host. Se for esse o caso, você pode ajustar o tempo limite do cache para melhor atender às suas necessidades.

Considere o seguinte cenário:

- Você criou um novo FlexGroup com 8 constituintes
- O tempo limite do cache para o novo FlexGroup é definido para 160 minutos

Nesse cenário, os primeiros 8 clones a serem concluídos serão cópias completas, não clones de arquivos locais. Qualquer clonagem adicional dessa VM antes que o tempo limite de 160 segundos expire usará o mecanismo de clone de arquivo dentro de cada constituinte de uma forma round-robin para criar cópias quase imediatas distribuídas uniformemente pelos volumes constituintes.

Cada novo trabalho clone que um volume recebe repõe o tempo limite. Se um volume constituinte no exemplo

FlexGroup não receber uma solicitação de clone antes do tempo limite, o cache para essa VM específica será limpo e o volume precisará ser preenchido novamente. Além disso, se a origem do clone original mudar (por exemplo, você atualizou o modelo), então o cache local em cada componente será invalidado para evitar qualquer conflito. Como dito anteriormente, o cache é ajustável e pode ser configurado para corresponder às necessidades do seu ambiente.

Para obter mais informações sobre como usar FlexGroups com VAAI, consulte este artigo da KB: "[VAAI: Como o armazenamento em cache funciona com o FlexGroup volumes?](#)"

Em ambientes onde você não consegue aproveitar ao máximo o cache FlexGroup, mas ainda exige clonagem rápida entre volumes, considere o uso de vVols. A clonagem entre volumes com vVols é muito mais rápida do que usar datastores tradicionais e não depende de um cache.

Definições de QoS

A configuração do QoS no nível FlexGroup usando o Gerenciador de sistema do ONTAP ou o shell do cluster é suportada, no entanto, não fornece reconhecimento de VM ou integração do vCenter.

O QoS (IOPS máx/min) pode ser definido em VMs individuais ou em todas as VMs em um datastore na IU do vCenter ou por meio de APIs REST usando ferramentas do ONTAP. A configuração de QoS em todas as VMs substitui todas as configurações separadas por VM. As configurações não se estendem a VMs novas ou migradas no futuro; defina a QoS nas novas VMs ou recrie QoS a todas as VMs no datastore.

Observe que o VMware vSphere trata todas as IO para um datastore NFS como uma única fila por host, e a regulação da QoS em uma VM pode afetar a performance de outras VMs no mesmo datastore. Isso está em contraste com o vVols, que pode manter suas configurações de política de QoS se eles migrarem para outro datastore e não afetarem o IO de outras VMs quando forem controlados.

Métricas

O ONTAP 9.8 também adicionou novas métricas de performance baseadas em arquivos (IOPS, taxa de transferência e latência) para o FlexGroup Files, e essas métricas podem ser visualizadas nas ferramentas do ONTAP para os relatórios de VM e painel do VMware vSphere. O plug-in das ferramentas do ONTAP para VMware vSphere também permite que você defina regras de qualidade do serviço (QoS) usando uma combinação de IOPS máximo e/ou mínimo. Eles podem ser configurados em todas as VMs em um datastore ou individualmente para VMs específicas.

Práticas recomendadas

- Use as ferramentas do ONTAP para criar datastores do FlexGroup para garantir que o FlexGroup seja criado de forma otimizada e que as políticas de exportação sejam configuradas para corresponder ao seu ambiente vSphere. No entanto, depois de criar o volume FlexGroup com as ferramentas do ONTAP, você descobrirá que todos os nós do cluster do vSphere estão usando um único endereço IP para montar o datastore. Isso pode resultar em um gargalo na porta de rede. Para evitar esse problema, desmonte o datastore e remonte-o usando o assistente padrão do vSphere datastore usando um nome DNS de round-robin que carrega o balanceamento entre LIFs no SVM. Depois de remontar, as ferramentas do ONTAP poderão gerenciar o datastore novamente. Se as ferramentas do ONTAP não estiverem disponíveis, use os padrões do FlexGroup e crie sua política de exportação seguindo as diretrizes do "[Datastores e protocolos - NFS](#)".
- Ao dimensionar um armazenamento de dados do FlexGroup, lembre-se de que o FlexGroup consiste em vários volumes FlexVol menores que criam um namespace maior. Como tal, dimensione o datastore para ter pelo menos 8x (assumindo os componentes 8 padrão) o tamanho do seu maior arquivo VMDK mais 10-20% de espaço livre não utilizado para permitir flexibilidade no rebalanceamento. Por exemplo, se você tiver um VMDK de 6TB TB no seu ambiente, dimensione o datastore do FlexGroup não menor que 52,8TB TB (6x8 a 10%).

- VMware e NetApp suportam entroncamento de sessão NFSv4,1 começando com ONTAP 9.14,1. Consulte as notas da matriz de interoperabilidade NetApp NFS 4,1 para obter detalhes específicos da versão. O NFSv3 não oferece suporte a vários caminhos físicos para um volume, mas oferece suporte ao nconnect a partir do vSphere 8.0U2. Mais informações sobre o nconnect podem ser encontradas no ["NFSv3 nLigue o recurso ao NetApp e VMware"](#).
- Use o plug-in NFS para VMware VAAI para descarga de cópia. Observe que, embora a clonagem seja aprimorada em um datastore FlexGroup, como mencionado anteriormente, o ONTAP não oferece vantagens significativas de desempenho em comparação com a cópia do host ESXi ao copiar VMs entre volumes FlexVol e/ou FlexGroup. Portanto, considere suas cargas de trabalho de clonagem ao decidir usar VAAI ou FlexGroups. Modificar o número de volumes constituintes é uma forma de otimizar a clonagem baseada em FlexGroup. Como está ajustando o tempo limite do cache mencionado anteriormente.
- Use as ferramentas do ONTAP para VMware vSphere 9,8 ou posterior para monitorar a performance das VMs FlexGroup usando métricas do ONTAP (painel e relatórios de VM) e gerenciar a QoS em VMs individuais. Essas métricas não estão disponíveis atualmente por meio de comandos ou APIs do ONTAP.
- O plug-in do SnapCenter para VMware vSphere versão 4,4 e posterior oferece suporte ao backup e à recuperação de VMs em um datastore FlexGroup no sistema de storage primário. A VCS 4,6 adiciona suporte ao SnapMirror para datastores baseados em FlexGroup. Usar snapshots e replicação baseados em array é a maneira mais eficiente de proteger seus dados.

Configuração de rede

A configuração das configurações de rede ao usar o vSphere com sistemas executando o ONTAP é simples e semelhante a outras configurações de rede.

Aqui estão algumas coisas a considerar:

- Separe o tráfego de rede de armazenamento de outras redes. Uma rede separada pode ser obtida usando uma VLAN dedicada ou switches separados para armazenamento. Se a rede de armazenamento partilhar caminhos físicos, como uplinks, poderá necessitar de portas de QoS ou uplink adicionais para garantir uma largura de banda suficiente. Não conecte os hosts diretamente ao storage; use os switches para ter caminhos redundantes e permitir que o VMware HA funcione sem intervenção. ["Ligação direta em rede"](#) Consulte para obter informações adicionais.
- Os frames grandes podem ser usados se desejado e suportados pela sua rede, especialmente ao usar iSCSI. Se forem usados, certifique-se de que estejam configurados de forma idêntica em todos os dispositivos de rede, VLANs e assim por diante no caminho entre o armazenamento e o host ESXi. Caso contrário, você pode ver problemas de desempenho ou conexão. A MTU também deve ser definida de forma idêntica no switch virtual ESXi, na porta VMkernel e também nas portas físicas ou grupos de interface de cada nó ONTAP.
- O NetApp recomenda apenas desativar o controle de fluxo de rede nas portas de rede do cluster dentro de um cluster ONTAP. O NetApp não faz outras recomendações sobre as práticas recomendadas para as portas de rede restantes usadas para tráfego de dados. Você deve ativá-lo ou desativá-lo conforme necessário. ["TR-4182"](#) Consulte para obter mais informações sobre o controle de fluxo.
- Quando os storages ESXi e ONTAP estão conectados a redes de armazenamento Ethernet, a NetApp recomenda configurar as portas Ethernet às quais esses sistemas se conectam como portas de borda de protocolo de árvore de expansão rápida (RSTP) ou usando o recurso Cisco PortFast. A NetApp recomenda ativar o recurso de tronco de porta de árvore de expansão rápida em ambientes que usam o recurso Cisco PortFast e que têm entroncamento de VLAN 802,1Q habilitado para o servidor ESXi ou para os storages ONTAP.
- A NetApp recomenda as seguintes práticas recomendadas para agregação de links:

- Use switches que suportam agregação de links de portas em dois chassis de switch separados usando uma abordagem de grupo de agregação de links de vários gabinetes, como o Virtual PortChannel (VPC) da Cisco.
- Desative o LACP para portas de switch conectadas ao ESXi a menos que você esteja usando dvSwitches 5,1 ou posterior com o LACP configurado.
- Use o LACP para criar agregados de link para sistemas de storage ONTAP com grupos de interface multimodo dinâmico com hash IP.
- Use uma política de agrupamento de hash IP no ESXi.

A tabela a seguir fornece um resumo dos itens de configuração de rede e indica onde as configurações são aplicadas.

Item	ESXi	Interrutor	Nó	SVM
Endereço IP	VMkernel	Não**	Não**	Sim
Agregação de links	Switch virtual	Sim	Sim	Não*
VLAN	Grupos de portas VMkernel e VM	Sim	Sim	Não*
Controle de fluxo	NIC	Sim	Sim	Não*
Spanning tree	Não	Sim	Não	Não
MTU (para quadros jumbo)	Switch virtual e porta VMkernel (9000)	Sim (definido para máx.)	Sim (9000)	Não*
Grupos de failover	Não	Não	Sim (criar)	Sim (selecione)

*Os LIFs SVM se conectam a portas, grupos de interfaces ou interfaces VLAN que têm VLAN, MTU e outras configurações. No entanto, as configurações não são gerenciadas no nível da SVM.

**Esses dispositivos têm endereços IP próprios para gerenciamento, mas esses endereços não são usados no contexto da rede de armazenamento ESXi.

SAN (FC, FCoE, NVMe/FC, iSCSI), RDM

No vSphere, há três maneiras de usar LUNs de armazenamento em bloco:

- Com datastores VMFS
- Com mapeamento de dispositivo bruto (RDM)
- Como um LUN acessado e controlado por um iniciador de software de um sistema operacional convidado VM

O VMFS é um sistema de arquivos em cluster de alto desempenho que fornece datastores que são pools de armazenamento compartilhado. Os armazenamentos de dados VMFS podem ser configurados com LUNs acessados por meio de namespaces FC, iSCSI, FCoE ou NVMe acessados pelo protocolo NVMe/FC. O VMFS permite que LUNs tradicionais sejam acessados simultaneamente por cada servidor ESX em um cluster. O tamanho máximo de LUN do ONTAP é geralmente 16TB; portanto, um armazenamento de dados VMFS 5 de tamanho máximo de 64TB GB (veja a primeira tabela nesta seção) é criado usando quatro LUNs 16TB (todos os sistemas de storage SAN suportam o tamanho máximo de LUN VMFS de 64TB GB). Como a arquitetura LUN do ONTAP não tem pequenas profundidades de fila individuais, os armazenamentos de dados VMFS no ONTAP podem ser dimensionados para um grau maior do que com arquiteturas de array tradicionais de uma maneira relativamente simples.

O vSphere inclui suporte integrado para vários caminhos para dispositivos de armazenamento, conhecido como NMP (NMP). O NMP pode detectar o tipo de storage para sistemas de storage compatíveis e configurar automaticamente a pilha NMP para suportar os recursos do sistema de storage em uso.

Tanto o NMP quanto o ONTAP suportam o Acesso lógico assimétrico (ALUA) para negociar caminhos otimizados e não otimizados. No ONTAP, um caminho otimizado para ALUA segue um caminho de dados direto, usando uma porta de destino no nó que hospeda o LUN que está sendo acessado. O ALUA é ativado por padrão no vSphere e no ONTAP. O NMP reconhece o cluster ONTAP como ALUA e usa o plug-in do tipo storage array ALUA (`VMW_SATP_ALUA`) e seleciona o plug-in de seleção de caminho round-robin (`VMW_PSP_RR`).

O ESXi 6 suporta até 256 LUNs e até 1.024 caminhos totais para LUNs. Quaisquer LUNs ou caminhos além desses limites não são vistos pelo ESXi. Assumindo o número máximo de LUNs, o limite de caminho permite quatro caminhos por LUN. Em um cluster ONTAP maior, é possível alcançar o limite de caminho antes do limite de LUN. Para lidar com essa limitação, o ONTAP oferece suporte ao mapa de LUN seletivo (SLM) na versão 8,3 e posterior.

O SLM limita os nós que anunciam caminhos para um determinado LUN. É uma prática recomendada do NetApp ter pelo menos um LIF por nó por SVM e usar o SLM para limitar os caminhos anunciados para o nó que hospeda o LUN e seu parceiro de HA. Embora existam outros caminhos, eles não são anunciados por padrão. É possível modificar os caminhos anunciados com os argumentos adicionar e remover nó de relatório dentro do SLM. Observe que os LUNs criados em versões anteriores ao 8,3 anunciam todos os caminhos e precisam ser modificados para anunciar apenas os caminhos para o par de HA de hospedagem. Para obter mais informações sobre o SLM, consulte a seção 5,9 do "[TR-4080](#)". O método anterior de portsets também pode ser usado para reduzir ainda mais os caminhos disponíveis para um LUN. Os Portsets ajudam reduzindo o número de caminhos visíveis através dos quais os iniciadores em um iggroup podem ver LUNs.

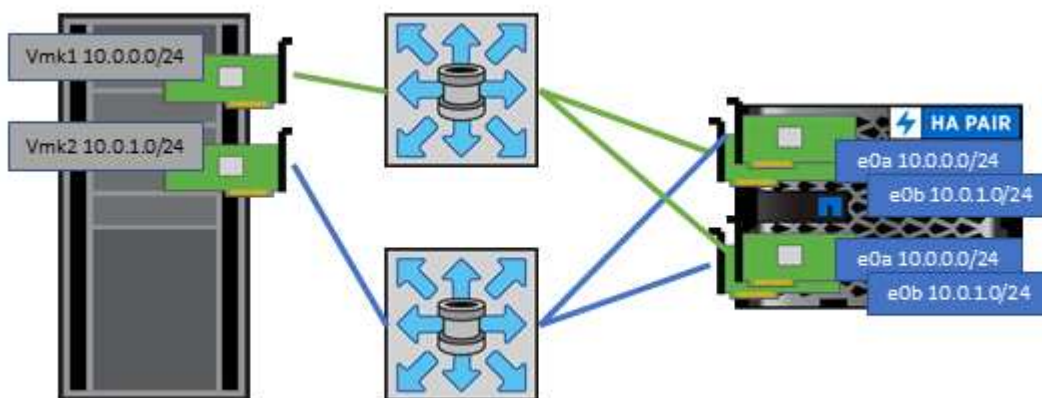
- O SLM está ativado por predefinição. A menos que você esteja usando portsets, nenhuma configuração adicional é necessária.
- Para LUNs criadas antes do Data ONTAP 8.3, aplique manualmente o SLM executando o `lun mapping remove-reporting-nodes` comando para remover os nós de relatórios de LUN e restringir o acesso LUN ao nó proprietário de LUN e ao seu parceiro de HA.

Os protocolos de bloqueio (iSCSI, FC e FCoE) acessam LUNs usando IDs de LUN e números de série, juntamente com nomes exclusivos. FC e FCoE usam nomes mundiais (WWNNs e WWPNs) e iSCSI usa nomes qualificados iSCSI (IQNs). O caminho para LUNs dentro do armazenamento não tem sentido para os protocolos de bloco e não é apresentado em nenhum lugar do protocolo. Portanto, um volume que contém apenas LUNs não precisa ser montado internamente, e um caminho de junção não é necessário para volumes que contêm LUNs usados em datastores. O subsistema NVMe no ONTAP funciona da mesma forma.

Outras práticas recomendadas a considerar:

- Certifique-se de que é criada uma interface lógica (LIF) para cada SVM em cada nó no cluster do ONTAP para obter disponibilidade e mobilidade máximas. A prática recomendada de SAN ONTAP é usar duas portas físicas e LIFs por nó, uma para cada malha. O ALUA é usado para analisar caminhos e identificar caminhos otimizados ativos (diretos) versus caminhos não otimizados ativos. O ALUA é usado para FC, FCoE e iSCSI.
- Para redes iSCSI, use várias interfaces de rede VMkernel em sub-redes de rede diferentes com agrupamento NIC quando vários switches virtuais estiverem presentes. Você também pode usar várias NICs físicas conetadas a vários switches físicos para fornecer HA e maior taxa de transferência. A figura a seguir fornece um exemplo de conectividade multipath. No ONTAP, use um grupo de interface de modo único com vários links para diferentes switches ou LACP com grupos de interface multimodo para alta disponibilidade e benefícios de agregação de links.

- Se o CHAP (Challenge-Handshake Authentication Protocol) for usado no ESXi para autenticação de destino, ele também deve ser configurado no ONTAP usando a CLI (`vserver iscsi security create`) ou com o Gerenciador de sistema (edite a segurança do iniciador em armazenamento > SVMs > Configurações da SVM > Protocolos > iSCSI).
- Use as ferramentas do ONTAP para o VMware vSphere para criar e gerenciar LUNs e grupos de pessoas. O plug-in determina automaticamente as WWPNs de servidores e cria grupos apropriados. Ele também configura LUNs de acordo com as melhores práticas e os mapeia para os grupos corretos.
- Use RDMS com cuidado porque eles podem ser mais difíceis de gerenciar e também usam caminhos, que são limitados como descrito anteriormente. Os LUNs ONTAP suportam ambos "[modo de compatibilidade física e virtual](#)" os RDMS.
- Para saber mais sobre como usar o NVMe/FC com o vSphere 7,0, consulte este "[Guia de configuração de host ONTAP NVMe/FC](#)" e "[TR-4684](#)". A figura a seguir mostra a conectividade multipath de um host vSphere a um LUN ONTAP.



NFS

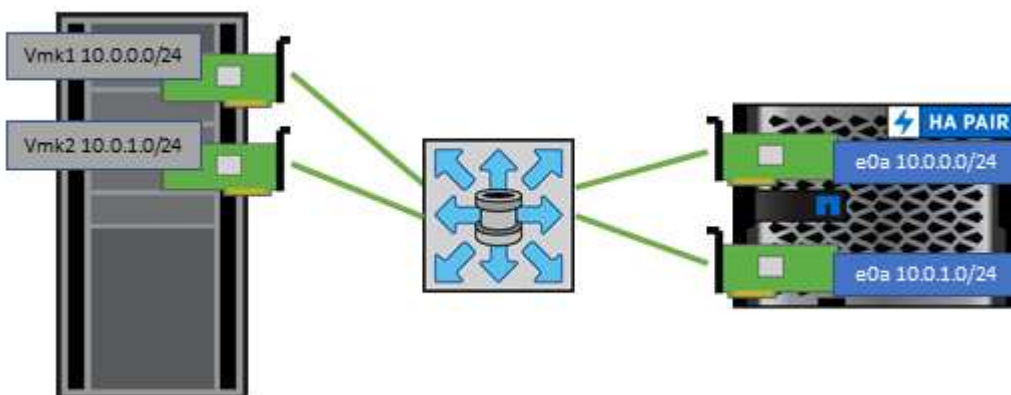
O vSphere permite que os clientes usem arrays NFS de classe empresarial para fornecer acesso simultâneo a datastores para todos os nós em um cluster ESXi. Como mencionado na seção datastore, há alguns benefícios de visibilidade de eficiência de uso e facilidade de uso ao usar o NFS com vSphere.

As práticas recomendadas a seguir são recomendadas ao usar o ONTAP NFS com vSphere:

- Use uma única interface lógica (LIF) para cada SVM em cada nó no cluster do ONTAP. As recomendações anteriores de um LIF por datastore não são mais necessárias. Embora o acesso direto (LIF e datastore no mesmo nó) seja o melhor, não se preocupe com o acesso indireto porque o efeito de desempenho geralmente é mínimo (microsegundos).
- Todas as versões do VMware vSphere com suporte no momento podem usar o NFS v3 e o v4,1. O suporte oficial para nconnect foi adicionado à atualização 2 do vSphere 8,0 para NFS v3. Para o NFS v4,1, o vSphere continua a oferecer suporte ao entroncamento de sessão, autenticação Kerberos e autenticação Kerberos com integridade. É importante notar que o entroncamento de sessão requer o ONTAP 9.14,1 ou uma versão posterior. Você pode saber mais sobre o recurso nconnect e como ele melhora o desempenho em "[NFSv3 nLigue o recurso ao NetApp e VMware](#)".

Vale a pena notar que NFSv3 e NFSv4,1 usam diferentes mecanismos de bloqueio. O NFSv3 usa o bloqueio do lado do cliente, enquanto o NFSv4,1 usa o bloqueio do lado do servidor. Embora um volume ONTAP possa ser exportado através de ambos os protocolos, o ESXi pode montar apenas um datastore através de um protocolo. No entanto, isso não significa que outros hosts ESXi não possam montar o mesmo datastore através de uma versão diferente. Para evitar quaisquer problemas, é essencial especificar a versão do protocolo a ser usada durante a montagem, garantindo que todos os hosts usem a mesma versão e, portanto,

o mesmo estilo de bloqueio. É essencial evitar misturar versões NFS entre hosts. Se possível, use perfis de host para verificar a conformidade. **Como não há conversão automática de datastore entre NFSv3 e NFSv4,1, crie um novo datastore NFSv4,1 e use o Storage vMotion para migrar VMs para o novo datastore.** Por favor, consulte as notas da tabela de interoperabilidade NFS v4,1 no "[Ferramenta Matriz de interoperabilidade NetApp](#)" para obter os níveis de patch ESXi específicos necessários para suporte. * As políticas de exportação NFS são usadas para controlar o acesso pelos hosts vSphere. Você pode usar uma diretiva com vários volumes (datastores). Com o NFSv3, o ESXi usa o estilo de segurança sys (UNIX) e requer a opção de montagem raiz para executar VMs. No ONTAP, essa opção é chamada de superusuário e, quando a opção superusuário é usada, não é necessário especificar o ID de usuário anônimo. Observe que regras de política de exportação com valores diferentes `-anon` e `-allow-suid` podem causar problemas de descoberta de SVM com as ferramentas do ONTAP. Aqui está uma política de exemplo: **Protocolo de Acesso: nfs3** especificação de correspondência do Cliente: 192.168.42.21 **regra de Acesso RO: Sys** regra de Acesso RW: Sys **UID anônimo** superusuário: Sys * se o plug-in NetApp NFS para VMware VAAI for usado, o protocolo deve ser definido como `nfs` quando a regra de política de exportação for criada ou modificada. O protocolo NFSv4 é necessário para que a descarga de cópia VAAI funcione, e especificar o protocolo como `nfs` inclui automaticamente as versões NFSv3 e NFSv4. * Os volumes do armazenamento de dados NFS são juntados do volume raiz do SVM; portanto, o ESXi também deve ter acesso ao volume raiz para navegar e montar volumes do armazenamento de dados. A política de exportação para o volume raiz e para quaisquer outros volumes em que a junção do volume do datastore esteja aninhada, deve incluir uma regra ou regras para os servidores ESXi concedendo acesso somente leitura. Aqui está uma política de exemplo para o volume raiz, também usando o plug-in VAAI: **Protocolo de acesso: nfs (que inclui nfs3 e nfs4)** especificação de correspondência do cliente: 192.168.42.21 **regra de acesso RO: Sys** regra de acesso RW: Nunca (melhor segurança para o volume raiz) **UID anônimo** superusuário: Sys (também necessário para o volume raiz com VAAI) * Use ferramentas ONTAP para o provisionamento de armazenamento de dados VMware (as melhores práticas de armazenamento de dados do VMware) **para o vSphere simplifica automaticamente as práticas de armazenamento de dados do VMware:** Para o provisionamento de dados VMware ONTAP). **Ao criar datastores para clusters VMware com o plug-in, selecione o cluster em vez de um único servidor ESX. Essa opção o aciona para montar automaticamente o datastore em todos os hosts do cluster.** Use a função de montagem de plug-in para aplicar datastores existentes a novos servidores. ** Quando não estiver usando as ferramentas do ONTAP para VMware vSphere, use uma única política de exportação para todos os servidores ou para cada cluster de servidores onde é necessário controle de acesso adicional. * Embora o ONTAP ofereça uma estrutura de namespace de volume flexível para organizar volumes em uma árvore usando junções, essa abordagem não tem valor para o vSphere. Ele cria um diretório para cada VM na raiz do datastore, independentemente da hierarquia do namespace do storage. Assim, a prática recomendada é simplesmente montar o caminho de junção para volumes para vSphere no volume raiz do SVM, que é como as ferramentas do ONTAP para VMware vSphere provisionam datastores. Não ter caminhos de junção aninhados também significa que nenhum volume é dependente de qualquer volume que não seja o volume raiz e que tirar um volume off-line ou destruí-lo, mesmo intencionalmente, não afeta o caminho para outros volumes. * Um tamanho de bloco de 4K é bom para partições NTFS em datastores NFS. A figura a seguir mostra a conectividade de um host vSphere para um datastore ONTAP NFS.



A tabela a seguir lista as versões de NFS e os recursos compatíveis.

Recursos do vSphere	NFSv3	NFSv4.1
VMotion e Storage vMotion	Sim	Sim
Alta disponibilidade	Sim	Sim
Tolerância de falhas	Sim	Sim
DRS	Sim	Sim
Perfis de host	Sim	Sim
Armazenamento DRS	Sim	Não
Controle de e/S de storage	Sim	Não
SRM	Sim	Não
Volumes virtuais	Sim	Não
Aceleração de hardware (VAAI)	Sim	Sim
Autenticação Kerberos	Não	Sim (aprimorado com o vSphere 6,5 e posterior para oferecer suporte a AES, krb5i)
Suporte multipathing	Não	Sim (ONTAP 9.14,1)

Ligação direta em rede

Às vezes, os administradores de storage preferem simplificar suas infraestruturas removendo switches de rede da configuração. Isso pode ser suportado em alguns cenários.

ISCSI e NVMe/TCP

Um host usando iSCSI ou NVMe/TCP pode ser conectado diretamente a um sistema de storage e operar normalmente. A razão é pathing. Conexões diretas a dois controladores de storage diferentes resultam em dois caminhos independentes para o fluxo de dados. A perda de caminho, porta ou controlador não impede que o outro caminho seja usado.

NFS

O armazenamento NFS com conexão direta pode ser usado, mas com uma limitação significativa - o failover não funcionará sem um esforço significativo de script, o que seria da responsabilidade do cliente.

O motivo pelo qual o failover sem interrupções é complicado com o storage NFS com conexão direta é o roteamento que ocorre no sistema operacional local. Por exemplo, suponha que um host tenha um endereço IP de 192.168.1.1/24 e esteja conectado diretamente a um controlador ONTAP com um endereço IP de 192.168.1.50/24. Durante o failover, esse endereço 192.168.1.50 pode fazer failover para a outra controladora e estará disponível para o host, mas como o host detecta sua presença? O endereço 192.168.1.1 original ainda existe na NIC host que não se conecta mais a um sistema operacional. O tráfego destinado a 192.168.1.50 continuaria a ser enviado para uma porta de rede inoperável.

A segunda NIC do SO poderia ser configurada como 192.168.1.2 e seria capaz de se comunicar com o endereço 192.168.1.50 com falha, mas as tabelas de roteamento local teriam um padrão de usar um endereço **e apenas um** para se comunicar com a sub-rede 192.168.1.0/24. Um sysadmin poderia criar uma estrutura de script que detectaria uma conexão de rede com falha e alteraria as tabelas de roteamento local ou colocaria

interfaces para cima e para baixo. O procedimento exato dependeria do SO em uso.

Na prática, os clientes da NetApp têm NFS com conexão direta, mas normalmente apenas para workloads em que as pausas de e/S durante failovers são aceitáveis. Quando os suportes rígidos são usados, não deve haver nenhum erro de e/S durante essas pausas. O IO deve travar até que os serviços sejam restaurados, seja por um failback ou intervenção manual para mover endereços IP entre NICs no host.

FC Direct Connect

Não é possível conectar diretamente um host a um sistema de storage ONTAP usando o protocolo FC. A razão é o uso de NPIV. A WWN que identifica uma porta ONTAP FC para a rede FC usa um tipo de virtualização chamado NPIV. Qualquer dispositivo conectado a um sistema ONTAP deve ser capaz de reconhecer um NPIV WWN. Não há fornecedores atuais de HBA que ofereçam um HBA que possa ser instalado em um host que possa suportar um destino NPIV.

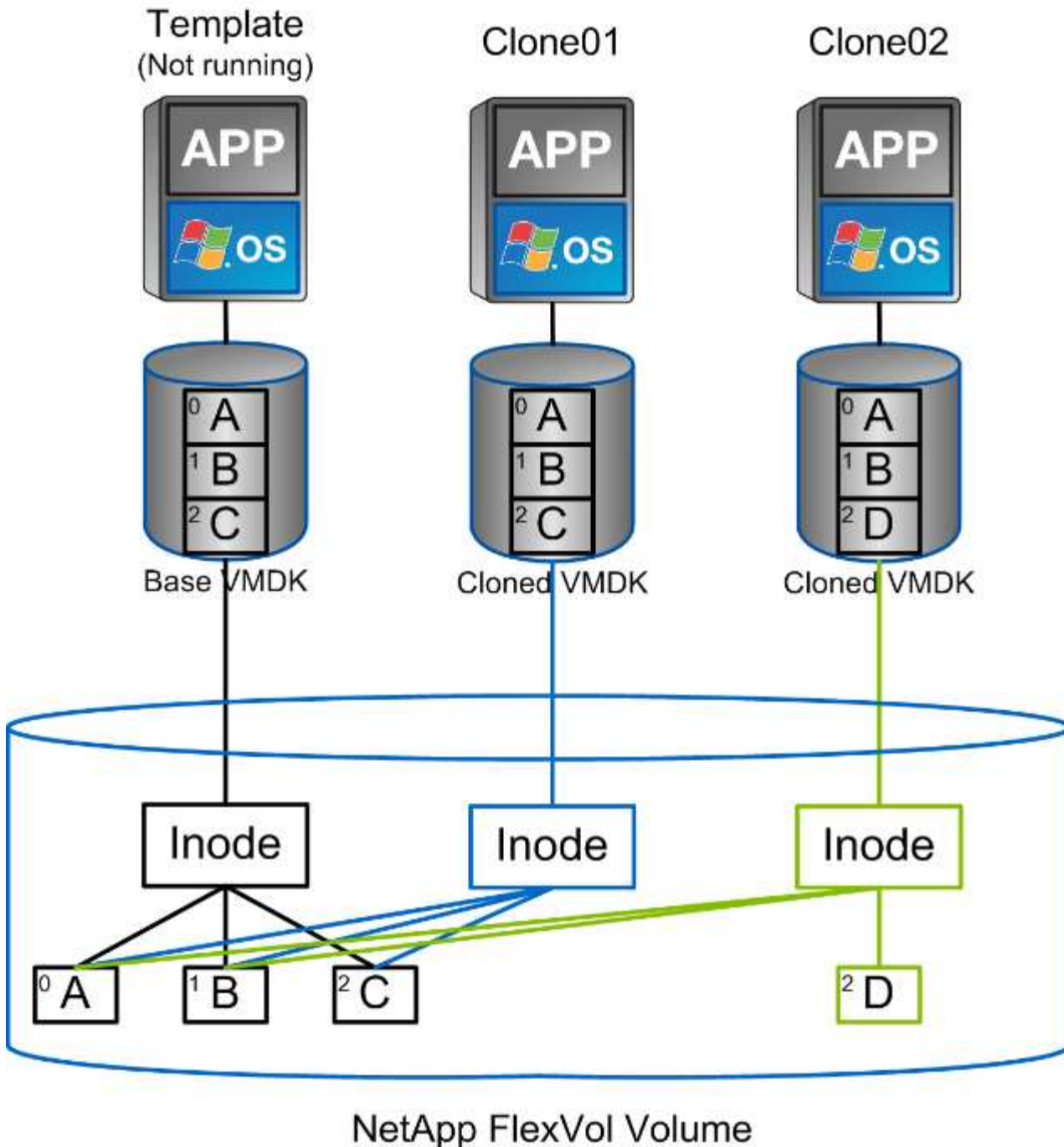
Clonagem de VM e datastore

Clonar um objeto de storage permite criar rapidamente cópias para uso adicional, como provisionamento de VMs adicionais, operações de backup/recuperação etc.

No vSphere, você pode clonar uma VM, um disco virtual, um vVol ou um datastore. Depois de clonado, o objeto pode ser ainda mais personalizado, muitas vezes por meio de um processo automatizado. O vSphere é compatível com clones de cópia completa, bem como clones vinculados, onde ele controla as alterações separadamente do objeto original.

Os clones vinculados são ótimos para economizar espaço, mas aumentam a quantidade de e/S que o vSphere lida com a VM, afetando a performance dessa VM e, talvez, o host geral. É por isso que os clientes da NetApp costumam usar clones baseados em sistema de storage para obter o melhor dos dois mundos: Uso eficiente do storage e maior performance.

A figura a seguir mostra a clonagem de ONTAP.



A clonagem pode ser descarregada para sistemas que executam o ONTAP por meio de vários mecanismos, normalmente no nível de VM, vVol ou datastore. Estes incluem o seguinte:

- VVols usando o fornecedor de APIs do NetApp vSphere para reconhecimento de armazenamento (VASA). Os clones do ONTAP são usados para dar suporte aos snapshots do VVol gerenciados pelo vCenter que são eficientes em termos de espaço com efeito de e/S mínimo para criá-los e excluí-los. As VMs também podem ser clonadas usando o vCenter, e elas também são descarregadas para o ONTAP, seja em um único datastore/volume ou entre datastores/volumes.
- Clonagem e migração do vSphere usando as APIs do vSphere – Array Integration (VAAI). As operações de clonagem de VM podem ser descarregadas para o ONTAP em ambientes SAN e nas (a NetApp fornece um plug-in ESXi para habilitar o VAAI para NFS). O vSphere apenas descarrega as operações em VMs frias (desativadas) em um datastore nas, enquanto as operações em VMs quentes (clonagem e armazenamento vMotion) também são descarregadas para SAN. O ONTAP usa a abordagem mais

eficiente com base nas licenças de produto de origem, destino e instalação. Esse recurso também é usado pelo VMware Horizon View.

- SRA (usado com o VMware Site Recovery Manager). Aqui, os clones são usados para testar a recuperação da réplica de DR sem interrupções.
- Backup e recuperação usando ferramentas do NetApp, como o SnapCenter. Os clones de VM são usados para verificar as operações de backup, bem como para montar um backup de VM para que arquivos individuais possam ser copiados.

A clonagem descarregada do ONTAP pode ser invocada por ferramentas VMware, NetApp e de terceiros. Clones que são descarregados para o ONTAP têm várias vantagens. Na maioria dos casos, elas usam espaço eficiente, precisando de storage somente para alterações no objeto. Não há efeito de desempenho adicional para lê-las e gravá-las e, em alguns casos, o desempenho é aprimorado ao compartilhar blocos em caches de alta velocidade. Eles também descarregam ciclos de CPU e e/S de rede do servidor ESXi. A descarga de cópia em um armazenamento de dados tradicional usando um FlexVol volume pode ser rápida e eficiente com a licença FlexClone, mas as cópias entre volumes FlexVol podem ser mais lentas. Se você mantiver os modelos de VM como uma fonte de clones, considere colocá-los no volume do datastore (use pastas ou bibliotecas de conteúdo para organizá-los) para clones rápidos e com uso eficiente de espaço.

Você também pode clonar um volume ou LUN diretamente no ONTAP para clonar um armazenamento de dados. Com os datastores NFS, a tecnologia FlexClone pode clonar um volume inteiro, e o clone pode ser exportado do ONTAP e montado pelo ESXi como outro datastore. Para armazenamentos de dados VMFS, o ONTAP pode clonar um LUN em um volume ou volume inteiro, incluindo um ou mais LUNs nele. Um LUN que contém um VMFS deve ser mapeado para um grupo de iniciadores ESXi (igroup) e, em seguida, demarcado pelo ESXi para ser montado e usado como um datastore regular. Para alguns casos de uso temporário, um VMFS clonado pode ser montado sem uma nova assinatura. Depois que um datastore é clonado, as VMs dentro dele podem ser registradas, reconfiguradas e personalizadas como se fossem VMs clonadas individualmente.

Em alguns casos, recursos licenciados adicionais podem ser usados para aprimorar a clonagem, como o SnapRestore para backup ou o FlexClone. Essas licenças são frequentemente incluídas em pacotes de licenças sem nenhum custo adicional. É necessária uma licença FlexClone para as operações de clonagem da Vevolve, bem como para dar suporte a snapshots gerenciados de uma VVol (que são descarregados do hipervisor para a ONTAP). Uma licença do FlexClone também pode melhorar certos clones baseados em VAAI quando usados em um datastore/volume (cria cópias instantâneas com uso eficiente de espaço em vez de cópias em bloco). Ele também é usado pelo SRA ao testar a recuperação de uma réplica de DR e SnapCenter para operações de clone, além de pesquisar cópias de backup para restaurar arquivos individuais.

Proteção de dados

Fazer backup de suas VMs e recuperá-las rapidamente estão entre os grandes pontos fortes do ONTAP para vSphere, e é fácil gerenciar essa capacidade no vCenter com o plug-in SnapCenter para VMware vSphere.

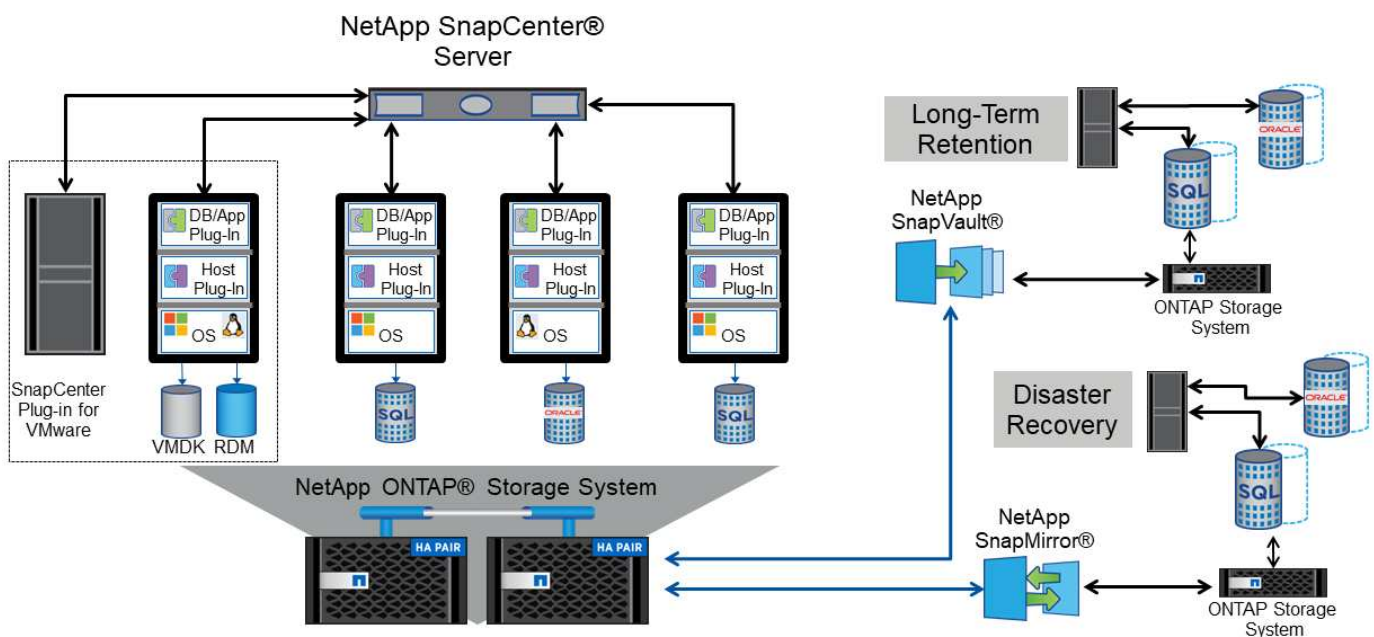
Use snapshots para fazer cópias rápidas da VM ou do armazenamento de dados sem afetar a performance e enviá-las para um sistema secundário usando o SnapMirror para proteção de dados externa a mais longo prazo. Essa abordagem minimiza o espaço de armazenamento e a largura de banda da rede, armazenando apenas informações alteradas.

O SnapCenter permite criar políticas de backup que podem ser aplicadas a vários trabalhos. Essas políticas podem definir agendamento, retenção, replicação e outros recursos. Elas continuam permitindo a seleção opcional de snapshots consistentes com VM, o que aproveita a capacidade do hipervisor de silenciar a I/O antes de tirar um snapshot da VMware. No entanto, devido ao efeito de desempenho dos snapshots VMware,

eles geralmente não são recomendados, a menos que você precise que o sistema de arquivos Guest seja encerrado. Em vez disso, use snapshots para proteção geral e use ferramentas de aplicativos, como plug-ins do SnapCenter, para proteger dados transacionais, como SQL Server ou Oracle. Esses snapshots são diferentes dos snapshots VMware (consistência) e são adequados para proteção de longo prazo. Os snapshots VMware são "recomendado" apenas para uso a curto prazo devido ao desempenho e a outros efeitos.

Esses plug-ins oferecem recursos estendidos para proteger os bancos de dados em ambientes físicos e virtuais. Com o vSphere, você pode usá-los para proteger bancos de dados SQL Server ou Oracle onde os dados são armazenados em LUNs RDM, LUNs iSCSI diretamente conectados ao SO convidado ou arquivos VMDK em armazenamentos de dados VMFS ou NFS. Os plug-ins permitem a especificação de diferentes tipos de backups de banco de dados, suporte a backup on-line ou off-line e proteção de arquivos de banco de dados juntamente com arquivos de log. Além de backup e recuperação, os plug-ins também dão suporte à clonagem de bancos de dados para fins de desenvolvimento ou teste.

A figura a seguir mostra um exemplo de implantação do SnapCenter.



Para obter recursos aprimorados de recuperação de desastres, considere usar o NetApp SRA para ONTAP com o VMware Site Recovery Manager. Além de oferecer suporte à replicação de datastores para um local de DR, ele também permite testes sem interrupções no ambiente de DR clonando os datastores replicados. A recuperação de um desastre e a reprotção da produção após a solução da interrupção também são facilitadas pela automação incorporada ao SRA.

Por fim, para obter o mais alto nível de proteção de dados, considere uma configuração vMSC (VMware vSphere Metro Storage Cluster) usando o NetApp MetroCluster. O vMSC é uma solução certificada pela VMware que combina replicação síncrona com clustering baseado em array, oferecendo os mesmos benefícios de um cluster de alta disponibilidade, mas distribuído em locais separados para proteger contra desastres no local. O NetApp MetroCluster oferece configurações econômicas para replicação síncrona com recuperação transparente a partir de qualquer falha única de componente de storage, bem como recuperação de comando único em caso de desastre no local. O vMSC é descrito em mais detalhes em "TR-4128".

Qualidade do serviço (QoS)

Os limites de taxa de transferência são úteis no controle de níveis de serviço, no

gerenciamento de cargas de trabalho desconhecidas ou para testar aplicativos antes da implantação para garantir que eles não afetem outras cargas de trabalho em produção. Eles também podem ser usados para restringir uma carga de trabalho bully depois que ela é identificada.

Suporte à política de QoS ONTAP

Os sistemas que executam o ONTAP podem usar o recurso de QoS de storage para limitar a taxa de transferência em Mbps e/ou e/os por segundo (IOPS) para diferentes objetos de storage, como arquivos, LUNs, volumes ou SVMs inteiras.

Níveis mínimos de serviço baseados em IOPS também são compatíveis para fornecer desempenho consistente para objetos SAN no ONTAP 9.2 e para objetos nas no ONTAP 9.3.

O limite máximo de taxa de transferência de QoS em um objeto pode ser definido em Mbps e/ou IOPS. Se ambos forem usados, o primeiro limite atingido é imposto pelo ONTAP. Um workload pode conter vários objetos e uma política de QoS pode ser aplicada a um ou mais workloads. Quando uma política é aplicada a vários workloads, os workloads compartilham o limite total da política. Objetos aninhados não são suportados (por exemplo, arquivos dentro de um volume não podem ter sua própria política). Os mínimos de QoS só podem ser definidos em IOPS.

As ferramentas a seguir estão disponíveis no momento para gerenciar políticas de QoS do ONTAP e aplicá-las a objetos:

- CLI do ONTAP
- Gerente do sistema da ONTAP
- OnCommand Workflow Automation
- Active IQ Unified Manager
- Kit de ferramentas do NetApp PowerShell para ONTAP
- Ferramentas do ONTAP para o provedor VMware vSphere VASA

Para atribuir uma política de QoS a um LUN, incluindo VMFS e RDM, o ONTAP SVM (exibido como SVM), caminho de LUN e número de série podem ser obtidos no menu sistemas de armazenamento na página inicial de ferramentas do ONTAP para VMware vSphere. Selecione o sistema de storage (SVM) e, em seguida, objetos relacionados > SAN. Use essa abordagem ao especificar QoS usando uma das ferramentas do ONTAP.

["Visão geral do gerenciamento e monitoramento de desempenho"](#) Consulte para obter mais informações.

Armazenamentos de dados NFS que não são vVols

Uma política de QoS do ONTAP pode ser aplicada a todo o datastore ou arquivos VMDK individuais dentro dele. No entanto, é importante entender que todas as VMs em um datastore NFS tradicional (não vVols) compartilham uma fila de e/S comum de um determinado host. Se qualquer VM for estrangulada por uma política de QoS do ONTAP, isso resultará na prática em que toda e/S desse datastore parecerá ser estrangulada para esse host.

Exemplo: * você configura um limite de QoS no VM1.vmdk para um volume que é montado como um armazenamento de dados NFS tradicional pelo host esxi-01. * O mesmo host (esxi-01) está usando vm2.vmdk e está no mesmo volume. * Se VM1.vmdk for estrangulado, então vm2.vmdk também parecerá ser estrangulado, pois compartilha a mesma fila de e/S com VM1.vmdk.



Isso não se aplica ao vVols.

A partir do vSphere 6,5, você pode gerenciar limites granulares de arquivos em datastores não vVols utilizando o Storage Policy-Based Management (SPBM) com Storage I/O Control (SIOC) v2.

Consulte os links a seguir para obter mais informações sobre como gerenciar o desempenho com as políticas SIOC e SPBM.

["Regras baseadas no host do SPBM: SIOC v2"](#) ["Gerencie os recursos de e/S de storage com o vSphere"](#)

Para atribuir uma política de QoS a um VMDK no NFS, observe as seguintes diretrizes:

- A política deve ser aplicada ao `vmname-flat.vmdk` que contém a imagem de disco virtual real, não ao `vmname.vmdk` (arquivo de descritor de disco virtual) ou `vmname.vmx` (arquivo de descritor de VM).
- Não aplique políticas a outros arquivos VM, como arquivos de swap virtuais (`vmname.vswp`).
- Ao usar o cliente da Web vSphere para encontrar caminhos de arquivo (datastore > Files), esteja ciente de que ele combina as informações do `-flat.vmdk` e `e.vmdk` simplesmente mostra um arquivo com o nome do `.vmdk`, mas o tamanho do `-flat.vmdk`. Adicione `-flat` ao nome do arquivo para obter o caminho correto.

Os armazenamentos de dados do FlexGroup oferecem recursos aprimorados de QoS ao usar as ferramentas do ONTAP para VMware vSphere 9,8 e posterior. Você pode facilmente definir QoS em todas as VMs em um datastore ou em VMs específicas. Consulte a seção FlexGroup deste relatório para obter mais informações. Esteja ciente de que as limitações de QoS mencionadas anteriormente com armazenamentos de dados NFS tradicionais ainda se aplicam.

Armazenamentos de dados VMFS

Usando LUNs ONTAP, as políticas de QoS podem ser aplicadas ao FlexVol volume que contém LUNs ou LUNs individuais, mas não arquivos VMDK individuais, porque o ONTAP não tem conhecimento do sistema de arquivos VMFS.

Armazenamentos de dados vVols

A QoS mínima e/ou máxima pode ser facilmente definida em VMs individuais ou VMDKs sem afetar qualquer outra VM ou VMDK usando o gerenciamento baseado em políticas de storage e vVols.

Ao criar o perfil de capacidade de storage para o contêiner da VVol, especifique um valor máximo e/ou mínimo de IOPS sob a capacidade de desempenho e, em seguida, faça referência a esse SCP com a política de armazenamento da VM. Use essa política ao criar a VM ou aplicar a diretiva a uma VM existente.



O vVols requer o uso de ferramentas do ONTAP para VMware vSphere, que funciona como o provedor VASA para ONTAP. ["VMware vSphere Virtual volumes \(vVols\) com o ONTAP"](#) Consulte para obter as práticas recomendadas do vVols.

QoS ONTAP e VMware SIOC

O ONTAP QoS e o VMware vSphere Storage I/O Control (SIOC) são tecnologias complementares que o vSphere e os administradores de storage podem usar em conjunto para gerenciar a performance de VMs vSphere hospedadas em sistemas executando o ONTAP. Cada ferramenta tem suas próprias forças, como mostrado na tabela a seguir. Devido aos diferentes escopos do VMware vCenter e do ONTAP, alguns objetos podem ser vistos e gerenciados por um sistema e não pelo outro.

Propriedade	QoS ONTAP	VMware SIOC
Quando ativo	A política está sempre ativa	Ativo quando existe contenção (latência do datastore sobre o limite)
Tipo de unidades	IOPS, Mbps	IOPS, compartilhamentos
Escopo do vCenter ou do aplicativo	Vários ambientes do vCenter, outros hipervisores e aplicações	Servidor vCenter único
Definir QoS na VM?	VMDK somente em NFS	VMDK em NFS ou VMFS
Definir QoS no LUN (RDM)?	Sim	Não
Definir QoS no LUN (VMFS)?	Sim	Sim (o datastore pode ser estrangulado)
Definir QoS no volume (armazenamento de dados NFS)?	Sim	Sim (o datastore pode ser estrangulado)
Definir QoS no SVM (locatário)?	Sim	Não
Abordagem baseada em políticas?	Sim. Pode ser compartilhado por todas as cargas de trabalho na política ou aplicado na íntegra a cada workload na política.	Sim, com o vSphere 6,5 e posterior.
Licença necessária	Incluído com ONTAP	Enterprise Plus

Programador de recursos distribuídos do VMware Storage

O VMware Storage Distributed Resource Scheduler (SDRS) é um recurso do vSphere que coloca as VMs no armazenamento com base na latência de e/S atual e no uso do espaço. Em seguida, ele move a VM ou VMDKs sem interrupções entre os armazenamentos de dados em um cluster de datastore (também chamado de pod), selecionando o melhor datastore no qual colocar a VM ou VMDKs no cluster do datastore. Um cluster de datastore é um conjunto de datastores semelhantes que são agregados em uma única unidade de consumo da perspectiva do administrador do vSphere.

Ao usar SDRS com ferramentas do ONTAP para VMware vSphere, primeiro você deve criar um datastore com o plug-in, usar o vCenter para criar o cluster do datastore e, em seguida, adicionar o datastore a ele. Após a criação do cluster do datastore, armazenamentos de dados adicionais podem ser adicionados ao cluster do datastore diretamente do assistente de provisionamento na página Detalhes.

Outras práticas recomendadas da ONTAP para SDRS incluem o seguinte:

- Todos os armazenamentos de dados no cluster devem usar o mesmo tipo de armazenamento (como SAS, SATA ou SSD), ser todos os armazenamentos de dados VMFS ou NFS e ter as mesmas configurações de replicação e proteção.
- Considere usar SDRS no modo padrão (manual). Essa abordagem permite que você analise as recomendações e decida se as aplicará ou não. Esteja ciente desses efeitos das migrações VMDK:
 - Quando OS SDRS migram VMDKs entre armazenamentos de dados, qualquer economia de espaço da clonagem ou deduplicação do ONTAP é perdida. Você pode executar novamente a deduplicação para recuperar essas economias.
 - Depois que OS SDRS movem VMDKs, o NetApp recomenda recriar os snapshots no datastore de origem porque o espaço é bloqueado pela VM que foi movida.

- Mover VMDKs entre armazenamentos de dados no mesmo agregado tem poucos benefícios, e OS SDRS não têm visibilidade de outras cargas de trabalho que possam compartilhar o agregado.

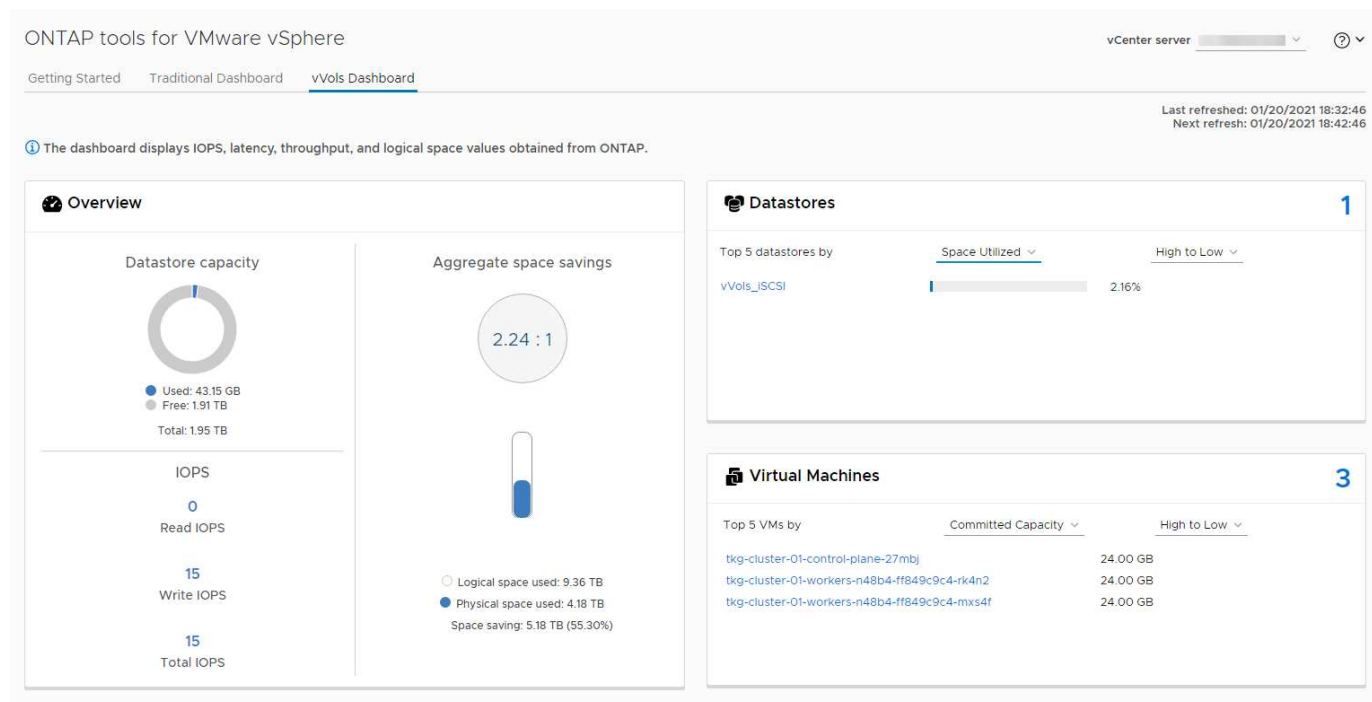
VVols e gerenciamento baseado em políticas de storage

As VMware vSphere APIs for Storage Awareness (VASA) facilitam para um administrador de storage configurar datastores com recursos bem definidos e permitem que o administrador da VM use-os sempre que necessário para provisionar VMs sem ter que interagir uns com os outros. Vale a pena dar uma olhada nessa abordagem para ver como ela pode otimizar suas operações de storage de virtualização e evitar muito trabalho trivial.

Antes do VASA, os administradores de VM podiam definir políticas de armazenamento de VM, mas precisavam trabalhar com o administrador de armazenamento para identificar armazenamentos de dados apropriados, geralmente usando documentação ou convenções de nomenclatura. Com o VASA, o administrador de storage pode definir uma variedade de recursos de storage, incluindo desempenho, disposição em camadas, criptografia e replicação. Um conjunto de recursos para um volume ou um conjunto de volumes é chamado de Perfil de capacidade de armazenamento (SCP).

O SCP suporta QoS mínimo e/ou máximo para vVols de dados de uma VM. A QoS mínima é suportada apenas em sistemas AFF. As ferramentas do ONTAP para VMware vSphere incluem um painel que exibe o desempenho granular da VM e a capacidade lógica para vVols em sistemas ONTAP.

A figura a seguir mostra as ferramentas do ONTAP para o painel vVols do VMware vSphere 9,8.



Após a definição do perfil de funcionalidade de storage, ele pode ser usado para provisionar VMs usando a política de storage que identifica seus requisitos. O mapeamento entre a política de armazenamento de VM e o perfil de capacidade de armazenamento de dados permite que o vCenter exiba uma lista de datastores compatíveis para seleção. Essa abordagem é conhecida como gerenciamento baseado em políticas de storage.

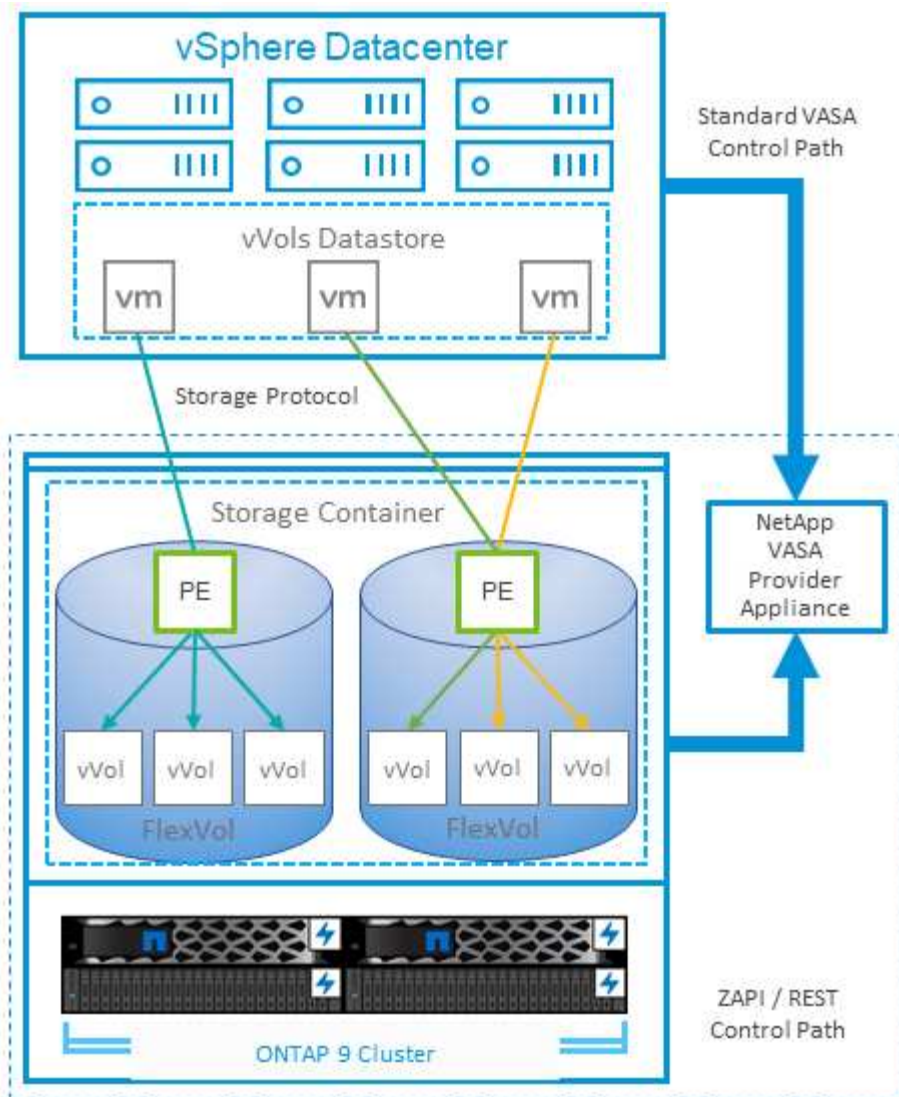
O VASA fornece a tecnologia para consultar o armazenamento e retornar um conjunto de recursos de armazenamento ao vCenter. Os fornecedores do VASA fornecem a tradução entre as APIs e construções do sistema de storage e as APIs da VMware que são entendidas pelo vCenter. O fornecedor VASA da NetApp

para ONTAP é oferecido como parte das ferramentas da ONTAP para a VM do dispositivo VMware vSphere, e o plug-in do vCenter fornece a interface para provisionar e gerenciar datastores vVol, bem como a capacidade de definir perfis de capacidade de armazenamento (SCPs).

O ONTAP dá suporte aos armazenamentos de dados VMFS e NFS VVol. O uso do vVols com armazenamentos de dados SAN traz alguns dos benefícios do NFS, como granularidade no nível da VM. Aqui estão algumas práticas recomendadas a serem consideradas e você pode encontrar informações adicionais em "[TR-4400](#)":

- Um datastore da VVol pode consistir em vários volumes FlexVol em vários nós dos clusters. A abordagem mais simples é um único datastore, mesmo quando os volumes têm capacidades diferentes. O SPBM garante que um volume compatível seja usado para a VM. No entanto, todos os volumes precisam fazer parte de um único SVM do ONTAP e acessá-los usando um único protocolo. Um LIF por nó para cada protocolo é suficiente. Evite o uso de várias versões do ONTAP em um único datastore da vVol porque as funcionalidades do storage podem variar entre lançamentos.
- Use as ferramentas do ONTAP para o plug-in do VMware vSphere para criar e gerenciar datastores da evolução. Além de gerenciar o datastore e seu perfil, ele cria automaticamente um endpoint de protocolo para acessar os vVols, se necessário. Se os LUNs forem usados, observe que os PES de LUN são mapeados usando IDs de LUN 300 e superiores. Verifique se a configuração do sistema avançado do host ESXi `Disk.MaxLUN` permite um número de ID LUN maior que 300 (o padrão é 1.024). Execute esta etapa selecionando o host ESXi no vCenter, a guia Configurar e localize `Disk.MaxLUN` na lista de Configurações avançadas do sistema.
- Não instale nem migre o provedor VASA, o vCenter Server (baseado em appliance ou Windows) ou as ferramentas do ONTAP para o próprio VMware vSphere em um datastore vVols, porque eles são mutuamente dependentes, limitando sua capacidade de gerenciá-los no caso de uma interrupção de energia ou outra interrupção do data center.
- Faça backup da VM do provedor VASA regularmente. No mínimo, crie instantâneos por hora do armazenamento de dados tradicional que contém o Fornecedor VASA. Para obter mais informações sobre como proteger e recuperar o provedor VASA, consulte este "[Artigo da KB](#)".

A figura a seguir mostra os componentes do vVols.



Backup e migração para a nuvem

Outro ponto forte da ONTAP é o amplo suporte à nuvem híbrida, unindo sistemas na nuvem privada local com funcionalidades de nuvem pública. Aqui estão algumas soluções de nuvem da NetApp que podem ser usadas em conjunto com o vSphere:

- **Ofertas de primeira parte.** O Amazon FSX para NetApp ONTAP, Google Cloud NetApp volumes e o Azure NetApp Files for ANF oferecem serviços de storage gerenciado de vários protocolos e alta performance nos principais ambientes de nuvem pública. Eles podem ser usados diretamente pelo VMware Cloud no AWS (VMC na AWS), Azure VMware Solution (AVS) e Google Cloud VMware Engine (GCVE) como armazenamentos de dados ou storage para sistemas operacionais convidados (GOS) e instâncias de computação.
- **Cloud Volumes ONTAP.** O software de gerenciamento de dados NetApp Cloud Volumes ONTAP oferece controle, proteção, flexibilidade e eficiência aos seus dados na nuvem escolhida. O Cloud Volumes ONTAP é um software de gerenciamento de dados nativo da nuvem desenvolvido com base no storage ONTAP. Use em conjunto com o Cloud Manager para implantar e gerenciar instâncias do Cloud Volumes ONTAP em conjunto com seus sistemas ONTAP locais. Aproveite os recursos avançados de SAN nas e iSCSI, juntamente com o gerenciamento de dados unificado, incluindo snapshots e replicação SnapMirror.
- **Serviços em nuvem.** Use o backup e a recuperação do BlueXP ou o SnapMirror Cloud para proteger

dados de sistemas locais usando storage de nuvem pública. O Cloud Sync ajuda a migrar e manter seus dados sincronizados em storage nas, armazenamentos de objetos e Cloud Volumes Service. A recuperação de desastres da BlueXP oferece uma solução econômica e eficiente para aproveitar as tecnologias NetApp como base para uma solução de recuperação de desastres robusta e capaz para recuperação de desastres na nuvem, DR no local e no local.

- **FabricPool.** O FabricPool oferece disposição em camadas rápida e fácil para dados do ONTAP. Os blocos inativos podem ser migrados para um armazenamento de objetos em nuvens públicas ou em um armazenamento de objetos StorageGRID privado e são recuperados automaticamente quando os dados do ONTAP são acessados novamente. Ou use a categoria objeto como um terceiro nível de proteção para dados que já são gerenciados pelo SnapVault. Com essa abordagem, você pode ["Armazene mais snapshots de suas VMs"](#) fazer isso em sistemas de storage ONTAP primário e/ou secundário.
- **ONTAP Select.** Use o storage definido por software da NetApp para estender sua nuvem privada pela Internet para instalações e escritórios remotos, onde você pode usar o ONTAP Select para oferecer suporte a serviços de bloco e arquivos, bem como os mesmos recursos de gerenciamento de dados do vSphere que você tem em seu data center empresarial.

Ao projetar suas aplicações baseadas em VM, considere a mobilidade futura na nuvem. Por exemplo, em vez de colocar arquivos de aplicativos e dados em conjunto, use uma exportação de LUN ou NFS separada para os dados. Isso permite que você migre a VM e os dados separadamente para os serviços de nuvem.

Para um mergulho profundo em mais tópicos de segurança, consulte os seguintes recursos.

- ["Documentação do Cloud Volumes ONTAP"](#)
- ["Documentação do ONTAP Select"](#)
- ["Documentação de backup e recuperação do BlueXP"](#)
- ["Documentação de recuperação de desastres do BlueXP"](#)
- ["Amazon FSX para NetApp ONTAP"](#)
- ["VMware Cloud na AWS"](#)
- ["O que é o Azure NetApp Files?"](#)
- ["Solução Azure VMware"](#)
- ["Google Cloud VMware Engine"](#)
- ["O que é o Google Cloud NetApp volumes?"](#)

Criptografia para dados do vSphere

Hoje, há cada vez mais demandas para proteger dados em repouso por meio da criptografia. Embora o foco inicial tenha sido em informações financeiras e de saúde, há um interesse crescente em proteger todas as informações, sejam elas armazenadas em arquivos, bancos de dados ou outros tipos de dados.

Os sistemas executando o ONTAP facilitam a proteção de quaisquer dados com criptografia em repouso. O NetApp Storage Encryption (NSE) usa unidades de disco com autcriptografia com o ONTAP para proteger dados SAN e nas. O NetApp também oferece o NetApp volume Encryption e o NetApp Aggregate Encryption como uma abordagem simples e baseada em software para encriptar volumes em qualquer unidade de disco. Esta criptografia de software não requer unidades de disco especiais ou gerenciadores de chaves externos e está disponível para clientes ONTAP sem nenhum custo adicional. É possível fazer upgrade e começar a usá-lo sem interromper seus clientes ou aplicações. Além disso, eles são validados para o padrão FIPS 140-2 nível 1, incluindo o gerenciador de chaves integrado.

Existem várias abordagens para proteger os dados de aplicativos virtualizados em execução no VMware vSphere. Uma abordagem é proteger os dados com software dentro da VM no nível do SO convidado. Os hipervisores mais recentes, como o vSphere 6,5, agora oferecem suporte à criptografia no nível da VM como outra alternativa. No entanto, a criptografia do software NetApp é simples e fácil e tem esses benefícios:

- **Nenhum efeito na CPU do servidor virtual.** Alguns ambientes de servidor virtual precisam de cada ciclo de CPU disponível para seus aplicativos, mas os testes mostraram que até 5x recursos de CPU são necessários com criptografia no nível do hipervisor. Mesmo que o software de criptografia ofereça suporte ao conjunto de instruções AES-NI da Intel para descarregar a carga de trabalho de criptografia (como a criptografia do software NetApp), essa abordagem pode não ser viável devido à exigência de novas CPUs que não são compatíveis com servidores mais antigos.
- **Gerenciador de chaves integrado incluído.** A criptografia do software NetApp inclui um gerenciador de chaves integrado sem custo adicional, o que facilita o início sem servidores de gerenciamento de chaves de alta disponibilidade que são complexos de comprar e usar.
- **Nenhum efeito na eficiência de armazenamento.** Técnicas de eficiência de storage, como deduplicação e compactação, são amplamente utilizadas atualmente e são essenciais para usar a Mídia de disco flash de maneira econômica. No entanto, os dados criptografados geralmente não podem ser deduplicados ou compactados. O hardware e a criptografia de storage da NetApp operam em um nível mais baixo e permitem o uso total dos recursos de eficiência de storage da NetApp líderes do setor, ao contrário de outras abordagens.
- **Criptografia granular fácil do datastore.** Com a criptografia de volume NetApp, cada volume recebe sua própria chave AES de 256 bits. Se você precisar alterá-lo, você pode fazê-lo com um único comando. Essa abordagem é ótima se você tiver vários locatários ou precisar provar criptografia independente para diferentes departamentos ou aplicativos. Essa criptografia é gerenciada no nível do datastore, o que é muito mais fácil do que gerenciar VMs individuais.

É simples começar a usar a criptografia de software. Depois que a licença for instalada, basta configurar o gerenciador de chaves integrado especificando uma senha e, em seguida, criar um novo volume ou fazer uma movimentação de volume no lado do storage para habilitar a criptografia. A NetApp está trabalhando para adicionar suporte mais integrado aos recursos de criptografia em versões futuras de suas ferramentas VMware.

Para um mergulho profundo em mais tópicos de segurança, consulte os seguintes recursos.

- ["Relatórios técnicos de segurança"](#)
- ["Guias de proteção de segurança"](#)
- ["Documentação do produto de segurança e criptografia de dados do ONTAP"](#)

Active IQ Unified Manager

O Active IQ Unified Manager fornece visibilidade das VMs em sua infraestrutura virtual e permite monitorar e solucionar problemas de storage e performance em seu ambiente virtual.

Uma implantação típica de infraestrutura virtual no ONTAP tem vários componentes espalhados pelas camadas de computação, rede e storage. Qualquer atraso de desempenho em uma aplicação de VM pode ocorrer devido a uma combinação de latências enfrentadas pelos vários componentes nas respectivas camadas.

A captura de tela a seguir mostra a exibição máquinas virtuais do Active IQ Unified Manager.

Active IQ Unified Manager

Virtual Machines

VIEW: Custom

Name	Status	Power State	Protocol	Capacity (Used Allocated)	4 IOPS	VM Latency (ms)	Host IOPS	Host Latency (ms)	Network Latency (ms)	Datastore IOPS	Datastore Latency (ms)
vCenter7	ON	ON	NFS	160 GB 712 GB	183	0	243	0	0	831	0.3

POWER: ON

VCENTER SERVER: vcenter7.stl.netapp.com

TOPOLOGY VIEW

Expand Topology

AD	ON	ON	NFS	8.05 GB 100 GB	167	0	306	0	0	831	0.3
BluePaddle-01	ON	ON	NFS	398 GB 2.26 TB	44	0	149	0	0	831	0.3
AIQUM	ON	ON	NFS	92 GB 400 GB	41	0	149	0	0	831	0.3
DirtWolf-02	ON	ON	NFS	138 GB 2.26 TB	39	0	306	0	0	831	0.3
BluePaddle-02	ON	ON	NFS	398 GB 2.26 TB	38	0	149	0	0	831	0.3

Showing all 44 Virtual Machines

O Unified Manager apresenta o subsistema subjacente de um ambiente virtual em uma visualização topológica para determinar se ocorreu um problema de latência no nó de computação, na rede ou no storage. A visualização também destaca o objeto específico que causa o atraso de desempenho para tomar medidas corretivas e solucionar o problema subjacente.

A captura de tela a seguir mostra a topologia expandida do AIQUM.

Active IQ Unified Manager

Expanded Topology for vCenter7

Close

VMs e gerenciamento baseado em políticas de storage

As VMware vSphere APIs for Storage Awareness (VASA) facilitam para um administrador de storage configurar datastores com recursos bem definidos e permitem

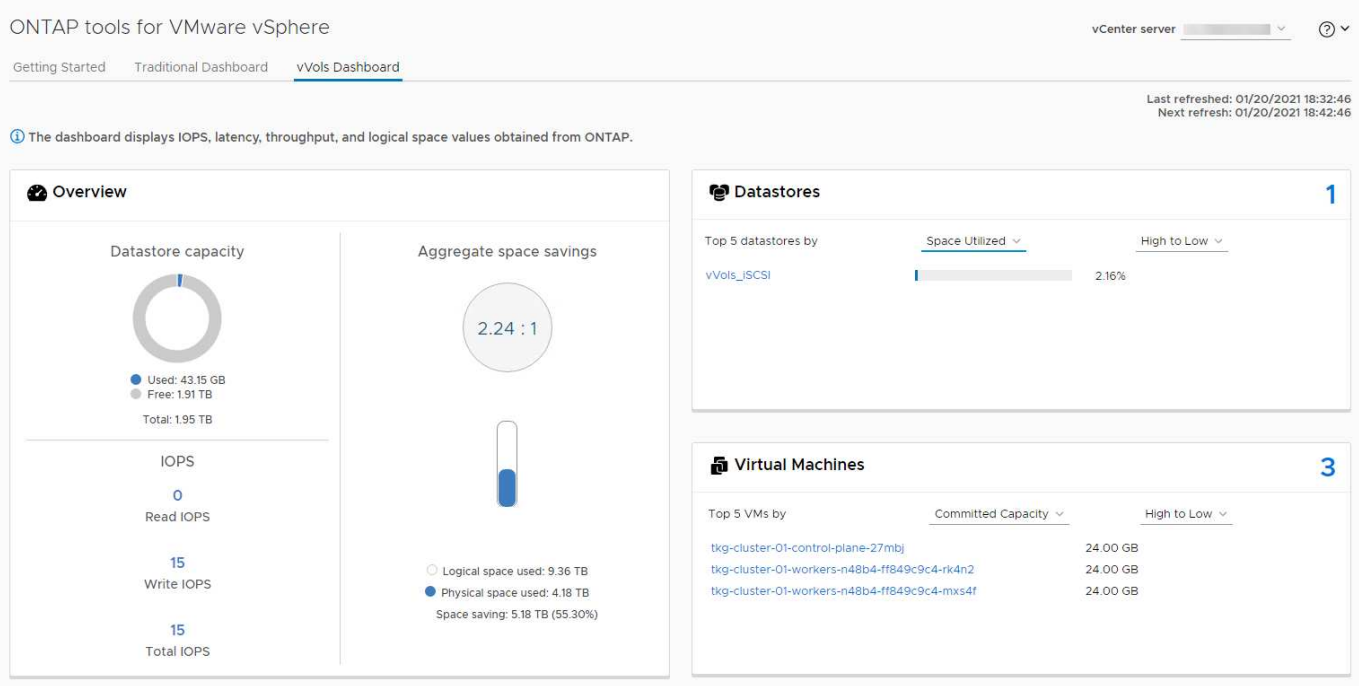
que o administrador da VM use-os sempre que necessário para provisionar VMs sem ter que interagir uns com os outros.

Vale a pena dar uma olhada nessa abordagem para ver como ela pode otimizar suas operações de storage de virtualização e evitar muito trabalho trivial.

Antes do VASA, os administradores de VM podiam definir políticas de armazenamento de VM, mas precisavam trabalhar com o administrador de armazenamento para identificar armazenamentos de dados apropriados, geralmente usando documentação ou convenções de nomenclatura. Com o VASA, o administrador de storage pode definir uma variedade de recursos de storage, incluindo desempenho, disposição em camadas, criptografia e replicação. Um conjunto de recursos para um volume ou um conjunto de volumes é chamado de Perfil de capacidade de armazenamento (SCP).

O SCP suporta QoS mínimo e/ou máximo para vVols de dados de uma VM. A QoS mínima é suportada apenas em sistemas AFF. As ferramentas do ONTAP para VMware vSphere incluem um painel que exibe o desempenho granular da VM e a capacidade lógica para vVols em sistemas ONTAP.

A figura a seguir mostra as ferramentas do ONTAP para o painel vVols do VMware vSphere 9,8.



Após a definição do perfil de funcionalidade de storage, ele pode ser usado para provisionar VMs usando a política de storage que identifica seus requisitos. O mapeamento entre a política de armazenamento de VM e o perfil de capacidade de armazenamento de dados permite que o vCenter exiba uma lista de datastores compatíveis para seleção. Essa abordagem é conhecida como gerenciamento baseado em políticas de storage.

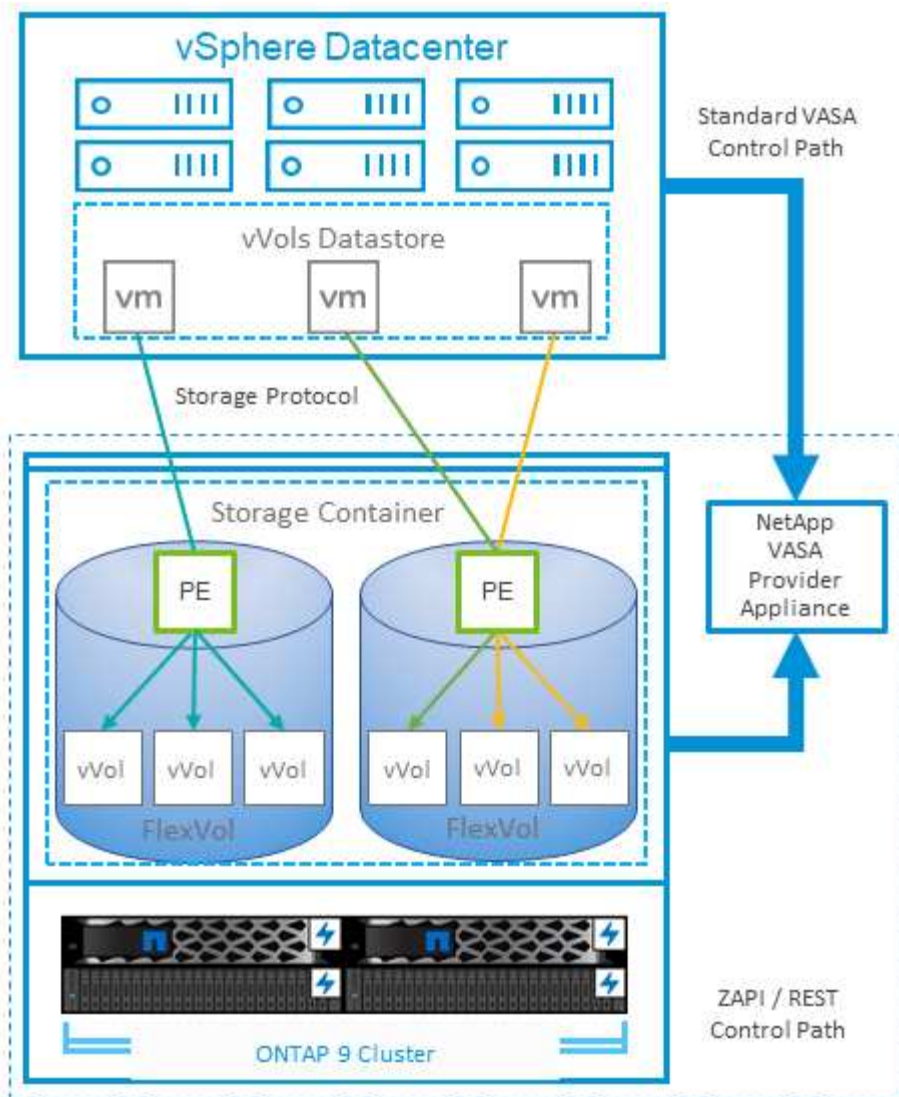
O VASA fornece a tecnologia para consultar o armazenamento e retornar um conjunto de recursos de armazenamento ao vCenter. Os fornecedores do VASA fornecem a tradução entre as APIs e construções do sistema de storage e as APIs da VMware que são entendidas pelo vCenter. O fornecedor VASA da NetApp para ONTAP é oferecido como parte das ferramentas da ONTAP para a VM do dispositivo VMware vSphere, e o plug-in do vCenter fornece a interface para provisionar e gerenciar datastores vVol, bem como a capacidade de definir perfis de capacidade de armazenamento (SCPs).

O ONTAP dá suporte aos armazenamentos de dados VMFS e NFS VVol. O uso do vVols com

armazenamentos de dados SAN traz alguns dos benefícios do NFS, como granularidade no nível da VM. Aqui estão algumas práticas recomendadas a serem consideradas e você pode encontrar informações adicionais em ["TR-4400"](#):

- Um datastore da VVol pode consistir em vários volumes FlexVol em vários nós dos clusters. A abordagem mais simples é um único datastore, mesmo quando os volumes têm capacidades diferentes. O SPBM garante que um volume compatível seja usado para a VM. No entanto, todos os volumes precisam fazer parte de um único SVM do ONTAP e acessá-los usando um único protocolo. Um LIF por nó para cada protocolo é suficiente. Evite o uso de várias versões do ONTAP em um único datastore da vVol porque as funcionalidades do storage podem variar entre lançamentos.
- Use as ferramentas do ONTAP para o plug-in do VMware vSphere para criar e gerenciar datastores da evolução. Além de gerenciar o datastore e seu perfil, ele cria automaticamente um endpoint de protocolo para acessar os vVols, se necessário. Se os LUNs forem usados, observe que os PES de LUN são mapeados usando IDs de LUN 300 e superiores. Verifique se a configuração do sistema avançado do host ESXi `Disk.MaxLUN` permite um número de ID LUN maior que 300 (o padrão é 1.024). Execute esta etapa selecionando o host ESXi no vCenter, a guia Configurar e localize `Disk.MaxLUN` na lista de Configurações avançadas do sistema.
- Não instale nem migre o provedor VASA, o vCenter Server (baseado em appliance ou Windows) ou as ferramentas do ONTAP para o próprio VMware vSphere em um datastore vVols, porque eles são mutuamente dependentes, limitando sua capacidade de gerenciá-los no caso de uma interrupção de energia ou outra interrupção do data center.
- Faça backup da VM do provedor VASA regularmente. No mínimo, crie instantâneos por hora do armazenamento de dados tradicional que contém o Fornecedor VASA. Para obter mais informações sobre como proteger e recuperar o provedor VASA, consulte este ["Artigo da KB"](#).

A figura a seguir mostra os componentes do vVols.



Programador de recursos distribuídos do VMware Storage

O VMware Storage Distributed Resource Scheduler (SDRS) é um recurso do vSphere que coloca as VMs no armazenamento com base na latência de e/S atual e no uso do espaço.

Em seguida, ele move a VM ou VMDKs sem interrupções entre os armazenamentos de dados em um cluster de datastore (também chamado de pod), selecionando o melhor datastore no qual colocar a VM ou VMDKs no cluster do datastore. Um cluster de datastore é um conjunto de datastores semelhantes que são agregados em uma única unidade de consumo da perspectiva do administrador do vSphere.

Ao usar SDRS com ferramentas do ONTAP para VMware vSphere, primeiro você deve criar um datastore com o plug-in, usar o vCenter para criar o cluster do datastore e, em seguida, adicionar o datastore a ele. Após a criação do cluster do datastore, armazenamentos de dados adicionais podem ser adicionados ao cluster do datastore diretamente do assistente de provisionamento na página Detalhes.

Outras práticas recomendadas da ONTAP para SDRS incluem o seguinte:

- Todos os armazenamentos de dados no cluster devem usar o mesmo tipo de armazenamento (como SAS, SATA ou SSD), ser todos os armazenamentos de dados VMFS ou NFS e ter as mesmas configurações de

replicação e proteção.

- Considere usar SDRS no modo padrão (manual). Essa abordagem permite que você analise as recomendações e decida se as aplicará ou não. Esteja ciente desses efeitos das migrações VMDK:
 - Quando OS SDRS migram VMDKs entre armazenamentos de dados, qualquer economia de espaço da clonagem ou deduplicação do ONTAP é perdida. Você pode executar novamente a deduplicação para recuperar essas economias.
 - Depois que OS SDRS movem VMDKs, o NetApp recomenda recriar os snapshots no datastore de origem porque o espaço é bloqueado pela VM que foi movida.
 - Mover VMDKs entre armazenamentos de dados no mesmo agregado tem poucos benefícios, e OS SDRS não têm visibilidade de outras cargas de trabalho que possam compartilhar o agregado.

Host ESXi recomendado e outras configurações do ONTAP

A NetApp desenvolveu um conjunto de configurações de host ESXi ideais para protocolos NFS e Block. Orientações específicas também são fornecidas para configurações de multipathing e tempo limite de HBA para o comportamento adequado com o ONTAP com base nos testes internos do NetApp e VMware.

Esses valores são facilmente definidos usando as ferramentas do ONTAP para VMware vSphere: No painel Resumo, clique em Editar configurações no portlet de sistemas host ou clique com o botão direito do Mouse no host no vCenter e navegue até Ferramentas do ONTAP > Definir valores recomendados.

Aqui estão as configurações de host atualmente recomendadas com as versões 9,8-9,13.

* Configuração do host*	Valor recomendado NetApp	Reinicialização necessária
Configuração Avançada do ESXi		
VMFS3.HardwareAcceleratedLocking	Manter padrão (1)	Não
VMFS3.EnableBlockDelete	Mantenha o padrão (0), mas pode ser alterado se necessário. Para obter mais informações, consulte "VMware KB 2007427"	Não
VMFS3.EnableVMFS6Unmap	Mantenha a predefinição (1) para obter mais informações, consulte "VMware vSphere APIs: Integração de array (VAAI)"	Não
Configurações NFS		
NET.TcpipHeapSize	vSphere 6,0 ou posterior, definido como 32. Todas as outras configurações NFS, definidas como 30	Sim
NET.TcpipHeapMax	Defina como 512MB para a maioria das versões do vSphere 6.X. Defina como padrão (1024MB) para 6.5U3, 6.7U3 e 7,0 ou posterior.	Sim

NFS.MaxVolumes	VSphere 6,0 ou posterior, defina como 256 todas as outras configurações NFS definidas como 64.	Não
NFS41.MaxVolumes	VSphere 6,0 ou posterior, definido como 256.	Não
1	VSphere 6,0 ou posterior, definido como 128	Sim
NFS.HeartbeatMaxFailures	Definido como 10 para todas as configurações NFS	Não
Frequência NFS.HeartbeatFrequency	Definido como 12 para todas as configurações NFS	Não
NFS.HeartbeatTimeout	Definido como 5 para todas as configurações NFS.	Não
Descrição: Sunrpc.MaxConnPerIP	VSphere 7,0 ou posterior, definido como 128.	Não
Configurações FC/FCoE		
Política de seleção de caminho	Defina como RR (round robin) quando os caminhos FC com ALUA são usados. Defina como FIXO para todas as outras configurações. Definir esse valor como RR ajuda a fornecer balanceamento de carga em todos os caminhos ativos/otimizados. O VALOR FIXO é para configurações mais antigas e não-ALUA e ajuda a impedir e/S de proxy Em outras palavras, ele ajuda a evitar que a e/S vá para o outro nó de um par de HA (high-availability) em um ambiente com Data ONTAP operando no modo 7D.	Não
Disk.QFullSampleSize	Defina como 32 para todas as configurações. Definir este valor ajuda a evitar erros de e/S.	Não
Disk.QFullThreshold	Defina como 8 para todas as configurações. Definir este valor ajuda a evitar erros de e/S.	Não
Tempos limite Emulex FC HBA	Use o valor padrão.	Não
Tempos limite do QLogic FC HBA	Use o valor padrão.	Não
Definições iSCSI		

Política de seleção de caminho	Defina como RR (round robin) para todos os caminhos iSCSI. Definir esse valor como RR ajuda a fornecer balanceamento de carga em todos os caminhos ativos/otimizados.	Não
Disk.QFullSampleSize	Defina como 32 para todas as configurações. Definir este valor ajuda a evitar erros de e/S	Não
Disk.QFullThreshold	Defina como 8 para todas as configurações. Definir este valor ajuda a evitar erros de e/S.	Não



1 - opção de configuração avançada NFS o MaxQueueDepth pode não funcionar como esperado ao usar o VMware vSphere ESXi 7.0.1 e o VMware vSphere ESXi 7,0.2. Consulte "[VMware KB 86331](#)" para obter mais informações.

As ferramentas do ONTAP também especificam certas configurações padrão ao criar volumes e LUNs do ONTAP FlexVol:

Ferramenta ONTAP	Predefinição
Reserva de snapshot (-percentagem-Snapshot-space)	0
Reserva fracionária (-reserva fracionária)	0
Atualização da hora de acesso (-atime-update)	Falso
Leitura mínima (-min-readahead)	Falso
Instantâneos programados	Nenhum
Eficiência de storage	Ativado
Garantia de volume	Nenhum (thin Provisioning)
Tamanho automático do volume	grow_shrink
Reserva de espaço LUN	Desativado
Alocação de espaço LUN	Ativado

Configurações de multipath para desempenho

Embora não esteja configurado atualmente pelas ferramentas ONTAP disponíveis, o NetApp sugere estas opções de configuração:

- Em ambientes de alto desempenho ou ao testar o desempenho com um único armazenamento de dados LUN, considere alterar a configuração de balanceamento de carga da política de seleção de caminho (PSP) round-robin (VMW_PSP_RR) da configuração padrão de IOPS de 1000 para um valor de 1. Consulte o KB da VMware "[2069356](#)" para obter mais informações.
- Na Atualização 1 do vSphere 6,7, a VMware introduziu um novo mecanismo de balanceamento de carga de latência para a PSP Round Robin. A nova opção considera a largura de banda de e/S e a latência do caminho ao selecionar o caminho ideal para e/S. Você pode se beneficiar ao usá-lo em ambientes com

conetividade de caminho não equivalente, como casos com mais saltos de rede em um caminho do que outro, ou ao usar um sistema de array NetApp All SAN. Consulte ["Plug-ins e políticas de seleção de caminho"](#) para obter mais informações.

Documentação adicional

Para FCP e iSCSI com vSphere, mais detalhes podem ser encontrados em [para FCP e iSCSI com vSphere 8](#), mais detalhes podem ser encontrados em ["Use o VMware vSphere 8.x com o ONTAP"](#) para NVMe-of com vSphere 7. Para NVMe-of com vSphere 8, mais detalhes podem ser encontrados em ["Para NVMe-of, mais detalhes podem ser encontrados em Configuração de host NVMe-of para ESXi 7.x com ONTAP"](#) para NVMe-of com vSphere 7 ["Use o VMware vSphere 7.x com o ONTAP"](#). Mais detalhes podem ser encontrados em ["Para NVMe-of, mais detalhes podem ser encontrados em Configuração de host NVMe-of para ESXi 8.x com ONTAP"](#)

Volumes virtuais (vVols) com ONTAP

Visão geral

O ONTAP é uma solução de storage líder para ambientes VMware vSphere há mais de duas décadas e continua adicionando recursos inovadores para simplificar o gerenciamento e reduzir custos.

Este documento abrange os recursos do ONTAP para volumes virtuais do VMware vSphere (vVols), incluindo as informações mais recentes sobre produtos e casos de uso, juntamente com as práticas recomendadas e outras informações para simplificar a implantação e reduzir erros.



Essa documentação substitui os relatórios técnicos publicados anteriormente *TR-4400: VMware vSphere Virtual volumes (vVols) com o ONTAP*

As práticas recomendadas complementam outros documentos, como guias e listas de compatibilidade. Eles são desenvolvidos com base em testes de laboratório e extensa experiência de campo por engenheiros e clientes da NetApp. Eles podem não ser as únicas práticas que funcionam ou são suportadas, mas são geralmente as soluções mais simples que atendem às necessidades da maioria dos clientes.



Este documento foi atualizado para incluir novos recursos do vVols encontrados na atualização 1 do vSphere 8,0, que são compatíveis com a versão 9,12 do ONTAP Tools.

Visão geral dos volumes virtuais (vVols)

A NetApp começou a trabalhar com a VMware para oferecer suporte às APIs do vSphere for Storage Awareness (VASA) para o vSphere 5 em 2012. Este provedor VASA inicial permitiu a definição de recursos de armazenamento em um perfil que poderia ser usado para filtrar datastores ao provisionar e verificar a conformidade com a política posteriormente. Com o tempo, isso evoluiu para adicionar novos recursos para permitir mais automação no provisionamento, bem como adicionar volumes virtuais ou vVols, onde objetos de storage individuais são usados para arquivos de máquina virtual e discos virtuais. Esses objetos podem ser LUNs, arquivos e agora com o vSphere 8 - NVMe Namespaces. NetApp trabalhou em conjunto com a VMware como parceiro de referência para vVols lançado com o vSphere 6 em 2015 e novamente como parceiro de design para vVols usando NVMe sobre Fabrics no vSphere 8. A NetApp continua a aprimorar os vVols para aproveitar os recursos mais recentes do ONTAP.

Existem vários componentes a ter em conta:

Fornecedor VASA

Este é o componente de software que lida com a comunicação entre o VMware vSphere e o sistema de storage. Para o ONTAP, o provedor VASA é executado em um dispositivo conhecido como ferramentas ONTAP para VMware vSphere (ferramentas ONTAP para abreviação). As ferramentas do ONTAP também incluem um plug-in do vCenter, um adaptador de replicação de armazenamento (SRA) para o VMware Site Recovery Manager e um servidor de API REST para criar sua própria automação. Depois que as ferramentas do ONTAP são configuradas e registradas no vCenter, há pouca necessidade de interagir diretamente com o sistema ONTAP, já que quase todas as suas necessidades de armazenamento podem ser gerenciadas diretamente na IU do vCenter ou por meio da automação da API REST.

Protocol Endpoint (PE)

O endpoint do protocolo é um proxy para e/S entre os hosts ESXi e o datastore vVols. O Fornecedor ONTAP VASA cria-os automaticamente, seja um LUN de ponto de extremidade de protocolo (4MB GB de tamanho) por FlexVol volume do armazenamento de dados vVols ou um ponto de montagem NFS por interface NFS (LIF) no nó de storage que hospeda um FlexVol volume no armazenamento de dados. O host ESXi monta esses pontos finais do protocolo diretamente em vez de LUNs e arquivos de disco virtuais individuais da Vevolve. Não é necessário gerenciar os endpoints do protocolo à medida que são criados, montados, desmontados e excluídos automaticamente pelo Provedor VASA, juntamente com quaisquer grupos de interface ou políticas de exportação necessárias.

Virtual Protocol Endpoint (VPE)

Novidade no vSphere 8: Ao usar o NVMe sobre Fabrics (NVMe-of) com vVols, o conceito de endpoint de protocolo não é mais relevante no ONTAP. Em vez disso, um PE virtual é instanciado automaticamente pelo host ESXi para cada grupo ANA assim que a primeira VM é ativada. O ONTAP cria automaticamente grupos ANA para cada FlexVol volume usado pelo datastore.

Uma vantagem adicional ao usar o NVMe-of para vVols é que não há solicitações de vinculação necessárias do Fornecedor VASA. Em vez disso, o host ESXi lida internamente com a funcionalidade de vinculação do VPE. Isso reduz a oportunidade de uma tempestade de Vevolve para impactar o serviço.

Para obter mais informações, consulte a. "[NVMe e Virtual volumes](https://www.vmware.com)" "[vmware.com](https://www.vmware.com)"

Armazenamento de dados de volume virtual

O Virtual volume datastore é uma representação lógica de datastore de um contentor vVols que é criado e mantido por um provedor VASA. O contentor representa um pool de capacidade de armazenamento provisionado a partir de sistemas de armazenamento gerenciados pelo Fornecedor VASA. As ferramentas do ONTAP dão suporte à alocação de vários volumes FlexVol (chamados de volumes de backup) a um único datastore vVols. Esses datastores vVols podem abranger vários nós em um cluster ONTAP, combinando flash e sistemas híbridos com recursos diferentes. O administrador pode criar novos volumes do FlexVol usando o assistente de provisionamento ou a API REST, ou selecionar volumes FlexVol pré-criados para backup de storage, se estiverem disponíveis.

Volumes virtuais (vVols)

VVols são os arquivos de máquina virtual e discos armazenados no datastore vVols. Usar o termo vVol (singular) está se referindo a um único arquivo específico, LUN ou namespace. O ONTAP cria namespaces, LUNs ou arquivos NVMe dependendo do protocolo que o armazenamento de dados usa. Existem vários tipos distintos de vVols; os mais comuns são Config (arquivos de metadados), dados (disco virtual ou VMDK) e Swap (criado quando a VM é ativada). VVols protegidos pela criptografia VMware VM será do tipo outro. A criptografia de VM VMware não deve ser confundida com o volume ONTAP ou a criptografia agregada.

Gerenciamento baseado em políticas

As VMware vSphere APIs for Storage Awareness (VASA) facilitam o uso dos recursos de armazenamento necessários para provisionar VMs sem ter que interagir com a equipe de armazenamento. Antes do VASA, os administradores de VM podiam definir políticas de armazenamento de VM, mas tiveram que trabalhar com seus administradores de storage para identificar armazenamentos de dados apropriados, geralmente usando documentação ou convenções de nomenclatura. Com o VASA, os administradores do vCenter com as permissões apropriadas podem definir uma variedade de recursos de armazenamento que os usuários do vCenter podem usar para provisionar VMs. O mapeamento entre a política de armazenamento de VM e o perfil de capacidade de armazenamento de dados permite que o vCenter exiba uma lista de armazenamentos de dados compatíveis para seleção, além de permitir que outras tecnologias como o Aria (anteriormente conhecido como vRealize) Automation ou o Tanzu Kubernetes Grid selecionem automaticamente o storage de uma política atribuída. Essa abordagem é conhecida como gerenciamento baseado em políticas de storage. Embora perfis e políticas de capacidade de storage também possam ser usados com datastores tradicionais, nosso foco aqui é nos datastores vVols.

Existem dois elementos:

Perfil de capacidade de armazenamento (SCP)
Um perfil de capacidade de armazenamento (SCP) é uma forma de modelo de armazenamento que permite que o administrador do vCenter defina quais recursos de armazenamento eles precisam sem realmente precisar entender como gerenciar esses recursos no ONTAP. Ao adotar uma abordagem de estilo de modelo, ele permite que o administrador forneça facilmente serviços de storage de uma forma consistente e previsível. As funcionalidades descritas em um SCP incluem desempenho, protocolo, eficiência de storage e outros recursos. Os recursos específicos variam de acordo com a versão. Eles são criados usando as ferramentas do ONTAP para o menu VMware vSphere na IU do vCenter. Você também pode usar APIs REST para criar SCPs. Eles podem ser criados manualmente selecionando recursos individuais ou gerados automaticamente a partir de armazenamentos de dados existentes (tradicionais).
Política de armazenamento de VM
As políticas de armazenamento de VM são criadas no vCenter em políticas e perfis. Para vVols, crie um conjunto de regras usando regras do provedor do tipo de storage do NetApp vVols. As ferramentas do ONTAP fornecem uma abordagem simplificada, permitindo que você simplesmente selecione um SCP em vez de forçá-lo a especificar regras individuais.

Como mencionado acima, o uso de políticas pode ajudar a simplificar a tarefa de provisionar um volume. Basta selecionar uma política apropriada e o Provedor VASA mostrará os datastores da vols que suportam essa política e colocam a evolução em um FlexVol volume individual compatível (Figura 1).

Implante a VM usando a Política de storage

New Virtual Machine

- ✓ 1 Select a creation type
- ✓ 2 Select a name and folder
- ✓ 3 Select a compute resource
- 4 Select storage**
- 5 Select compatibility
- 6 Select a guest OS
- 7 Customize hardware
- 8 Ready to complete

Select storage

Select the storage for the configuration and disk files

Encrypt this virtual machine (Requires Key Management Server)

VM Storage Policy Platinum

Disable Storage DRS for this virtual machine

	Name	Storage Compatibility	Capacity	Provisioned	Free	Type	Clu
<input checked="" type="radio"/>	vVolsiSCSI	Compatible	100 GB	40.74 GB	64.88 GB	vVol	
<input type="radio"/>	vVolsNFS2202...	Compatible	2 TB	36.88 GB	1.96 TB	vVol	
<input type="radio"/>	local-esx01	Incompatible	3.63 TB	1.46 GB	3.63 TB	VMFS 6	
<input type="radio"/>	local-esx07	Incompatible	1.81 TB	3.85 GB	1.81 TB	VMFS 6	
<input type="radio"/>	local-esx08	Incompatible	1.69 TB	1.43 GB	1.69 TB	VMFS 6	
<input type="radio"/>	local-esx09	Incompatible	1.81 TB	3.85 GB	1.81 TB	VMFS 6	
<input type="radio"/>	local-esx15	Incompatible	3.63 TB	1.46 GB	3.63 TB	VMFS 6	
<input type="radio"/>	tier001_ds	Incompatible	22 TB	23.73 TB	18.09 TB	NFS v3	

CANCEL

BACK

NEXT

Uma vez que uma VM é provisionada, o Fornecedor VASA continuará a verificar a conformidade e alertará o administrador da VM com um alarme no vCenter quando o volume de backup não for mais compatível com a política (Figura 2).

Conformidade com a política de storage da VM

Storage Policies

VM Storage Policies

AFF_VASA10

VM Storage Policy Compliance

⊗ Noncompliant

Last Checked Date

5/20/2022, 12:59:35 PM

VM Replication Groups

[CHECK COMPLIANCE](#)

Suporte para NetApp vVols

A ONTAP tem suportado a especificação VASA desde o seu lançamento inicial em 2012. Embora outros sistemas de armazenamento NetApp possam suportar VASA, este documento se concentra nas versões atualmente suportadas do ONTAP 9.

ONTAP

Além do ONTAP 9 nos sistemas AFF, ASA e FAS, o NetApp oferece suporte a cargas de trabalho VMware no ONTAP Select, Amazon FSX for NetApp com VMware Cloud na AWS, Azure NetApp Files com solução Azure VMware, Cloud Volumes Service com o VMware Engine e NetApp Private Storage na Equinix, mas a funcionalidade específica pode variar de acordo com o provedor de serviços e a conectividade de rede disponível. O acesso dos convidados do vSphere aos dados armazenados nessas configurações, bem como ao Cloud Volumes ONTAP também está disponível.

No momento da publicação, os ambientes de hyperscaler são limitados apenas aos armazenamentos de dados tradicionais NFS v3. Portanto, o vVols só está disponível com sistemas ONTAP no local ou sistemas conectados na nuvem que oferecem todos os recursos de sistemas no local, como os hospedados por parceiros e fornecedores de serviços da NetApp no mundo todo.

_Para obter mais informações sobre o ONTAP, "[Documentação do produto ONTAP](#)" consulte _

_Para obter mais informações sobre as práticas recomendadas do ONTAP e do VMware vSphere, "[TR-4597](#)" consulte _

Benefícios de usar vVols com ONTAP

Quando a VMware introduziu o suporte ao vVols com o VASA 2,0 em 2015, ela descreveu-o como "uma estrutura de integração e gerenciamento que fornece um novo modelo operacional para storage externo (SAN/nas)." Esse modelo operacional oferece vários benefícios em conjunto com o storage ONTAP.

Gerenciamento baseado em políticas

Conforme abordado na seção 1,2, o gerenciamento baseado em políticas permite que as VMs sejam provisionadas e gerenciadas posteriormente usando políticas pré-definidas. Isso pode ajudar as operações DE TI de várias maneiras:

- **Aumente a velocidade.** As ferramentas do ONTAP eliminam a necessidade de o administrador do vCenter abrir tickets com a equipe de storage para atividades de provisionamento de storage. No entanto, as funções de RBAC das ferramentas do ONTAP no vCenter e no sistema ONTAP ainda permitem que equipes independentes (como equipes de storage) ou atividades independentes da mesma equipe restrinjam o acesso a funções específicas, se desejado.
- **Provisionamento mais inteligente.** Os recursos do sistema de storage podem ser expostos por meio das APIs VASA, permitindo que os fluxos de trabalho de provisionamento aproveitem recursos avançados sem que o administrador da VM precise entender como gerenciar o sistema de storage.
- **Provisionamento mais rápido.** Diferentes recursos de storage podem ser suportados em um único armazenamento de dados e selecionados automaticamente, conforme apropriado, para uma VM com base na política de VM.
- **Evite erros.** As políticas de storage e VM são desenvolvidas com antecedência e aplicadas conforme necessário, sem precisar personalizar o storage sempre que uma VM é provisionada. Os alarmes de conformidade são gerados quando as funcionalidades de storage são desviadas das políticas definidas. Como mencionado anteriormente, as SCPs tornam o provisionamento inicial previsível e repetível, ao mesmo tempo que basear as políticas de armazenamento de VM nos SCPs garante um posicionamento preciso.

- * Melhor gerenciamento de capacidade. * As ferramentas VASA e ONTAP possibilitam visualizar a capacidade de armazenamento até o nível de agregado individual, se necessário, e fornecem várias camadas de alertas no evento que a capacidade começa a ficar baixa.

Gerenciamento granular de VM na SAN moderna

Os sistemas de STORAGE SAN que usam Fibre Channel e iSCSI foram os primeiros a serem suportados pelo VMware para ESX, mas não têm a capacidade de gerenciar arquivos e discos individuais de VM a partir do sistema de storage. Em vez disso, os LUNs são provisionados e o VMFS gerencia os arquivos individuais. Isso dificulta o sistema de storage gerenciar diretamente a performance, a clonagem e a proteção de storage de VM individuais. O vVols traz granularidade de storage que os clientes que usam o storage NFS já aproveitam, com os recursos de SAN robustos e de alta performance do ONTAP.

Agora, com as ferramentas vSphere 8 e ONTAP para VMware vSphere 9,12 e posterior, os mesmos controles granulares usados pelo vVols para protocolos baseados em SCSI legados agora estão disponíveis na moderna SAN Fibre Channel usando NVMe over Fabrics para obter desempenho ainda maior em escala. Com a atualização 1 do vSphere 8,0, agora é possível implantar uma solução NVMe completa usando vVols sem conversão de e/S na pilha de storage do hipervisor.

Maiores funcionalidades de descarga de storage

Embora a VAAI ofereça uma variedade de operações que são descarregadas para o armazenamento, existem algumas lacunas que são abordadas pelo Provedor VASA. O SAN VAAI não consegue descarregar snapshots gerenciados da VMware para o sistema de storage. O NFS VAAI pode descarregar snapshots gerenciados pela VM, mas há limitações colocadas uma VM com snapshots nativos de storage. Como o vVols usa LUNs, namespaces ou arquivos individuais para discos de máquina virtual, o ONTAP pode clonar os arquivos ou LUNs de forma rápida e eficiente para criar snapshots granular de VM que não exigem mais arquivos delta. O NFS VAAI também não dá suporte a operações de descarregamento de clones para migrações de storage vMotion hot (ativado). A VM deve ser desligada para permitir a descarga da migração ao usar o VAAI com datastores NFS tradicionais. O fornecedor VASA nas ferramentas do ONTAP permite clones quase instantâneos com eficiência de armazenamento para migrações quentes e frias, e também suporta cópias quase instantâneas para migrações entre volumes de vVols. Devido a esses benefícios significativos de eficiência de storage, você pode aproveitar ao máximo os workloads vVols no ["Garantia de eficiência"](#) programa. Da mesma forma, se os clones entre volumes que usam VAAI não atenderem aos seus requisitos, você provavelmente poderá resolver seus desafios de negócios graças às melhorias na experiência de cópia com vVols.

Casos de uso comuns para vVols

Além desses benefícios, também vemos esses casos de uso comuns para o storage da evolução:

- **Provisionamento sob demanda de VMs**
 - IaaS provedor de serviços ou nuvem privada.
 - Aproveite a automação e a orquestração por meio do pacote Aria (anteriormente vRealize), OpenStack, etc.
- **Discos de primeira Classe (FCDs)**
 - Volumes persistentes do VMware Tanzu Kubernetes Grid [TKG].
 - Fornecer serviços semelhantes ao Amazon EBS através do gerenciamento de ciclo de vida independente do VMDK.
- **Provisionamento sob demanda de VMs temporárias**
 - Laboratórios de teste/desenvolvimento

- Ambientes de treinamento

Benefícios comuns com vVols

Quando usados em seu pleno benefício, como nos casos de uso acima, o vVols fornece as seguintes melhorias específicas:

- Os clones são rapidamente criados em um único volume ou em vários volumes em um cluster do ONTAP, o que é uma vantagem em comparação com os clones habilitados para VAAI tradicionais. Eles também são eficientes em storage. Os clones dentro de um volume usam o clone de arquivo ONTAP, que são como os volumes do FlexClone e armazenam somente alterações do arquivo VVol/LUN/namespace de origem. Assim, as VMs de longo prazo para produção ou outras aplicações são criadas rapidamente, ocupam o mínimo de espaço e podem se beneficiar da proteção no nível da VM (usando o plug-in NetApp SnapCenter para VMware vSphere, snapshots gerenciados VMware ou backup VADP) e do gerenciamento de performance (com QoS ONTAP).
- Os vVols são a tecnologia de armazenamento ideal ao usar o TKG com o vSphere CSI, fornecendo classes de armazenamento discretas e capacidades gerenciadas pelo administrador do vCenter.
- Os serviços do Amazon EBS podem ser fornecidos por meio de FCDs porque um FCD VMDK, como o nome sugere, é um cidadão de primeira classe no vSphere e tem um ciclo de vida que pode ser gerenciado independentemente de VMs às quais ele pode ser anexado.

Usando vVols com ONTAP

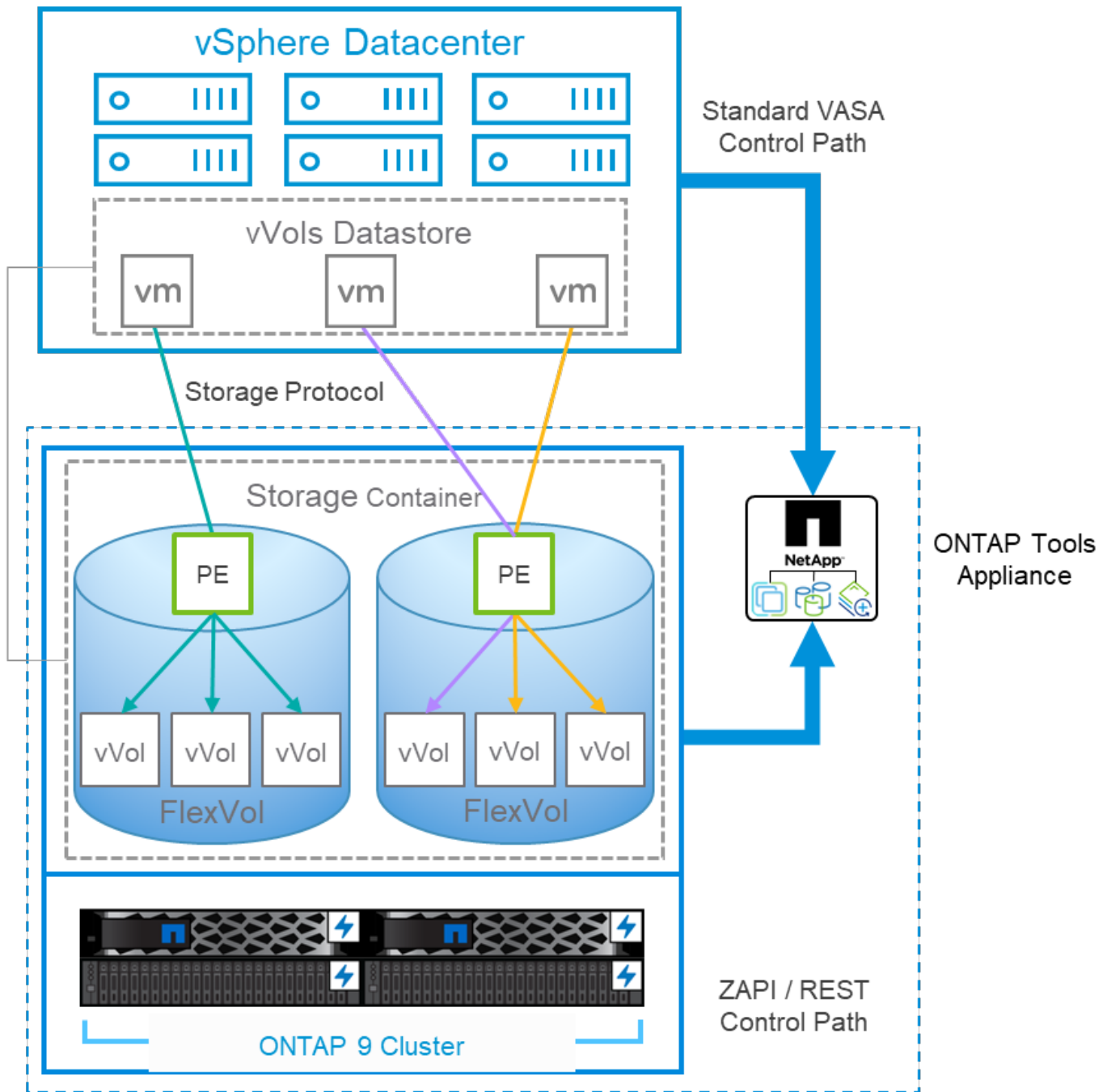
A chave para usar o vVols com o ONTAP é o software provedor VASA incluído como parte das ferramentas do ONTAP para o dispositivo virtual VMware vSphere.

As ferramentas do ONTAP também incluem as extensões de IU do vCenter, o servidor de API REST, o adaptador de replicação de armazenamento para o Gerenciador de recuperação de site da VMware, as ferramentas de configuração de monitoramento e host e uma matriz de relatórios que ajudam você a gerenciar melhor seu ambiente VMware.

Produtos e Documentação

A licença ONTAP FlexClone (incluída no ONTAP One) e o dispositivo ONTAP Tools são os únicos produtos adicionais necessários para usar vVols com o ONTAP. Versões recentes das ferramentas ONTAP são fornecidas como um único dispositivo unificado que é executado no ESXi, fornecendo a funcionalidade do que anteriormente eram três dispositivos e servidores diferentes. Para vVols, é importante usar as ferramentas do ONTAP vCenter UI Extensions ou REST APIs como ferramentas gerais de gerenciamento e interfaces de usuário para funções do ONTAP com o vSphere, juntamente com o provedor VASA, que fornece funcionalidade específica do vVols. O componente SRA está incluído nos datastores tradicionais, mas o VMware Site Recovery Manager não usa o SRA para vVols, em vez disso, implementando novos serviços no SRM 8,3 e posterior que utilizam o provedor VASA para replicação vVols.

Arquitetura do provedor VASA ao usar iSCSI ou FCP ONTAP



Instalação do produto

Para novas instalações, implante o dispositivo virtual em seu ambiente vSphere. As versões atuais das ferramentas do ONTAP se registrarão automaticamente no vCenter e habilitarão o provedor VASA por padrão. Além das informações do host ESXi e do vCenter Server, você também precisará dos detalhes de configuração do endereço IP do dispositivo. Como já foi dito anteriormente, o Fornecedor VASA requer que a licença ONTAP FlexClone já esteja instalada em quaisquer clusters ONTAP que você planeja usar para vVols. O dispositivo tem um watchdog integrado para garantir a disponibilidade e, como prática recomendada, deve ser configurado com recursos VMware High Availability e, opcionalmente, com tolerância a falhas. Consulte a seção 4,1 para obter mais detalhes. Não instale nem mova o dispositivo ONTAP Tools ou o vCenter Server Appliance (VCSA) para o storage vVols, pois isso pode impedir que os dispositivos sejam reiniciados.

As atualizações no local das ferramentas do ONTAP são suportadas usando o arquivo ISO de atualização disponível para download no site de suporte da NetApp (NSS). Siga as instruções do Guia de implementação

e configuração para atualizar o dispositivo.

Para dimensionar seu dispositivo virtual e entender os limites de configuração, consulte este artigo da base de conhecimento: ["Guia de dimensionamento das ferramentas do ONTAP para VMware vSphere"](#)

Documentação do produto

A documentação a seguir está disponível para ajudá-lo a implantar ferramentas do ONTAP.

["Para obter o repositório de documentação completo nº 44; visite este link para docs.NetApp.com"](#)

Comece agora

- ["Notas de lançamento"](#)
- ["Saiba mais sobre as ferramentas do ONTAP para VMware vSphere"](#)
- ["Ferramentas ONTAP Início rápido"](#)
- ["Implantar ferramentas do ONTAP"](#)
- ["Atualizar as ferramentas do ONTAP"](#)

Use as ferramentas do ONTAP

- ["Provisionar armazenamentos de dados tradicionais"](#)
- ["Provisione armazenamentos de dados vVols"](#)
- ["Configurar controles de acesso baseados em função"](#)
- ["Configure o diagnóstico remoto"](#)
- ["Configurar a alta disponibilidade"](#)

Proteja e gerencie armazenamentos de dados

- ["Proteger armazenamentos de dados tradicionais" Com SRM](#)
- ["Proteja máquinas virtuais baseadas em vVols" Com SRM](#)
- ["Monitore armazenamentos de dados tradicionais e máquinas virtuais"](#)
- ["Monitore armazenamentos de dados e máquinas virtuais do vVols"](#)

Além da documentação do produto, existem artigos da base de conhecimento de suporte que podem ser úteis.

- ["Como executar uma recuperação de desastres do provedor VASA - Guia de resolução"](#)

Painel do Fornecedor VASA

O provedor VASA inclui um painel com informações de desempenho e capacidade para VMs vVols individuais. Essas informações são fornecidas diretamente do ONTAP para os arquivos e LUNs da VEV, incluindo latência, IOPS, taxa de transferência e tempo de atividade das 5 principais VMs, além de latência e IOPS para os 5 principais datastores. Ele é habilitado por padrão ao usar o ONTAP 9.7 ou posterior. Pode levar até 30 minutos para que os dados iniciais sejam recuperados e exibidos no painel de instrumentos.

Painel do ONTAP Tools vVols

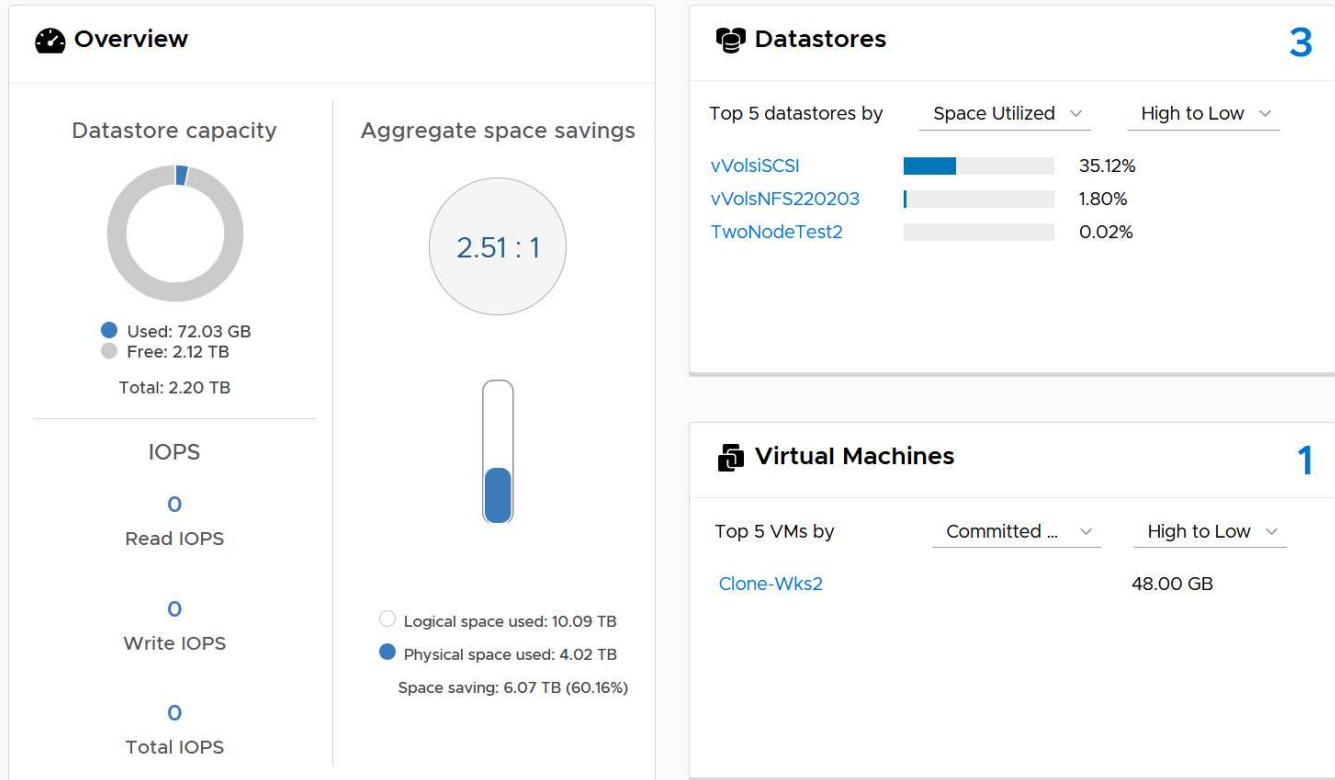
ONTAP tools for VMware vSphere

vCenter server [vm-is-vcenter01.vtme.netapp.com](#) ?

Getting Started Traditional Dashboard **vVols Dashboard**

Last refreshed: 05/20/2022 15:00:57
Next refresh: 05/20/2022 15:10:57

i The dashboard displays IOPS, latency, throughput, and logical space values obtained from ONTAP.



Práticas recomendadas

O uso do ONTAP vVols com o vSphere é simples e segue os métodos do vSphere publicados (consulte trabalhando com volumes virtuais no vSphere Storage na documentação da sua versão do ESXi). Aqui estão algumas práticas adicionais a serem consideradas em conjunto com o ONTAP.

Limites

Em geral, o ONTAP oferece suporte aos limites do vVols conforme definido pela VMware (consulte publicado "[Máximos de configuração](#)"). A tabela a seguir resume limites específicos de ONTAP em tamanho e número de vVols. Verifique sempre o "[NetApp Hardware Universe](#)" para obter limites atualizados sobre números e tamanhos de LUNs e arquivos.

Limites do ONTAP vVols

Capacidade/recurso	SAN (SCSI ou NVMe-of)	NFS
Tamanho máximo de vVols	62 TIB*	62 TIB*
Número máximo de vVols por FlexVol volume	1024	2 bilhões

Capacidade/recurso	SAN (SCSI ou NVMe-of)	NFS
Número máximo de vVols por nó ONTAP	Até 12.288**	50 bilhões
Número máximo de vVols por par ONTAP	Até 24.576**	50 bilhões
Número máximo de vVols por cluster ONTAP	Até 98.304**	Sem limite específico do cluster
Máximo de objetos QoS (grupo de políticas compartilhadas e nível de serviço vVols individuais)	12.000 até ONTAP 9.3; 40.000 com ONTAP 9.4 e posterior	

- Limite de tamanho com base em sistemas ASA ou em sistemas AFF e FAS que executam o ONTAP 9.12.1P2 e posterior.
 - O número de vVols SAN (namespaces NVMe ou LUNs) varia de acordo com a plataforma. Verifique sempre o ["NetApp Hardware Universe"](#) para obter limites atualizados sobre números e tamanhos de LUNs e arquivos.

Use as ferramentas do ONTAP para extensões de IU do VMware vSphere ou APIs REST para provisionar armazenamentos de dados vVols e Protocol Endpoints.*

Embora seja possível criar armazenamentos de dados vVols com a interface geral do vSphere, o uso de ferramentas do ONTAP criará automaticamente pontos de extremidade do protocolo conforme necessário e criará volumes FlexVol usando as práticas recomendadas do ONTAP e em conformidade com os perfis de funcionalidades de storage definidos. Basta clicar com o botão direito do Mouse no host/cluster/data center e selecionar *ONTAP Tools* e *provision datastore*. A partir daí, basta escolher as opções vVols desejadas no assistente.

Nunca armazene o dispositivo ONTAP Tools ou o vCenter Server Appliance (VCSA) em um datastore vVols que eles estejam gerenciando.

Isso pode resultar em uma "situação de frango e ovo" se você precisar reiniciar os aparelhos, porque eles não serão capazes de revincular seus próprios vVols enquanto eles estão reiniciando. Você pode armazená-los em um datastore vVols gerenciado por diferentes ferramentas do ONTAP e implantação do vCenter.

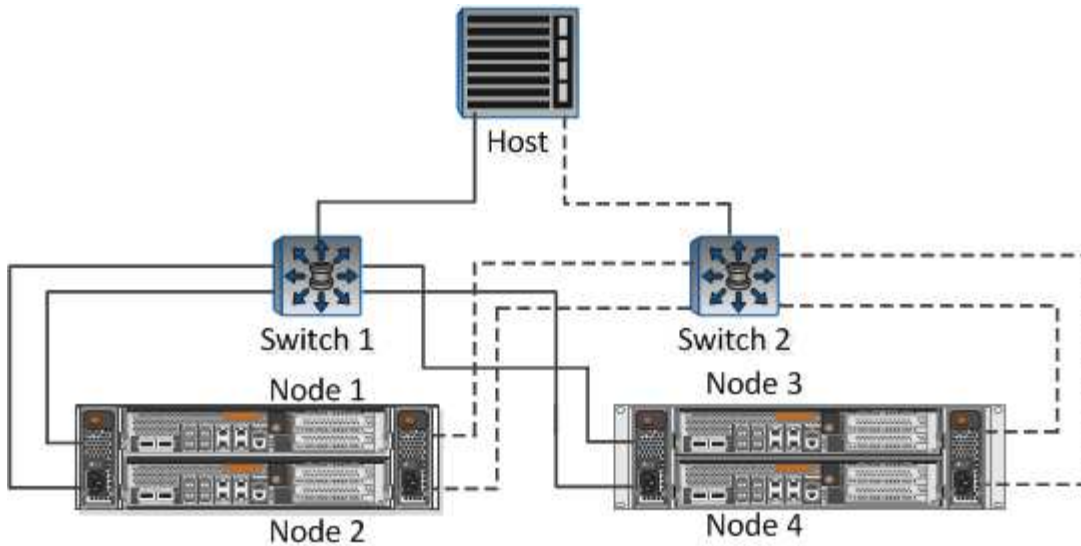
Evite operações vVols em diferentes versões do ONTAP.

Funcionalidades de storage compatíveis, como QoS, personalidade e muito mais, mudaram em várias versões do fornecedor VASA, e algumas dependem do lançamento do ONTAP. Usar versões diferentes em um cluster ONTAP ou mover vVols entre clusters com versões diferentes pode resultar em alarmes de conformidade ou comportamento inesperado.

Marque sua malha Fibre Channel antes de usar o NVMe/FC ou FCP para vVols.

O fornecedor de ferramentas ONTAP VASA cuida do gerenciamento de grupos FCP e iSCSI, bem como subsistemas NVMe no ONTAP com base em iniciadores descobertos de hosts ESXi gerenciados. No entanto, ele não se integra com switches Fibre Channel para gerenciar o zoneamento. O zoneamento deve ser feito de acordo com as melhores práticas antes que qualquer provisionamento possa ocorrer. O seguinte é um exemplo de zoneamento de iniciador único para quatro sistemas ONTAP:

Zoneamento do iniciador único:



Consulte os seguintes documentos para obter mais práticas recomendadas:

["TR-4080 melhores práticas para SAN ONTAP 9 moderna"](#)

["TR-4684 implementação e configuração de SANs modernas com NVMe-oF"](#)

Planeie o seu apoio FlexVols de acordo com as suas necessidades.

Pode ser desejável adicionar vários volumes de backup ao armazenamento de dados vVols para distribuir a carga de trabalho no cluster ONTAP, oferecer suporte a diferentes opções de política ou aumentar o número de LUNs ou arquivos permitidos. No entanto, se for necessária eficiência máxima de storage, coloque todos os volumes de backup em um único agregado. Ou, se for necessária a performance máxima de clonagem, considere usar um único FlexVol volume e manter seus modelos ou biblioteca de conteúdo no mesmo volume. O fornecedor VASA descarrega muitas operações de storage vVols para o ONTAP, incluindo migração, clonagem e snapshots. Quando isso é feito em um único FlexVol volume, clones de arquivo com uso eficiente de espaço são usados e ficam quase instantaneamente disponíveis. Quando isso é feito em volumes do FlexVol, as cópias ficam rapidamente disponíveis e usam deduplicação e compactação in-line, mas a eficiência máxima de storage não pode ser recuperada até que as tarefas em segundo plano sejam executadas em volumes usando deduplicação e compactação em segundo plano. Dependendo da origem e destino, alguma eficiência pode ser degradada.

Mantenha os Perfis de capacidade de armazenamento (SCPs) simples.

Evite especificar recursos que não são necessários definindo-os como nenhum. Isso minimizará os problemas ao selecionar ou criar volumes FlexVol. Por exemplo, com o provedor VASA 7,1 e anteriores, se a compressão for deixada na configuração padrão SCP de não, ela tentará desativar a compressão, mesmo em um sistema AFF.

Use os SCPs padrão como modelos de exemplo para criar seus próprios.

Os SCPs incluídos são adequados para a maioria dos usos gerais, mas seus requisitos podem ser diferentes.

Considere usar o máximo de IOPS para controlar VMs desconhecidas ou testar.

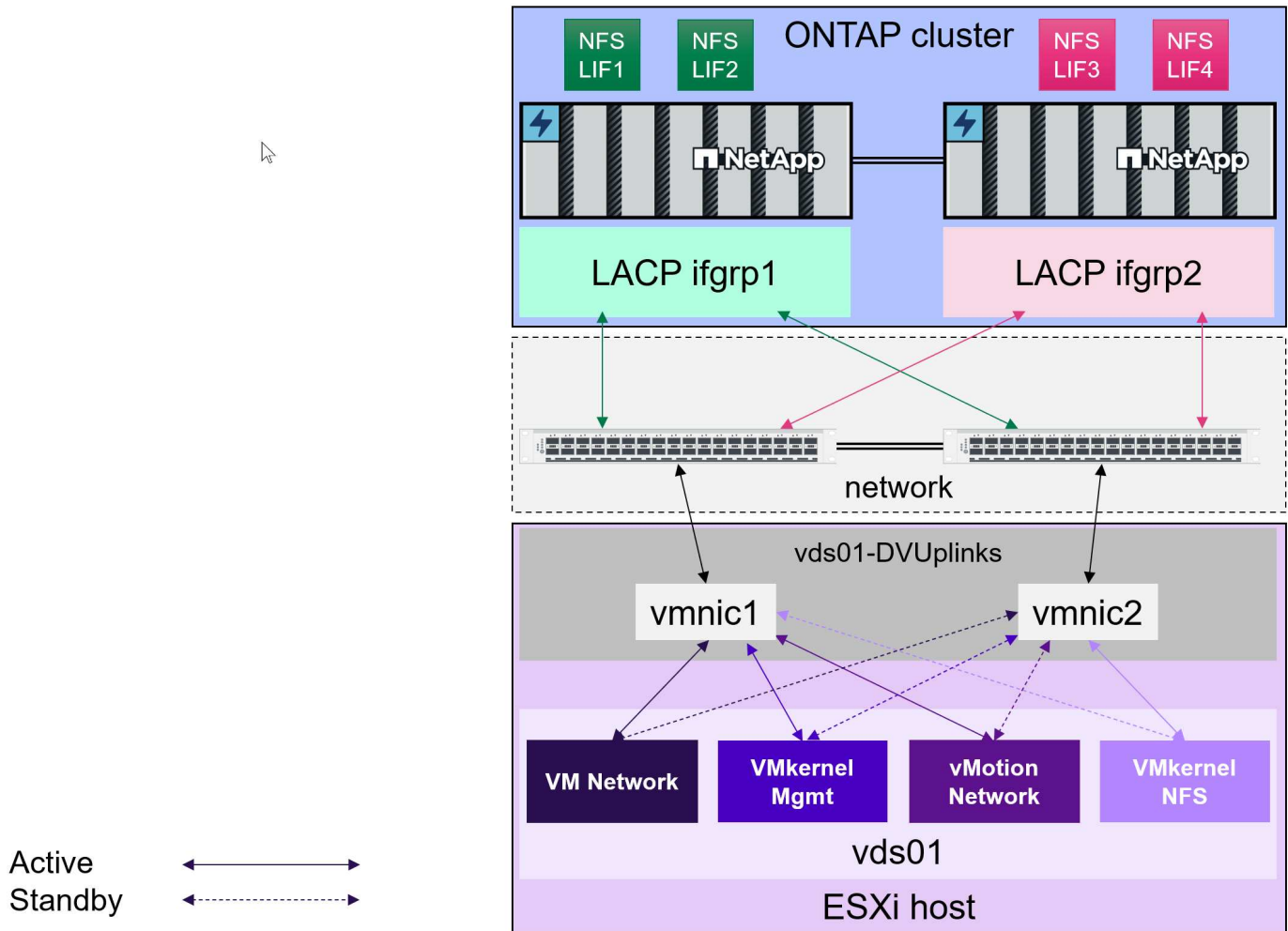
Disponível pela primeira vez no provedor VASA 7,1, o IOPS máximo pode ser usado para limitar as IOPS a uma evolução específica para uma carga de trabalho desconhecida, a fim de evitar impactos em outras cargas de trabalho mais críticas. Consulte a Tabela 4 para obter mais informações sobre o gerenciamento de desempenho.

Certifique-se de ter LIFs de dados suficientes. Crie pelo menos dois LIFs por nó por par de HA. Pode ser necessário mais com base na sua carga de trabalho.

Siga todas as melhores práticas de protocolo.

Consulte os outros guias de práticas recomendadas da NetApp e da VMware específicos do protocolo selecionado. Em geral, não existem outras alterações além das já mencionadas.

- Exemplo de configuração de rede usando vVols sobre NFS v3*



Implantando o storage vVols

Há várias etapas para criar o armazenamento do vVols para suas VMs.

As duas primeiras etapas podem não ser necessárias para um ambiente vSphere existente que usa o ONTAP para datastores tradicionais. Talvez você já esteja usando ferramentas do ONTAP para gerenciar, automatizar e gerar relatórios com seu VMFS ou storage tradicional baseado em NFS. Estas etapas são abordadas com mais detalhes na seção a seguir.

1. Crie a Storage Virtual Machine (SVM) e sua configuração de protocolo. Você selecionará NVMe/FC, NFSv3, NFSv4,1, iSCSI, FCP ou uma combinação dessas opções. Você pode usar assistentes do Gerenciador de sistema do ONTAP ou a linha de comando do shell do cluster.
 - Pelo menos um LIF por nó para cada conexão de switch/malha. Como prática recomendada, crie dois ou mais por nó para protocolos baseados em FCP, iSCSI ou NVMe.

- Os volumes podem ser criados no momento, mas é mais simples permitir que o assistente *provision datastore* os crie. A única exceção a esta regra é se você planeja usar a replicação do vVols com o VMware Site Recovery Manager. Isso é mais fácil de configurar com volumes FlexVol pré-existentes com relacionamentos SnapMirror existentes. Lembre-se de não ativar o QoS em nenhum volume a ser usado para vVols, pois isso deve ser gerenciado por ferramentas SPBM e ONTAP.
2. Implante as ferramentas do ONTAP para VMware vSphere usando o OVA baixado do site de suporte da NetApp.
 3. Configure as ferramentas do ONTAP para o seu ambiente.
 - Adicione o cluster do ONTAP às ferramentas do ONTAP em *sistemas de armazenamento*
 - Embora as ferramentas do ONTAP e o SRA ofereçam suporte a credenciais em nível de cluster e de SVM, o Fornecedor VASA dá suporte apenas a credenciais em nível de cluster para sistemas de storage. Isso ocorre porque muitas das APIs usadas para vVols só estão disponíveis no nível do cluster. Portanto, se você planeja usar o vVols, adicione seus clusters do ONTAP usando credenciais com escopo de cluster.
 - Se os LIFs de dados do ONTAP estiverem em sub-redes diferentes dos adaptadores do VMkernel, você deverá adicionar as sub-redes do adaptador VMkernel à lista de sub-redes selecionadas no menu de configurações das ferramentas do ONTAP. Por padrão, as ferramentas do ONTAP protegerão o tráfego de armazenamento permitindo apenas o acesso à sub-rede local.
 - As ferramentas do ONTAP vêm com várias políticas predefinidas que podem ser usadas ou consulte [Gerenciamento de VMs com políticas](#) para obter orientação sobre a criação de SCPs.
 4. Use o menu *ONTAP Tools* no vCenter para iniciar o assistente *provision datastore*.
 5. Forneça um nome significativo e selecione o protocolo desejado. Você também pode fornecer uma descrição do datastore.
 6. Selecione um ou mais SCPs compatíveis com o armazenamento de dados vVols. Isto irá filtrar quaisquer sistemas ONTAP que não possam corresponder ao perfil. Na lista resultante, selecione o cluster e o SVM desejados.
 7. Use o assistente para criar novos volumes FlexVol para cada um dos SCPs especificados ou use volumes existentes selecionando o botão de opção apropriado.
 8. Crie políticas de VM para cada SCP que será usado no armazenamento de dados no menu *políticas e Perfis* na IU do vCenter.
 9. Escolha o conjunto de regras de armazenamento "NetApp.Clustered.Data.ONTAP.VP.vvol". O conjunto de regras de armazenamento "NetApp.Clustered.Data.ONTAP.VP.VASA10" é compatível com armazenamentos de dados não vVols
 10. Você especificará o Perfil de capacidade de armazenamento por nome ao criar uma Política de armazenamento de VM. Enquanto estiver nesta etapa, você também pode configurar a correspondência de políticas do SnapMirror usando a guia replicação e a correspondência baseada em tags usando a guia tags. Observe que as tags já devem ser criadas para serem selecionáveis.
 11. Crie suas VMs, selecionando a Política de armazenamento de VM e o datastore compatível em Selecionar armazenamento.

Migração de VMs de datastores tradicionais para vVols

A migração de VMs de datastores tradicionais para um datastore vVols é tão simples quanto mover VMs entre datastores tradicionais. Basta selecionar a(s) VM(s), depois selecionar migrar da lista de ações e selecionar um tipo de migração de *change storage only*. As operações de cópia de migração serão descarregadas com o vSphere 6,0 e posterior para migrações SAN VMFS para vVols, mas não de VMDKs nas para vVols.

Gerenciamento de VMs com políticas

Para automatizar o provisionamento de storage com gerenciamento baseado em políticas, precisamos:

- Defina as capacidades do armazenamento (nó ONTAP e FlexVol volume) com os perfis de capacidade de armazenamento (SCPs).
- Crie políticas de armazenamento de VM que se mapeiam para as SCPs definidas.

A NetApp simplificou os recursos e o mapeamento começando com o provedor VASA 7,2, com melhorias contínuas em versões posteriores. Esta seção se concentra nessa nova abordagem. Versões anteriores suportavam um número maior de recursos e permitiam que eles fossem mapeados individualmente para políticas de storage, mas essa abordagem não é mais suportada.

Lançamento das funcionalidades de perfil de funcionalidade de storage das ferramentas do ONTAP

Capacidade SCP	Valores de capacidade	Lançamento suportado	Notas
Compressão	Sim, não, qualquer	Tudo	Obrigatório para AFF em 7,2 e mais tarde.
Desduplicação	Sim, não, qualquer	Tudo	M obrigatório para AFF em 7,2 e mais tarde.
Encriptação	Sim, não, qualquer	7,2 e mais tarde	Seleciona/cria FlexVol volume encriptados. É necessária licença ONTAP.
IOPS máximo	<number>	7,1 e mais tarde, mas diferenças	Listado em Grupo de políticas de QoS para 7,2 e posterior. Consulte 10 e posteriores para obter mais informações.
Personalidade	A FF, FAS	7,2 e mais tarde	O FAS também inclui outros sistemas que não sejam AFF, como o ONTAP Select. O AFF inclui o ASA.
Protocolo	NFS, NFS 4,1, iSCSI, FCP, NVMe/FC, qualquer	7,1 e anterior, 9,10 e posterior	7,2-9,8 é efetivamente "qualquer". Começando novamente em 9,10, onde NFS 4,1 e NVMe/FC foram adicionados à lista original.
Reserva de espaço (provisionamento fino)	Fino, grosso, (qualquer)	Tudo, mas diferenças	Chamado Thin Provisioning em 7,1 e anteriores, que também permitia valor de qualquer. Chamada reserva Espacial em 7,2. Todas as versões são padrão para thin.

Capacidade SCP	Valores de capacidade	Lançamento suportado	Notas
Política de disposição em camadas	Qualquer, nenhum, Instantâneo, Automático	7,2 e mais tarde	Usado para FabricPool - requer AFF ou ASA com ONTAP 9.4 ou posterior. Somente o instantâneo é recomendado a menos que use uma solução S3 no local como o NetApp StorageGRID.

Criando perfis de capacidade de armazenamento

O Fornecedor NetApp VASA vem com vários SCPs predefinidos. Novos SCPs podem ser criados manualmente, usando a IU do vCenter ou por meio de automação usando APIs REST. Especificando recursos em um novo perfil, clonando um perfil existente ou gerando perfis automaticamente de datastores tradicionais existentes. Isso é feito usando os menus em Ferramentas do ONTAP. Use *Perfis de capacidade de armazenamento* para criar ou clonar um perfil e *Mapeamento de armazenamento* para gerar automaticamente um perfil.

Funcionalidades de storage para ferramentas ONTAP 9,10 e posteriores

Create Storage Capability Profile

- 1 General
- 2 Platform
- 3 Protocol
- 4 Performance
- 5 Storage attributes
- 6 Summary

General

Specify a name and description for the storage capability profile. ?

Name:

Description:

CANCEL
NEXT

Create Storage Capability Profile

- 1 General
- 2 Platform**
- 3 Protocol
- 4 Performance
- 5 Storage attributes
- 6 Summary

Platform

Platform: All Flash FAS (AFF)

CANCEL

BACK

NEXT

Create Storage Capability Profile

- 1 General
- 2 Platform
- 3 Protocol**
- 4 Performance
- 5 Storage attributes
- 6 Summary

Protocol

Protocol: Any

- Any
- FCP
- NFS
- NFS 4.1
- iSCSI
- NVMe/FC

CANCEL

BACK

NEXT

Create Storage Capability Profile

- 1 General
- 2 Platform
- 3 Protocol
- 4 Performance**
- 5 Storage attributes
- 6 Summary

Performance

None ⓘ

QoS policy group ⓘ

Min IOPS:

Max IOPS:

Unlimited

CANCEL

BACK

NEXT

Create Storage Capability Profile

- 1 General
- 2 Platform
- 3 Protocol
- 4 Performance
- 5 Storage attributes**
- 6 Summary

Storage attributes

Deduplication: ▼

Compression: ▼

Space reserve: ▼

Encryption: ▼

Tiering policy (FabricPool): ▼

CANCEL

BACK

NEXT

Create Storage Capability Profile

- 1 General
- 2 Platform
- 3 Protocol
- 4 Performance
- 5 Storage attributes
- 6 Summary

Summary

Name:	New_SCP
Description:	N/A
Platform:	All Flash FAS (AFF)
Protocol:	Any
Min IOPS:	1000 IOPS
Max IOPS:	Unlimited
Space reserve:	Thin
Deduplication:	Yes
Compression:	Yes
Encryption:	Yes
Tiering policy (FabricPool):	Snapshot

CANCEL
BACK
FINISH

Criando datastores vVols uma vez que os SCPs necessários tenham sido criados, eles podem ser usados para criar o datastore vVols (e, opcionalmente, volumes FlexVol para o datastore). Clique com o botão direito do Mouse no host, cluster ou datacenter no qual você deseja criar o datastore vVols e selecione *ONTAP Tools > provision datastore*. Selecione um ou mais SCPs compatíveis com o armazenamento de dados e, em seguida, selecione um dos volumes FlexVol existentes e/ou provisione novos volumes FlexVol para o armazenamento de dados. Finalmente, especifique o SCP padrão para o datastore, que será usado para VMs que não têm um SCP especificado por política, bem como para vVols swap (isso não requer armazenamento de alto desempenho).

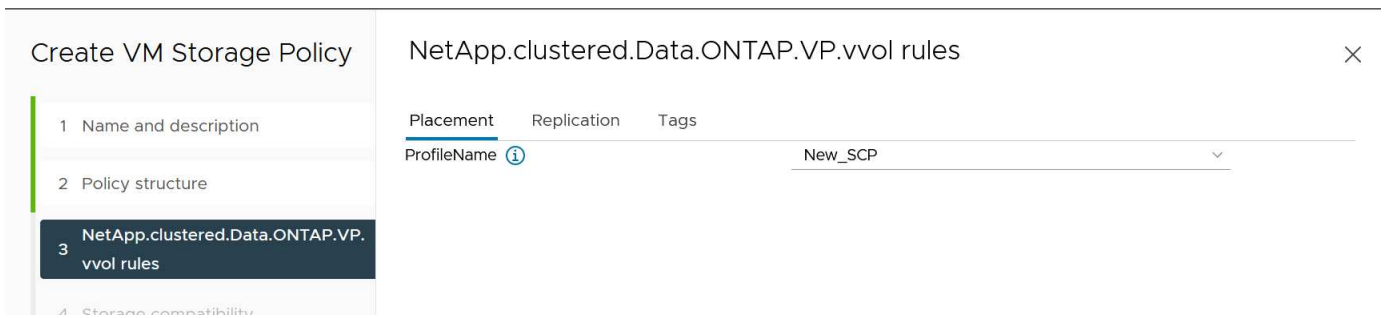
Criando políticas de armazenamento de VM

As políticas de armazenamento de VM são usadas no vSphere para gerenciar recursos opcionais, como Storage I/O Control ou vSphere Encryption. Eles também são usados com vVols para aplicar funcionalidades de storage específicas à VM. Use o tipo de storage "NetApp.Clustered.Data.ONTAP.VP.vvol" e a regra "profilename" para aplicar um SCP específico às VMs por meio do uso da Política. Veja o [xref:./vmware/vmware-vvols-ONTAP.exemplo de configuração de rede usando vVols sobre NFS v3](#)] para um exemplo disso com as ferramentas do provedor VASA da ONTAP. As regras para o storage "NetApp.Clustered.Data.ONTAP.VP.VASA10" devem ser usadas com datastores não baseados em vVols.

Versões anteriores são semelhantes, mas como mencionado no [Lançamento das funcionalidades de perfil de funcionalidade de storage das ferramentas do ONTAP](#), suas opções variam.

Uma vez criada a política de storage, ela pode ser usada ao provisionar novas VMs, conforme mostrado na ["Implante a VM usando a Política de storage"](#). As diretrizes para o uso de recursos de gerenciamento de desempenho com o provedor VASA 7,2 são abordadas [10 e posteriores](#)no .

Criação de políticas de armazenamento de VM com ferramentas ONTAP Forneceador VASA 9,10



Gerenciamento de desempenho com as ferramentas ONTAP 9,10 e posteriores

- O ONTAP Tools 9,10 usa seu próprio algoritmo de colocação equilibrada para colocar uma nova evolução no melhor FlexVol volume dentro de um datastore vVols. A colocação é baseada no SCP especificado e volumes FlexVol correspondentes. Isso garante que o armazenamento de dados e o armazenamento de backup possam atender aos requisitos de desempenho especificados.
- A alteração dos recursos de performance, como IOPS mín. E máx., requer alguma atenção à configuração específica.
 - **IOPS mínimo e máximo** podem ser especificados em um SCP e usados em uma Política de VM.
 - A alteração do IOPS no SCP não mudará a QoS nos vVols até que a Política de VM seja editada e, em seguida, reaplicada às VMs que a usam ([10 e posteriores](#) consulte). Ou crie um novo SCP com o IOPS desejado e altere a política para usá-lo (e reapplye às VMs). Geralmente, é recomendável simplesmente definir políticas separadas de armazenamento de VM e SCPs para diferentes níveis de serviço e simplesmente alterar a política de armazenamento de VM na VM.
 - As personalidades do AFF e do FAS têm configurações diferentes de IOPs. Tanto o mínimo quanto o máximo estão disponíveis no AFF. No entanto, sistemas que não sejam AFF só podem usar configurações de IOPs máximas.
- Em alguns casos, uma evolução pode precisar ser migrada após uma mudança de política (manual ou automaticamente pelo provedor VASA e ONTAP):
 - Algumas alterações não exigem migração (como a alteração do IOPS máximo, que pode ser aplicada imediatamente à VM, conforme descrito acima).
 - Se a alteração de política não puder ser suportada pelo FlexVol volume atual que armazena a evolução (por exemplo, a plataforma não oferece suporte à política de criptografia ou disposição em camadas solicitada), você precisará migrar manualmente a VM no vCenter.
- As ferramentas do ONTAP criam políticas individuais de QoS não compartilhadas com versões atualmente suportadas do ONTAP. Portanto, cada VMDK individual receberá sua própria alocação de IOPs.

Replicar a política de armazenamento de VM

VM Storage Policies

CREATE CHECK EDIT CLONE **REAPPLY** DELETE

Filter

<input type="checkbox"/>	Name	VC
<input type="checkbox"/>	Management Storage Policy - Large	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	VVol No Requirements Policy	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	Management Storage Policy - Stretched Lite	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	VM Encryption Policy	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	Management Storage policy - Encryption	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	Management Storage Policy - Single Node	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	Management Storage policy - Thin	vm-is-vcenter01.vtme.netapp.com
<input checked="" type="checkbox"/>	AFF_ISCSI_VMSP	vm-is-vcenter01.vtme.netapp.com
<input type="checkbox"/>	Host-local PMem Default Storage Policy	vm-is-vcenter01.vtme.netapp.com
<input checked="" type="checkbox"/>	1	

14 items

Protegendo vVols

As seções a seguir descrevem os procedimentos e as práticas recomendadas para o uso do VMware vVols com storage ONTAP.

Alta disponibilidade do fornecedor VASA

O Fornecedor NetApp VASA é executado como parte do dispositivo virtual juntamente com o plug-in do vCenter e o servidor de API REST (anteriormente conhecido como o console de armazenamento virtual [VSC]) e o adaptador de replicação de armazenamento. Se o provedor VASA não estiver disponível, as VMs que usam vVols continuarão a ser executadas. No entanto, novos datastores vVols não podem ser criados e o vVols não pode ser criado ou vinculado pelo vSphere. Isso significa que as VMs que usam vVols não podem ser ativadas, uma vez que o vCenter não poderá solicitar a criação da vVol de swap. E as VMs em execução não podem usar o vMotion para migração para outro host porque os vVols não podem ser vinculados ao novo host.

O VASA Provider 7,1 e posterior suportam novos recursos para garantir que os serviços estejam disponíveis quando necessário. Inclui novos processos watchdog que monitorizam o fornecedor VASA e os serviços de base de dados integrados. Se detectar uma falha, ele atualiza os arquivos de log e reinicia os serviços automaticamente.

Proteção adicional deve ser configurada pelo administrador do vSphere usando os mesmos recursos de disponibilidade usados para proteger outras VMs de missão crítica contra falhas no software, hardware do host e rede. Nenhuma configuração adicional é necessária no dispositivo virtual para usar esses recursos; basta configurá-los usando abordagens padrão do vSphere. Eles foram testados e são suportados pelo NetApp.

O vSphere High Availability é facilmente configurado para reiniciar uma VM em outro host no cluster de host em caso de falha. A tolerância a falhas do vSphere fornece maior disponibilidade criando uma VM secundária que é continuamente replicada e pode assumir o controle a qualquer momento. Informações adicionais sobre esses recursos estão disponíveis no ["Ferramentas do ONTAP para documentação do VMware vSphere \(Configurar alta disponibilidade para ferramentas do ONTAP\)"](#), bem como na documentação do VMware vSphere (procure disponibilidade do vSphere em ESXi e vCenter Server).

O ONTAP provedor VASA faz o backup automático da configuração do vVols em tempo real para sistemas ONTAP gerenciados, onde as informações do vVols são armazenadas nos metadados do FlexVol volume. Caso o ONTAP Tools Appliance fique indisponível por qualquer motivo, você pode implantar uma nova ferramenta de forma fácil e rápida e importar a configuração. Consulte este artigo da KB para obter mais informações sobre as etapas de recuperação do provedor VASA:

["Como executar uma recuperação de desastres do provedor VASA - Guia de resolução"](#)

Replicação do vVols

Muitos clientes da ONTAP replicam seus armazenamentos de dados tradicionais para sistemas de storage secundário usando o NetApp SnapMirror e, em seguida, usam o sistema secundário para recuperar VMs individuais ou um local inteiro em caso de desastre. Na maioria dos casos, os clientes usam uma ferramenta de software para gerenciar isso, como um produto de software de backup, como o plug-in do NetApp SnapCenter para VMware vSphere ou uma solução de recuperação de desastres, como o Gerenciador de recuperação de site da VMware (juntamente com o adaptador de replicação de armazenamento nas ferramentas do ONTAP).

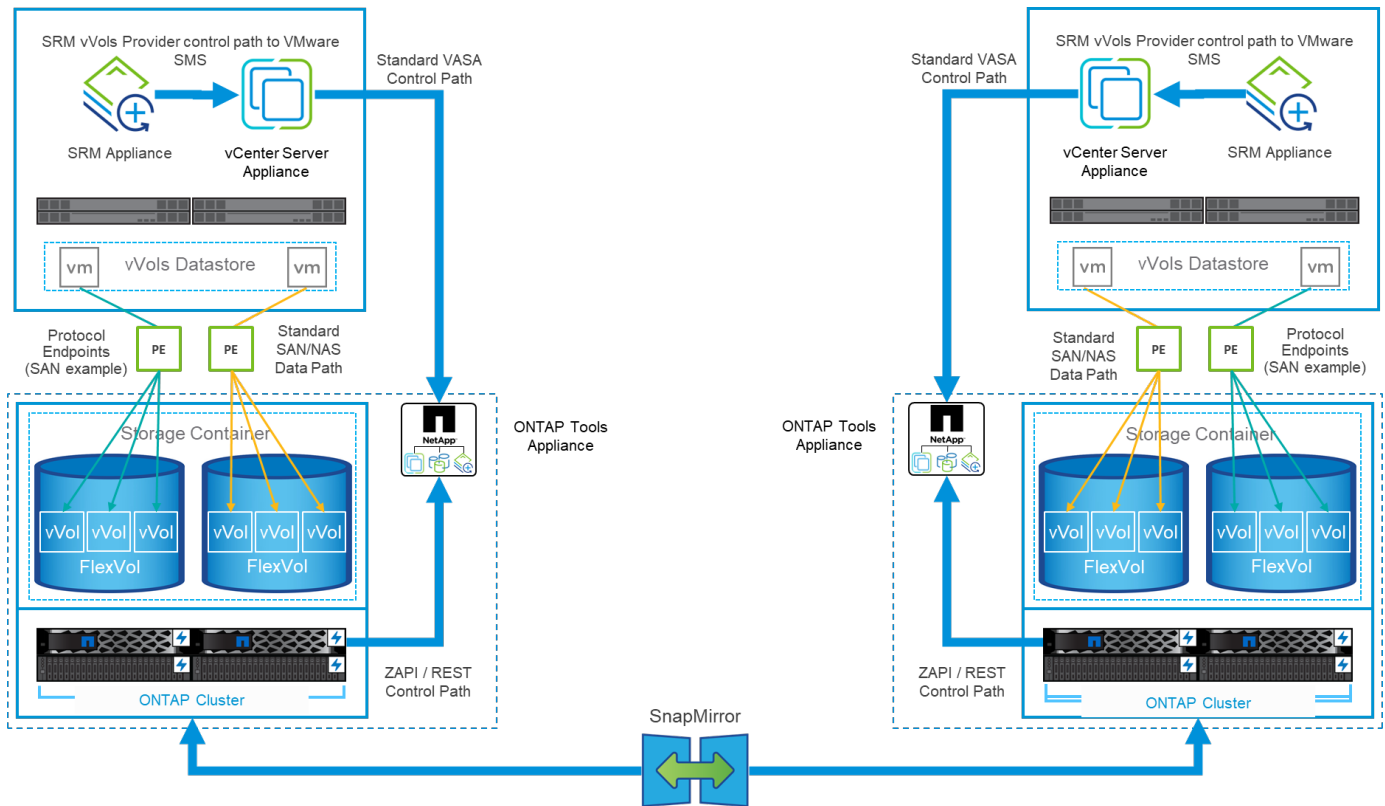
Esse requisito para uma ferramenta de software é ainda mais importante para gerenciar a replicação do vVols. Embora alguns aspectos possam ser gerenciados por recursos nativos (por exemplo, os snapshots gerenciados do vVols da VMware são descarregados para o ONTAP, que usa clones rápidos e eficientes de arquivos ou LUN), em geral, a orquestração é necessária para gerenciar a replicação e a recuperação. Os metadados sobre o vVols são protegidos pelo ONTAP, bem como pelo Fornecedor VASA, mas é necessário um processamento adicional para usá-los em um local secundário.

As ferramentas do ONTAP 9.7.1 em conjunto com a versão 8,3 do VMware Site Recovery Manager (SRM) adicionou suporte para orquestração do fluxo de trabalho de migração e recuperação de desastres aproveitando a tecnologia NetApp SnapMirror.

Na versão inicial do suporte do SRM com as ferramentas do ONTAP 9.7.1, era necessário pré-criar FlexVols e habilitar a proteção SnapMirror antes de usá-los como volumes de backup para um armazenamento de dados vVols. A partir das ferramentas ONTAP 9,10, esse processo não é mais necessário. Agora, você pode adicionar proteção SnapMirror aos volumes de backup existentes e atualizar suas políticas de storage de VM para aproveitar o gerenciamento baseado em políticas com orquestração e recuperação de desastres e automação integradas ao SRM.

Atualmente, o VMware SRM é a única solução de automação de recuperação de desastres e migração para vVols suportada pelo NetApp, e as ferramentas ONTAP verificarão a existência de um servidor SRM 8,3 ou posterior registrado no vCenter antes de permitir a replicação do vVols, embora seja possível aproveitar as APIS REST do ONTAP Tools para criar seus próprios serviços.

Replicação do vVols com o SRM



Suporte à MetroCluster

Embora as ferramentas do ONTAP não sejam capazes de acionar um switchover do MetroCluster, ele oferece suporte a sistemas NetApp MetroCluster para vVols que suportam volumes em uma configuração uniforme de cluster de storage metro do vSphere (vmsc). O switchover de um sistema MetroCluster é Tratado da maneira normal.

Embora o NetApp SnapMirror Business Continuity (SM-BC) também possa ser usado como base para uma configuração vmsc, ele não é atualmente suportado com vVols.

Consulte estes guias para obter mais informações sobre o NetApp MetroCluster:

["Arquitetura e design da solução IP MetroCluster TR-4689"](#)

["TR-4705 arquitetura e design da solução NetApp MetroCluster"](#)

["VMware KB 2031038 suporte ao VMware vSphere com NetApp MetroCluster"](#)

Visão geral do vVols Backup

Existem várias abordagens para proteger VMs, como usar agentes de backup in-Guest, anexar arquivos de dados da VM a um proxy de backup ou usar APIs definidas como VMware VADP. O vVols pode ser protegido usando os mesmos mecanismos e muitos parceiros da NetApp suportam backups de VM, incluindo vVols.

Como mencionado anteriormente, os snapshots gerenciados do VMware vCenter são descarregados para ONTAP clones de arquivo/LUN rápidos e eficientes em termos de espaço. Eles podem ser usados para backups rápidos e manuais, mas são limitados pelo vCenter a um máximo de 32 snapshots. Você pode usar o vCenter para tirar snapshots e reverter conforme necessário.

A partir do plug-in SnapCenter para VMware vSphere (SCV) 4,6, quando usado em conjunto com as ferramentas ONTAP 9,10 e posteriores, adiciona suporte para backup e recuperação consistentes com falhas

de VMs baseadas em vVols, utilizando snapshots ONTAP FlexVol volume com suporte para replicação SnapMirror e SnapVault. Suporte para até 1023 instantâneos por volume. O SCV também pode armazenar mais snapshots com retenção mais longa em volumes secundários usando o SnapMirror com uma política de cofre-espelho.

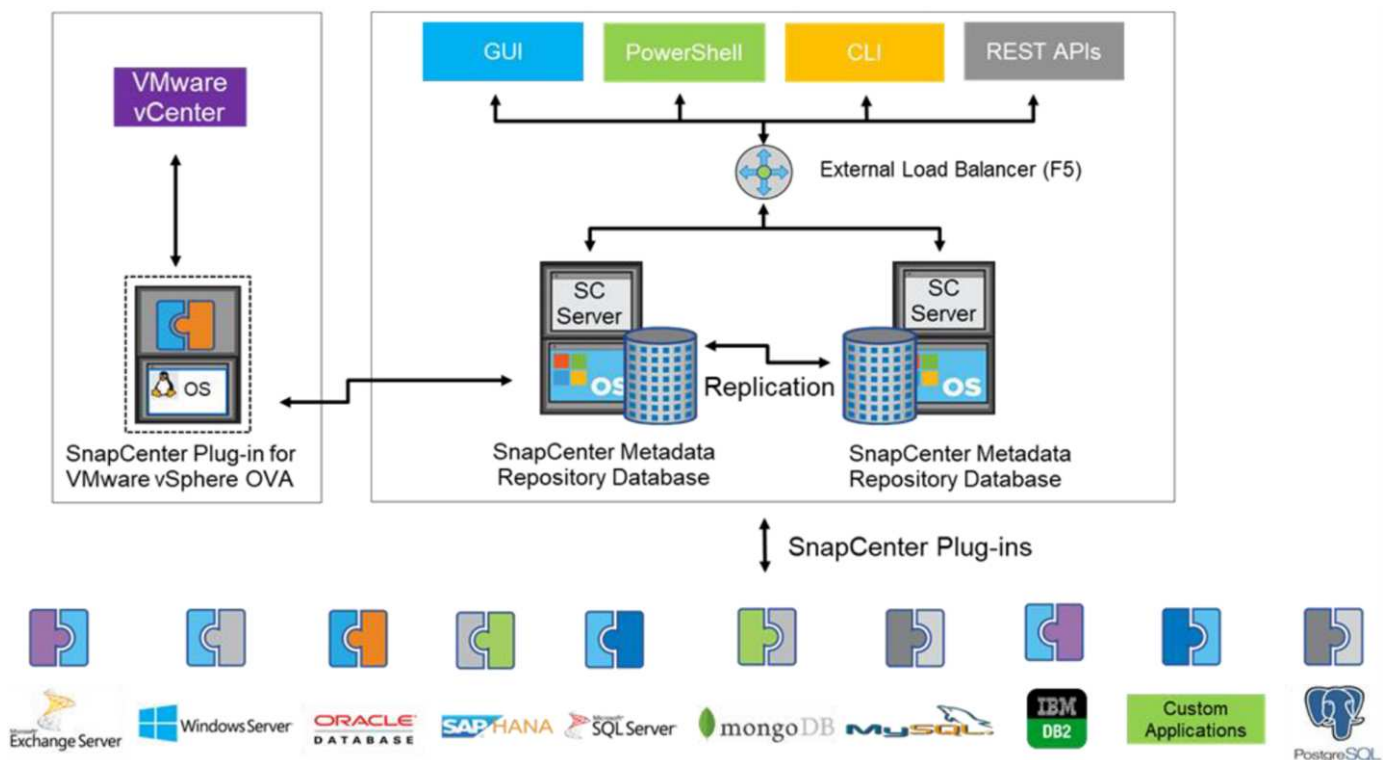
O suporte ao vSphere 8,0 foi introduzido com o SCV 4,7, que usou uma arquitetura de plug-in local isolada. O suporte ao vSphere 8.0U1 foi adicionado ao SCV 4,8, que fez a transição completa para a nova arquitetura de plug-in remoto.

Backup do vVols com o plug-in do SnapCenter para VMware vSphere

Com o NetApp SnapCenter, agora você pode criar grupos de recursos para vVols com base em tags e/ou pastas para aproveitar automaticamente os snapshots baseados no FlexVol do ONTAP para VMs baseadas em vVols. Isso permite que você defina serviços de backup e recuperação que protegerão as VMs automaticamente à medida que elas são provisionadas dinamicamente em seu ambiente.

O plug-in do SnapCenter para VMware vSphere é implantado como um dispositivo autônomo registrado como uma extensão do vCenter, gerenciado por meio da IU do vCenter ou por meio de APIs REST para automação de serviços de backup e recuperação.

Arquitetura da SnapCenter



Como os outros plugins do SnapCenter ainda não suportam vVols no momento desta escrita, vamos nos concentrar no modelo de implantação independente neste documento.

Como o SnapCenter usa snapshots do ONTAP FlexVol, não há sobrecarga no vSphere, nem há nenhuma penalidade de desempenho como se pode ver nas VMs tradicionais que usam snapshots gerenciados do vCenter. Além disso, como a funcionalidade do SCV é exposta por meio de APIs REST, ele facilita a criação de fluxos de trabalho automatizados usando ferramentas como VMware Aria Automation, Ansible, Terraform e praticamente qualquer outra ferramenta de automação capaz de usar APIs REST padrão.

Para obter informações sobre APIs REST do SnapCenter, consulte ["Visão geral das APIs REST"](#)

Para obter informações sobre o plug-in do SnapCenter para APIs REST do VMware vSphere, consulte ["Plug-in do SnapCenter para APIs REST do VMware vSphere"](#)

Práticas recomendadas

As práticas recomendadas a seguir podem ajudá-lo a aproveitar ao máximo sua implantação do SnapCenter.

- O SCV oferece suporte ao vCenter Server RBAC e ao ONTAP RBAC e inclui funções do vCenter predefinidas que são criadas automaticamente para você quando o plug-in é registrado. Você pode ler mais sobre os tipos compatíveis de RBAC ["aqui."](#)
 - Use a IU do vCenter para atribuir acesso a contas com menos privilégios usando as funções predefinidas descritas ["aqui"](#).
 - Se você usar o SCV com o servidor SnapCenter, deverá atribuir a função *SnapCenterAdmin*.
 - ONTAP RBAC refere-se à conta de usuário usada para adicionar e gerenciar os sistemas de storage usados pela SCV. O ONTAP RBAC não se aplica a backups baseados em vVols. Leia mais sobre ONTAP RBAC e SCV ["aqui"](#).
- Replique seus conjuntos de dados de backup para um segundo sistema usando o SnapMirror para réplicas completas de volumes de origem. Como mencionado anteriormente, você também pode usar políticas de espelhamento de cofre para retenção de dados de backup a longo prazo, independentemente das configurações de retenção de snapshot do volume de origem. Ambos os mecanismos são suportados com vVols.
- Como o SCV também requer ferramentas do ONTAP para a funcionalidade vVols do VMware vSphere, verifique sempre a ferramenta de Matriz de interoperabilidade (IMT) do NetApp para obter compatibilidade de versões específicas
- Se você estiver usando a replicação do vVols com o VMware SRM, lembre-se de sua política de RPO e agendamento de backup
- Crie suas políticas de backup com configurações de retenção que atendem aos objetivos do ponto de restauração (RPOs) definidos pela organização
- Configure as configurações de notificação em seus grupos de recursos para ser notificado sobre o status quando os backups são executados (veja a figura 10 abaixo)

Opções de notificação do grupo de recursos

Edit Resource Group

1. General info & notification

2. Resource

3. Spanning disks

4. Policies

5. Schedules

6. Summary

vCenter Server:

Name:

Description:

Notification:

Email send from:

Email send to:

Email subject:

Latest Snapshot name Enable _recent suffix for latest Snapshot Copy ⓘ

Custom snapshot format: Use custom name format for Snapshot copy

Note that the Plug-in for VMware vSphere cannot do the following:

BACK

NEXT

FINISH

CANCEL

Comece a usar a SCV usando esses documentos

["Saiba mais sobre o plug-in do SnapCenter para VMware vSphere"](#)

["Implante o plug-in do SnapCenter para VMware vSphere"](#)

Solução de problemas

Existem vários recursos de solução de problemas disponíveis com informações adicionais.

Site de suporte da NetApp

Além de uma variedade de artigos da base de conhecimento para produtos de virtualização NetApp, o site de suporte da NetApp também oferece uma página de destino conveniente para o ["Ferramentas do ONTAP para VMware vSphere"](#) produto. Este portal fornece links para artigos, downloads, relatórios técnicos e discussões sobre soluções VMware na Comunidade NetApp. Está disponível em:

["Site de suporte da NetApp"](#)

A documentação adicional da solução está disponível aqui:

["Soluções NetApp para virtualização com VMware da Broadcom"](#)

Solução de problemas do produto

Os vários componentes das ferramentas do ONTAP, como o plug-in do vCenter, o provedor VASA e o adaptador de replicação de armazenamento, estão todos documentados juntos no repositório de documentos do NetApp. No entanto, cada um tem uma subseção separada da base de dados de Conhecimento e pode ter

procedimentos específicos de solução de problemas. Estes abordam os problemas mais comuns que podem ser encontrados com o Fornecedor VASA.

Problemas de UI do provedor VASA

Ocasionalmente, o Cliente Web do vCenter vSphere encontra problemas com os componentes do Serenity, fazendo com que os itens do menu Fornecedor VASA para ONTAP não sejam exibidos. Consulte Resolução de problemas de Registro do provedor VASA no Guia de implantação ou nesta base de conhecimento ["artigo"](#).

Falha no provisionamento do armazenamento de dados do vVols

Ocasionalmente, os serviços do vCenter podem ter tempo limite ao criar o armazenamento de dados do vVols. Para corrigi-lo, reinicie o serviço vmware-sps e reinstale o datastore vVols usando os menus do vCenter (Storage > New datastore). Isso é coberto em falhas de provisionamento de armazenamento de dados do vVols com o vCenter Server 6,5 no Guia de administração.

A atualização do Unified Appliance falha ao montar o ISO

Devido a um bug no vCenter, o ISO usado para atualizar o dispositivo unificado de uma versão para a próxima pode não ser montado. Se o ISO puder ser anexado ao dispositivo no vCenter, siga o processo nesta base de conhecimento ["artigo"](#) para resolver.

VMware Site Recovery Manager com ONTAP

VMware Site Recovery Manager com ONTAP

A ONTAP tem sido uma solução de storage líder para ambientes VMware vSphere desde sua introdução ao data center moderno em 2002, e continua adicionando recursos inovadores para simplificar o gerenciamento e reduzir custos.

Este documento apresenta a solução ONTAP para o VMware Site Recovery Manager (SRM), o software de recuperação de desastres (DR) líder do setor da VMware, incluindo as informações mais recentes sobre produtos e as melhores práticas para simplificar a implantação, reduzir riscos e simplificar o gerenciamento contínuo.



Esta documentação substitui o relatório técnico publicado anteriormente *TR-4900: VMware Site Recovery Manager com o ONTAP*

As práticas recomendadas complementam outros documentos, como guias e ferramentas de compatibilidade. Eles são desenvolvidos com base em testes de laboratório e extensa experiência de campo por engenheiros e clientes da NetApp. Em alguns casos, as melhores práticas recomendadas podem não ser a opção certa para o seu ambiente; no entanto, geralmente são as soluções mais simples que atendem às necessidades da maioria dos clientes.

Este documento se concentra nos recursos das versões recentes do ONTAP 9 quando usado em conjunto com as ferramentas do ONTAP para VMware vSphere 9,12 (que inclui o adaptador de replicação de armazenamento NetApp [SRA] e o provedor VASA [VP]), bem como o VMware Site Recovery Manager 8,7.

Por que usar o ONTAP com o SRM?

As plataformas de gerenciamento de dados NetApp baseadas no ONTAP são algumas das soluções de storage mais adotadas para o SRM. Muitos são os motivos: Uma plataforma de gerenciamento de dados segura, de alto desempenho e unificada (nas e SAN juntos) que fornece ao setor a definição de eficiência de

storage, alocação a vários clientes, controle de qualidade do serviço, proteção de dados com snapshots com uso eficiente de espaço e replicação com o SnapMirror. Todos utilizam a integração nativa de multicloud híbrida para a proteção dos workloads da VMware e uma infinidade de ferramentas de automação e orquestração na ponta dos dedos.

Ao usar o SnapMirror para replicação baseada em array, você aproveita uma das tecnologias mais comprovadas e maduras da ONTAP. O SnapMirror oferece a vantagem de transferências de dados seguras e altamente eficientes, copiando apenas blocos de sistema de arquivos alterados, não VMs inteiras ou armazenamentos de dados. Mesmo esses blocos aproveitam a economia de espaço, como deduplicação, compressão e compactação. Os sistemas ONTAP modernos agora usam SnapMirror independentes da versão, permitindo flexibilidade na seleção dos clusters de origem e destino. O SnapMirror realmente se tornou uma das ferramentas mais poderosas disponíveis para recuperação de desastres.

Não importa se você está usando armazenamentos de dados tradicionais anexados a NFS, iSCSI ou Fibre Channel (agora com suporte para armazenamentos de dados vVols), o SRM oferece uma oferta exclusiva robusta que aproveita o melhor dos recursos do ONTAP para orquestração e Planejamento de recuperação de desastres ou migração de data center.

Como o SRM utiliza o ONTAP 9

O SRM aproveita as tecnologias avançadas de gerenciamento de dados dos sistemas ONTAP integrando-se às ferramentas do ONTAP para VMware vSphere, um dispositivo virtual que inclui três componentes principais:

- O plug-in do vCenter, anteriormente conhecido como Virtual Storage Console (VSC), simplifica o gerenciamento de storage e os recursos de eficiência, aprimora a disponibilidade e reduz os custos de storage e a sobrecarga operacional, independentemente de você estar usando SAN ou nas. Ele usa as práticas recomendadas para provisionar armazenamentos de dados e otimiza as configurações de host ESXi para ambientes de storage de bloco e NFS. Para todos esses benefícios, a NetApp recomenda esse plug-in ao usar o vSphere com sistemas que executam o ONTAP.
- O provedor VASA para ONTAP oferece suporte à estrutura VMware vStorage APIs for Storage Awareness (VASA). O provedor VASA conecta o vCenter Server com o ONTAP para auxiliar no provisionamento e monitoramento do armazenamento de VM. Ele permite o suporte do VMware Virtual volumes (vVols), o gerenciamento de perfis de funcionalidades de storage (incluindo recursos de replicação do vVols) e a performance individual do VM vVols. Ele também fornece alarmes para monitorar a capacidade e conformidade com os perfis. Quando usado em conjunto com o SRM, o Fornecedor VASA para ONTAP permite o suporte a máquinas virtuais baseadas em vVols sem exigir a instalação de um adaptador SRA no servidor SRM.
- O SRA é usado em conjunto com o SRM para gerenciar a replicação de dados de VM entre locais de produção e recuperação de desastres para armazenamentos de dados VMFS e NFS tradicionais e também para testes sem interrupções de réplicas de DR. Ele ajuda a automatizar as tarefas de descoberta, recuperação e reprotção. Inclui um dispositivo de servidor SRA e adaptadores SRA para o servidor SRM do Windows e o dispositivo SRM.

Depois de instalar e configurar os adaptadores SRA no servidor SRM para proteger datastores não vVols e/ou a replicação vVols ativada nas configurações do provedor VASA, você pode iniciar a tarefa de configurar seu ambiente vSphere para recuperação de desastres.

O provedor SRA e VASA fornecem uma interface de comando e controle para o servidor SRM gerenciar os FlexVols ONTAP que contêm suas máquinas virtuais (VMs) VMware, bem como a replicação SnapMirror que os protege.

A partir do SRM 8,3, um novo caminho de controle do SRM vVols Provider foi introduzido no servidor SRM, permitindo que ele se comunique com o vCenter Server e, por meio dele, com o provedor VASA sem precisar

de um SRA. Isso permitiu que o servidor SRM aproveitasse um controle muito mais profundo sobre o cluster ONTAP do que era possível antes, porque o VASA fornece uma API completa para integração estreatamente acoplada.

O SRM pode testar seu plano de recuperação de desastres sem interrupções, usando a tecnologia FlexClone proprietária da NetApp para fazer clones quase instantâneos de seus datastores protegidos no local de recuperação de desastres. O SRM cria uma sandbox para testar com segurança para que sua organização e seus clientes sejam protegidos em caso de um verdadeiro desastre, dando a você confiança na capacidade de suas organizações executar um failover durante um desastre.

No caso de um verdadeiro desastre ou mesmo de uma migração planejada, o SRM permite que você envie quaisquer alterações de última hora para o conjunto de dados por meio de uma atualização final do SnapMirror (se você optar por fazê-lo). Em seguida, ele quebra o espelho e monta o datastore em seus hosts de DR. Nesse ponto, suas VMs podem ser automaticamente ativadas em qualquer ordem de acordo com sua estratégia pré-planejada.

SRM com ONTAP e outros casos de uso: Nuvem híbrida e migração

A integração da sua implantação SRM com os recursos avançados de gerenciamento de dados do ONTAP permite uma escala e desempenho muito melhores quando comparados com as opções de armazenamento local. Mas, além disso, ela oferece a flexibilidade da nuvem híbrida. Com a nuvem híbrida, você economiza dinheiro categorizando blocos de dados não utilizados do array de alta performance para o hyperscaler preferido usando o FabricPool, que pode ser um armazenamento S3 no local, como o NetApp StorageGRID. Você também pode usar o SnapMirror para sistemas baseados em borda com ONTAP Select definido por software ou DR baseado em nuvem usando o Cloud Volumes ONTAP (CVO) ou ["Storage privado da NetApp na Equinix"](#) para Amazon Web Services (AWS), Microsoft Azure e Google Cloud Platform (GCP) para criar uma pilha de serviços de storage, rede e computação totalmente integrada na nuvem.

Depois, você pode executar failover de teste no data center de um fornecedor de serviços de nuvem com espaço físico de storage quase zero graças ao FlexClone. Proteger sua organização agora pode custar menos do que nunca.

O SRM também pode ser usado para executar migrações planejadas utilizando o SnapMirror para transferir eficientemente suas VMs de um data center para outro ou até mesmo dentro do mesmo data center, seja seu, ou por meio de qualquer número de provedores de serviços parceiros da NetApp.

Práticas recomendadas de implantação

As seções a seguir descrevem as práticas recomendadas de implantação com o ONTAP e o VMware SRM.

Layout e segmentação do SVM para SMT

Com o ONTAP, o conceito de máquina virtual de storage (SVM) fornece segmentação rigorosa em ambientes multitenant seguros. Os usuários do SVM em um SVM não podem acessar ou gerenciar recursos de outro. Dessa forma, você pode utilizar a tecnologia ONTAP criando SVMs separadas para diferentes unidades de negócios que gerenciam seus próprios fluxos de trabalho SRM no mesmo cluster para maior eficiência geral de storage.

Considere o gerenciamento do ONTAP usando contas com escopo SVM e LIFs de gerenciamento de SVM para aprimorar não apenas os controles de segurança, mas também a performance. O desempenho é inerentemente maior ao usar conexões com escopo SVM, pois o SRA não é necessário para processar todos os recursos em um cluster inteiro, incluindo recursos físicos. Em vez disso, ele só precisa entender os ativos lógicos que são abstraídos para o SVM específico.

Ao usar apenas protocolos nas (sem acesso SAN), você pode até mesmo aproveitar o novo modo otimizado nas definindo o seguinte parâmetro (observe que o nome é tal porque SRA e VASA usam os mesmos serviços de back-end no dispositivo):

1. Faça login no painel de controle em `https://<IP address>:9083` e clique em interface CLI baseada na Web.
2. Executar o comando `vp updateconfig -key=enable.qtree.discovery -value=true`.
3. Executar o comando `vp updateconfig -key=enable.optimised.sra -value=true`.
4. Executar o comando `vp reloadconfig`.

Implante ferramentas e considerações do ONTAP para vVols

Se você pretende usar o SRM com vVols, você deve gerenciar o storage usando credenciais com escopo de cluster e um LIF de gerenciamento de cluster. Isso ocorre porque o provedor VASA deve entender a arquitetura física subjacente para satisfazer as exigências da política para políticas de armazenamento de VM. Por exemplo, se você tiver uma política que exija armazenamento all-flash, o provedor VASA deve ser capaz de ver quais sistemas são all-flash.

Outra prática recomendada de implantação é nunca armazenar seu dispositivo ONTAP Tools em um datastore vVols que ele está gerenciando. Isso pode levar a uma situação em que você não pode ligar o provedor VASA, porque você não pode criar o vVol swap para o dispositivo porque o dispositivo está offline.

Práticas recomendadas para gerenciamento de sistemas ONTAP 9

Como mencionado anteriormente, você pode gerenciar clusters do ONTAP usando credenciais de escopo do cluster ou SVM e LIFs de gerenciamento. Para um desempenho ideal, você pode considerar o uso de credenciais com escopo SVM sempre que não estiver usando vVols. No entanto, ao fazer isso, você deve estar ciente de alguns requisitos, e que você perde alguma funcionalidade.

- A conta padrão do vsadmin SVM não tem o nível de acesso necessário para executar tarefas de ferramentas do ONTAP. Portanto, você precisa criar uma nova conta SVM.
- Se você estiver usando o ONTAP 9.8 ou posterior, o NetApp recomenda criar uma conta de usuário com menos privilégios RBAC usando o menu usuários do Gerenciador de sistemas do ONTAP juntamente com o arquivo JSON disponível no seu dispositivo de ferramentas do ONTAP em `https://<IP address>:9083/vsc/config/`. Use sua senha de administrador para baixar o arquivo JSON. Isso pode ser usado para contas com escopo de SVM ou cluster.

Se você estiver usando o ONTAP 9.6 ou anterior, use a ferramenta Criador de Usuário RBAC (RUC) disponível no "[Caixa de ferramentas do site de suporte da NetApp](#)".

- Como o plug-in da IU do vCenter, o provedor VASA e o servidor SRA são todos serviços totalmente integrados, você deve adicionar armazenamento ao adaptador SRA no SRM da mesma forma que você adiciona armazenamento nas ferramentas da IU do vCenter para ONTAP. Caso contrário, o servidor SRA pode não reconhecer as solicitações enviadas do SRM através do adaptador SRA.
- A verificação de caminho NFS não é realizada ao usar credenciais com escopo da SVM. Isso ocorre porque a localização física é logicamente abstraída do SVM. No entanto, isso não é motivo de preocupação, já que os sistemas ONTAP modernos não sofrem mais nenhum declínio de desempenho perceptível ao usar caminhos indiretos.
- Economias de espaço agregado devido à eficiência de storage podem não ser relatadas.
- Quando suportado, os espelhos de partilha de carga não podem ser atualizados.

- O log do EMS pode não ser realizado em sistemas ONTAP gerenciados com credenciais de escopo da SVM.

Práticas recomendadas operacionais

As seções a seguir descrevem as práticas recomendadas operacionais para o storage VMware SRM e ONTAP.

Armazenamentos de dados e protocolos

- Se possível, sempre use ferramentas do ONTAP para provisionar armazenamentos de dados e volumes. Isso garante que volumes, caminhos de junção, LUNs, grupos, políticas de exportação e outras configurações sejam configurados de maneira compatível.
- O SRM dá suporte a iSCSI, Fibre Channel e NFS versão 3 com ONTAP 9 ao usar replicação baseada em array por meio do SRA. O SRM não dá suporte à replicação baseada em array para NFS versão 4,1 com datastores tradicionais ou vVols.
- Para confirmar a conectividade, verifique sempre se é possível montar e desmontar um novo datastore de teste no local de DR do cluster do ONTAP de destino. Teste cada protocolo que você pretende usar para a conectividade do datastore. Uma prática recomendada é usar as ferramentas do ONTAP para criar seu datastore de teste, já que ele está fazendo toda a automação do datastore, conforme indicado pelo SRM.
- Os protocolos SAN devem ser homogêneos para cada local. Você pode misturar NFS e SAN, mas os protocolos SAN não devem ser misturados em um local. Por exemplo, você pode usar FCP no local A e iSCSI no local B. você não deve usar FCP e iSCSI no local A. o motivo para isso é que o SRA não cria grupos mistos no local de recuperação e o SRM não filtra a lista de iniciadores fornecida ao SRA.
- Guias anteriores aconselharam a criação de LIF para localidade de dados. Ou seja, monte sempre um datastore usando um LIF localizado no nó que possui fisicamente o volume. Isso não é mais um requisito nas versões modernas do ONTAP 9. Sempre que possível, e se forem dadas credenciais com escopo de cluster, as ferramentas do ONTAP ainda escolherão o balanceamento de carga entre LIFs locais para os dados, mas não será um requisito de alta disponibilidade ou desempenho.
- O ONTAP 9 pode ser configurado para remover automaticamente snapshots para preservar o tempo de atividade em caso de uma condição fora do espaço quando o dimensionamento automático não é capaz de fornecer capacidade de emergência suficiente. A configuração padrão para esse recurso não exclui automaticamente os snapshots criados pelo SnapMirror. Se os snapshots do SnapMirror forem excluídos, o NetApp SRA não poderá reverter e resincronizar a replicação para o volume afetado. Para impedir que o ONTAP exclua snapshots do SnapMirror, configure o recurso snapshot autodelete a ser testado.

```
snap autodelete modify -volume -commitment try
```

- O dimensionamento automático de volume deve ser definido como `grow` para volumes que contêm armazenamentos de dados SAN e `grow_shrink` para armazenamentos de dados NFS. Saiba mais "[configuração de volumes para crescer ou diminuir automaticamente](#)" sobre o .
- O SRM tem melhor desempenho quando o número de datastores e, portanto, os grupos de proteção são minimizados em seus planos de recuperação. Portanto, você deve considerar a otimização para a densidade da VM em ambientes protegidos pelo SRM, onde o rto é de importância fundamental.
- Use o DRS (Distributed Resource Scheduler) para ajudar a equilibrar a carga nos clusters ESXi protegidos e de recuperação. Lembre-se de que, se você planeja fazer o failback, ao executar uma reprotção, os clusters anteriormente protegidos se tornarão os novos clusters de recuperação. O DRS ajudará a equilibrar a colocação em ambas as direções.

- Sempre que possível, evite usar a personalização de IP com o SRM, pois isso pode aumentar seu rto.

Gerenciamento baseado em políticas de storage (SPBM) e vVols

A partir do SRM 8,3, a proteção de VMs usando datastores vVols é suportada. As programações do SnapMirror são expostas às políticas de armazenamento de VM pelo provedor VASA quando a replicação vVols é ativada no menu de configurações de ferramentas do ONTAP, como mostrado nas capturas de tela a seguir.

O exemplo a seguir mostra a capacitação da replicação do vVols.

Manage Capabilities



Enable VASA Provider

vStorage APIs for Storage Awareness (VASA) is a set of application program interfaces (APIs) that enables vSphere vCenter to recognize the capabilities of storage arrays.



Enable vVols replication

Enables replication of vVols when used with VMware Site Recovery Manager 8.3 or later.



Enable Storage Replication Adapter (SRA)

Storage Replication Adapter (SRA) allows VMware Site Recovery Manager (SRM) to integrate with third party storage array technology.

Enter authentication details for VASA Provider and SRA server:

IP address or hostname: 192.168.64.7
Username: Administrator
Password: _____

CANCEL

APPLY

A captura de tela a seguir fornece um exemplo de programações do SnapMirror exibidas no assistente criar política de armazenamento de VM.

Create VM Storage Policy

- 1 Name and description
- 2 Policy structure
- 3 NetApp.clustered.Data.ONTAP.VP...
- 4 Storage compatibility
- 5 Review and finish

NetApp.clustered.Data.ONTAP.VP.vvol rules

Placement **Replication** Tags

Disabled
 Custom

Provider: NetApp.clustered.Data.ONTAP.VP.vvolReplication ▾

Replication ⓘ Asynchronous ▾ REMOVE

Replication Schedule ⓘ [Select Value] ▾ REMOVE

[Select Value]
hourly

CANCEL BACK NEXT

O Fornecedor ONTAP VASA suporta failover para armazenamento dissimilar. Por exemplo, o sistema pode fazer failover do ONTAP Select em um local de borda para um sistema AFF no data center central. Independentemente da semelhança do storage, você sempre deve configurar mapeamentos de políticas de storage e mapeamentos reversos para políticas de armazenamento de VM habilitadas para replicação para garantir que os serviços fornecidos no local de recuperação atendam às expectativas e aos requisitos. A captura de tela a seguir destaca um exemplo de mapeamento de políticas.

New Storage Policy Mappings

- 1 Creation mode
- 2 Recovery storage policies
- 3 Reverse mappings
- 4 Ready to complete

Recovery storage policies

Configure recovery storage policy mappings for one or more storage policies.

Search...

vc1.demo.netapp.com

- Host-local PMem Default Storage Policy
- VC1 Storage Policy *
- VM Encryption Policy
- vSAN Default Storage Policy
- VVol No Requirements Policy

ADD MAPPINGS

vc2.demo.netapp.com

- Host-local PMem Default Storage Policy
- VC2 Storage Policy
- VM Encryption Policy
- vSAN Default Storage Policy

vc1.demo.netapp.com	vc2.demo.netapp.com
<input checked="" type="radio"/> VC1 Storage Policy	<input checked="" type="radio"/> VC2 Storage Policy

1 mapping(s)

CANCEL BACK NEXT

Criar volumes replicados para datastores vVols

Ao contrário dos armazenamentos de dados vVols anteriores, os armazenamentos de dados vVols replicados devem ser criados desde o início com a replicação habilitada e devem usar volumes pré-criados nos sistemas ONTAP com relacionamentos SnapMirror. Isso requer pré-configuração de coisas como peering de cluster e peering SVM. Essas atividades devem ser realizadas pelo administrador do ONTAP, pois isso facilita uma separação estrita de responsabilidades entre aqueles que gerenciam os sistemas ONTAP em vários locais e aqueles que são os principais responsáveis pelas operações do vSphere.

Isso vem com um novo requisito em nome do administrador do vSphere. Como os volumes estão sendo criados fora do escopo das ferramentas do ONTAP, ele não tem conhecimento das alterações feitas pelo administrador do ONTAP até o período de redescoberta agendado regularmente. Por esse motivo, é uma prática recomendada sempre executar a redescoberta sempre que você criar um relacionamento de volume ou SnapMirror para ser usado com o vVols. Basta clicar com o botão direito do rato no anfitrião ou cluster e selecionar Ferramentas do ONTAP > Atualizar dados do anfitrião e armazenamento, como mostrado na seguinte captura de ecrã.

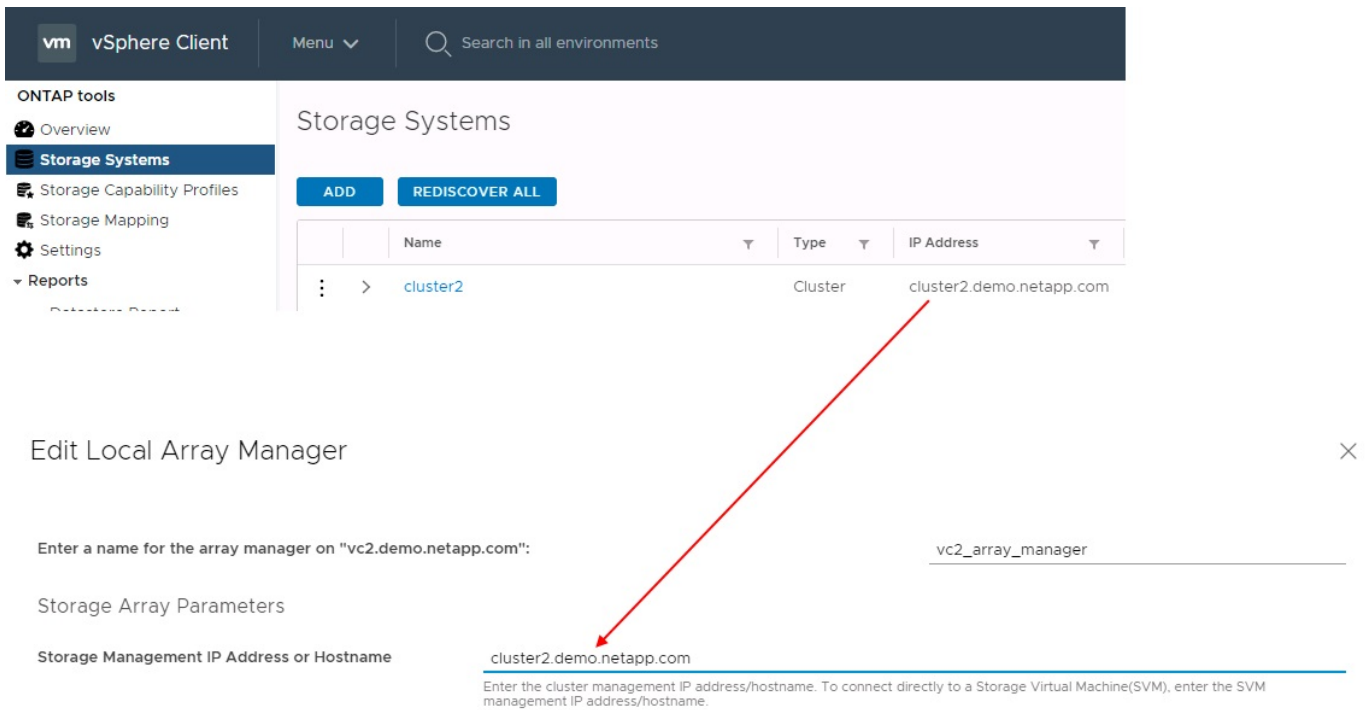


Uma precaução deve ser tomada quando se trata de vVols e SRM. Nunca misture VMs protegidas e desprotegidas no mesmo datastore vVols. O motivo disso é que, quando você usa o SRM para fazer failover para o local de DR, apenas as VMs que fazem parte do grupo de proteção são colocadas on-line no DR. Portanto, quando você reprotége (inverte o SnapMirror da recuperação de desastres para produção novamente), você pode substituir as VMs que não falharam e podem conter dados valiosos.

Sobre os pares de array

Um gerenciador de array é criado para cada par de array. Com as ferramentas SRM e ONTAP, cada emparelhamento de array é feito com o escopo de uma SVM, mesmo que você esteja usando credenciais de cluster. Isso permite segmentar fluxos de trabalho de DR entre locatários com base em quais SVMs eles foram atribuídos a gerenciar. Você pode criar vários gerenciadores de array para um determinado cluster, e eles podem ser assimétricos. Você pode fazer fan-out ou fan-out entre diferentes clusters do ONTAP 9. Por exemplo, você pode fazer a replicação do SVM-A e do SVM-B no Cluster-1 para SVM-C no Cluster-2, SVM-D no Cluster-3 ou vice-versa.

Ao configurar pares de matrizes no SRM, deve sempre adicioná-los no SRM da mesma forma que os adicionou às Ferramentas do ONTAP, ou seja, devem utilizar o mesmo nome de utilizador, palavra-passe e LIF de gestão. Esse requisito garante que o SRA se comunique adequadamente com o array. A captura de tela a seguir ilustra como um cluster pode aparecer nas Ferramentas do ONTAP e como ele pode ser adicionado a um gerenciador de array.



vm vSphere Client Menu Search in all environments

ONTAP tools

- Overview
- Storage Systems**
- Storage Capability Profiles
- Storage Mapping
- Settings
- Reports

Storage Systems

ADD REDISCOVER ALL

Name	Type	IP Address
cluster2	Cluster	cluster2.demo.netapp.com

Edit Local Array Manager ✕

Enter a name for the array manager on "vc2.demo.netapp.com":

Storage Array Parameters

Storage Management IP Address or Hostname

Enter the cluster management IP address/hostname. To connect directly to a Storage Virtual Machine(SVM), enter the SVM management IP address/hostname.

Sobre os grupos de replicação

Os grupos de replicação contêm coleções lógicas de máquinas virtuais que são recuperadas juntas. O provedor VASA de ferramentas do ONTAP cria automaticamente grupos de replicação para você. Como a replicação do ONTAP SnapMirror ocorre no nível do volume, todas as VMs em um volume estão no mesmo grupo de replicação.

Há vários fatores a serem considerados nos grupos de replicação e como você distribui VMs pelos volumes do FlexVol. Agrupar VMs semelhantes no mesmo volume pode aumentar a eficiência de storage com sistemas ONTAP mais antigos que não possuem deduplicação em nível de agregado, mas o agrupamento aumenta o tamanho do volume e reduz a simultaneidade de e/S do volume. O melhor equilíbrio entre performance e eficiência de storage pode ser obtido em sistemas ONTAP modernos, distribuindo máquinas virtuais por volumes FlexVol no mesmo agregado, aproveitando a deduplicação em nível de agregado e obtendo maior paralelização de e/S em vários volumes. Você pode recuperar VMs nos volumes juntos porque um grupo de proteção (discutido abaixo) pode conter vários grupos de replicação. A desvantagem desse layout é que os blocos podem ser transmitidos por cabo várias vezes, porque o volume SnapMirror não leva em conta a deduplicação agregada.

Uma consideração final para grupos de replicação é que cada um é, por sua natureza, um grupo de consistência lógica (não deve ser confundido com grupos de consistência SRM). Isso ocorre porque todas as VMs no volume são transferidas juntas usando o mesmo snapshot. Portanto, se você tiver VMs que precisam ser consistentes umas com as outras, considere armazená-las no mesmo FlexVol.

Sobre grupos de proteção

Os grupos de proteção definem VMs e datastores em grupos que são recuperados juntos do site protegido. O local protegido é onde as VMs configuradas em um grupo de proteção existem durante operações normais de estado estacionário. É importante notar que, embora o SRM possa exibir vários gerenciadores de matriz para um grupo de proteção, um grupo de proteção não pode abranger vários gerenciadores de matriz. Por esse motivo, você não deve estender arquivos de VM entre armazenamentos de dados em diferentes SVMs.

Sobre planos de recuperação

Os planos de recuperação definem quais grupos de proteção são recuperados no mesmo processo. Vários grupos de proteção podem ser configurados no mesmo plano de recuperação. Além disso, para permitir mais opções para a execução de planos de recuperação, um único grupo de proteção pode ser incluído em vários planos de recuperação.

Os planos de recuperação permitem que os administradores do SRM definam fluxos de trabalho de recuperação atribuindo VMs a um grupo de prioridades de 1 (mais alto) a 5 (mais baixo), sendo 3 (médio) o padrão. Dentro de um grupo de prioridade, as VMs podem ser configuradas para dependências.

Por exemplo, sua empresa pode ter um aplicativo essencial para negócios de nível 1 que depende de um servidor Microsoft SQL para seu banco de dados. Então, você decide colocar suas VMs no grupo de prioridades 1. No grupo de prioridade 1, você começa a Planejar o pedido para abrir serviços. Você provavelmente quer que o controlador de domínio do Microsoft Windows seja inicializado antes do servidor Microsoft SQL, que precisaria estar online antes do servidor de aplicativos, e assim por diante. Você adicionaria todas essas VMs ao grupo de prioridade e, em seguida, definiria as dependências porque as dependências se aplicam somente a um determinado grupo de prioridade.

A NetApp recomenda fortemente que você trabalhe com suas equipes de aplicações para entender a ordem das operações necessárias em um cenário de failover e para construir seus planos de recuperação adequadamente.

Failover de teste

Como prática recomendada, sempre execute um failover de teste sempre que uma alteração for feita na configuração de um storage de VM protegido. Isso garante que, no caso de um desastre, você possa confiar que o Site Recovery Manager pode restaurar serviços dentro do destino de rto esperado.

O NetApp também recomenda confirmar ocasionalmente a funcionalidade do aplicativo in-Guest, especialmente depois de reconfigurar o armazenamento de VM.

Quando uma operação de recuperação de teste é executada, uma rede privada de bolhas de teste é criada no host ESXi para as VMs. No entanto, essa rede não é conetada automaticamente a nenhum adaptador de rede físico e, portanto, não fornece conetividade entre os hosts ESXi. Para permitir a comunicação entre VMs que estão sendo executadas em diferentes hosts ESXi durante o teste de DR, uma rede privada física é criada entre os hosts ESXi no local de DR. Para verificar se a rede de teste é privada, a rede de bolhas de teste pode ser separada fisicamente ou usando VLANs ou marcação de VLAN. Essa rede deve ser segregada da rede de produção porque, à medida que as VMs são recuperadas, elas não podem ser colocadas na rede de produção com endereços IP que podem entrar em conflito com os sistemas de produção reais. Quando um plano de recuperação é criado no SRM, a rede de teste criada pode ser selecionada como a rede privada para conetar as VMs durante o teste.

Depois que o teste tiver sido validado e não for mais necessário, execute uma operação de limpeza. A limpeza em execução retorna as VMs protegidas ao seu estado inicial e redefine o plano de recuperação para o estado Pronto.

Considerações sobre failover

Há várias outras considerações quando se trata de falhar em um local, além da ordem de operações mencionada neste guia.

Um problema que você pode ter que lidar com as diferenças de rede entre sites. Alguns ambientes podem ser capazes de usar os mesmos endereços IP de rede no local principal e no local de DR. Essa capacidade é referida como uma LAN virtual (VLAN) estendida ou configuração de rede estendida. Outros ambientes

podem ter um requisito para usar endereços IP de rede diferentes (por exemplo, em VLANs diferentes) no local principal em relação ao local de DR.

A VMware oferece várias maneiras de resolver esse problema. Por um lado, tecnologias de virtualização de rede como o VMware NSX-T Data Center abstraem toda a pilha de rede das camadas 2 a 7 do ambiente operacional, permitindo soluções mais portáteis. Saiba mais "[Opções NSX-T com SRM](#)" sobre o .

O SRM também lhe dá a capacidade de alterar a configuração de rede de uma VM à medida que ela é recuperada. Essa reconfiguração inclui configurações como endereços IP, endereços de gateway e configurações de servidor DNS. Diferentes configurações de rede, que são aplicadas a VMs individuais à medida que são recuperadas, podem ser especificadas nas configurações da propriedade de uma VM no plano de recuperação.

Para configurar o SRM para aplicar diferentes configurações de rede a várias VMs sem ter que editar as propriedades de cada uma no plano de recuperação, a VMware fornece uma ferramenta chamada DR-ip-Customizer. Saiba como usar este utilitário, "[Documentação da VMware](#)" consulte .

Reproteger

Após uma recuperação, o local de recuperação se torna o novo local de produção. Como a operação de recuperação quebrou a replicação do SnapMirror, o novo local de produção não fica protegido de nenhum desastre futuro. Uma prática recomendada é proteger o novo local de produção para outro local imediatamente após uma recuperação. Se o local de produção original estiver operacional, o administrador da VMware poderá usar o local de produção original como um novo local de recuperação para proteger o novo local de produção, invertendo efetivamente o sentido de proteção. A reprotção está disponível apenas em falhas não catastróficas. Portanto, os vCenter Servers originais, os servidores ESXi, os servidores SRM e os bancos de dados correspondentes devem ser eventualmente recuperáveis. Se eles não estiverem disponíveis, um novo grupo de proteção e um novo plano de recuperação devem ser criados.

Failback

Uma operação de failback é fundamentalmente um failover em uma direção diferente do anterior. Como prática recomendada, você verifica se o site original está de volta aos níveis aceitáveis de funcionalidade antes de tentar failback ou, em outras palavras, failover para o site original. Se o local original ainda estiver comprometido, você deve atrasar o failback até que a falha seja suficientemente remediada.

Outra prática recomendada de failback é sempre executar um failover de teste após concluir a reprotção e antes de fazer seu failback final. Isso verifica se os sistemas no local original podem concluir a operação.

Reproteger o site original

Após o failback, você deve confirmar com todas as partes interessadas que seus serviços foram devolvidos ao normal antes de executar o reprotect novamente,

A execução do reprotect After failback coloca essencialmente o ambiente de volta ao estado em que estava no início, com a replicação do SnapMirror sendo executada novamente do local de produção para o local de recuperação.

Topologias de replicação

No ONTAP 9, os componentes físicos de um cluster são visíveis para os administradores de cluster, mas não são visíveis diretamente para os aplicativos e hosts que usam o cluster. Os componentes físicos fornecem um pool de recursos compartilhados a partir do qual os recursos lógicos do cluster são construídos. As aplicações e os hosts

acessam dados somente por meio de SVMs que contêm volumes e LIFs.

Cada SVM do NetApp é Tratado como um array no VMware vCenter Site Recovery Manager. O SRM é compatível com certos layouts de replicação de array a array (ou SVM para SVM).

Uma única VM não pode possuir dados – Virtual Machine Disk (VMDK) ou RDM – em mais de um array SRM pelos seguintes motivos:

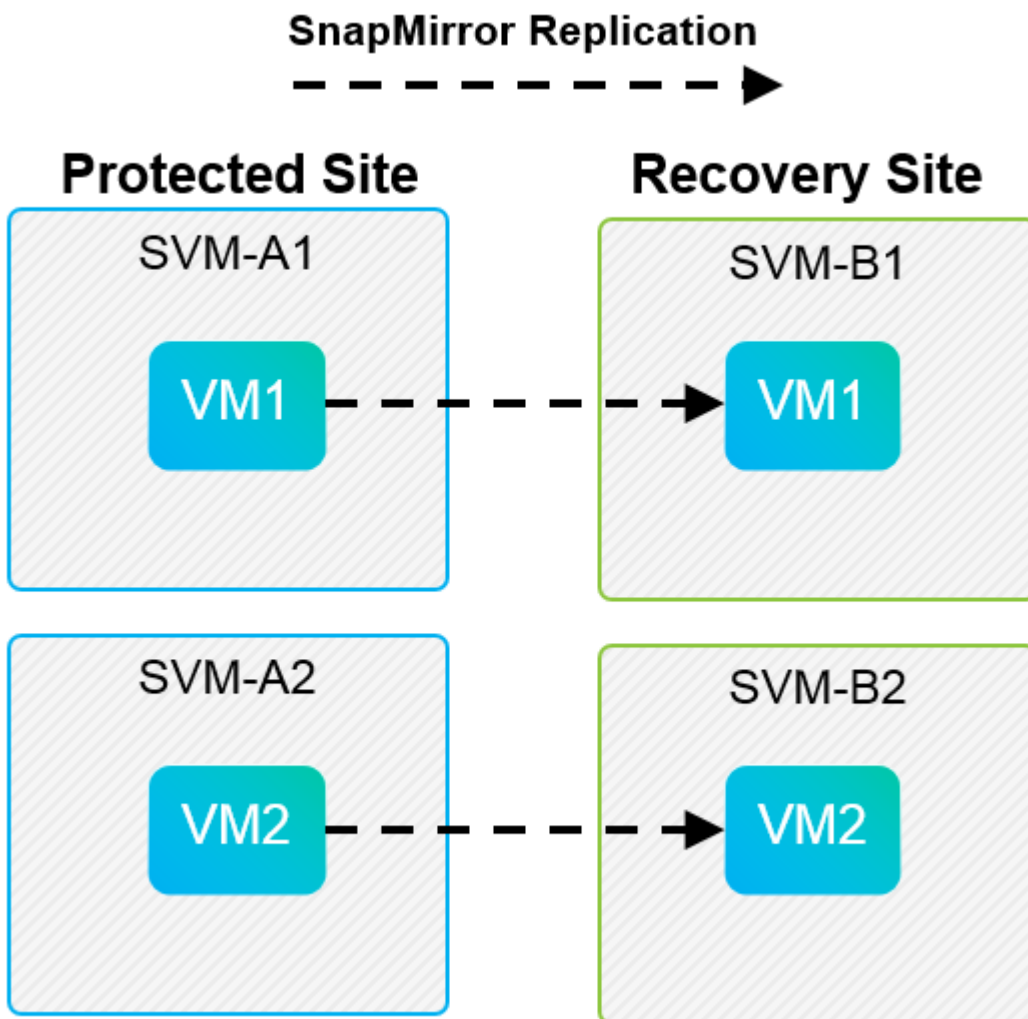
- O SRM vê apenas o SVM, e não um controlador físico individual.
- Um SVM pode controlar LUNs e volumes que abrangem vários nós em um cluster.

Prática recomendada

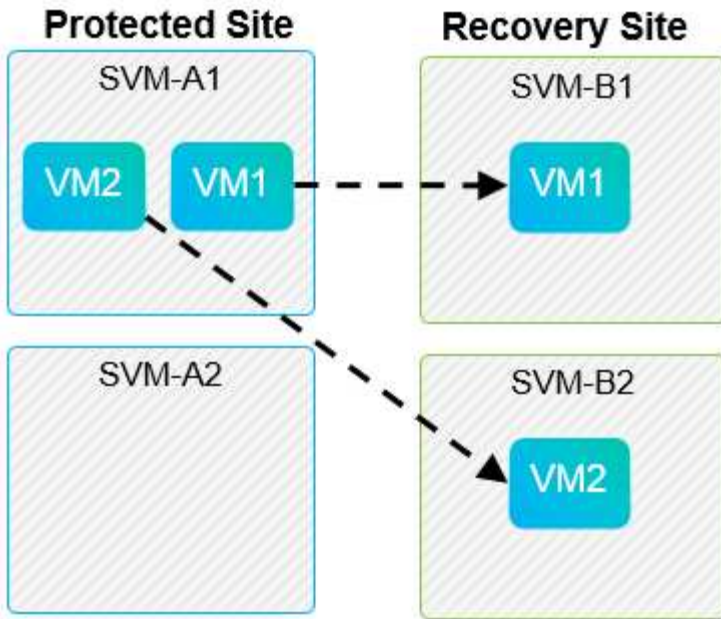
Para determinar a capacidade de suporte, tenha em mente esta regra: Para proteger uma VM usando o SRM e o NetApp SRA, todas as partes da VM devem existir em apenas uma SVM. Esta regra aplica-se tanto no local protegido como no local de recuperação.

Layouts SnapMirror suportados

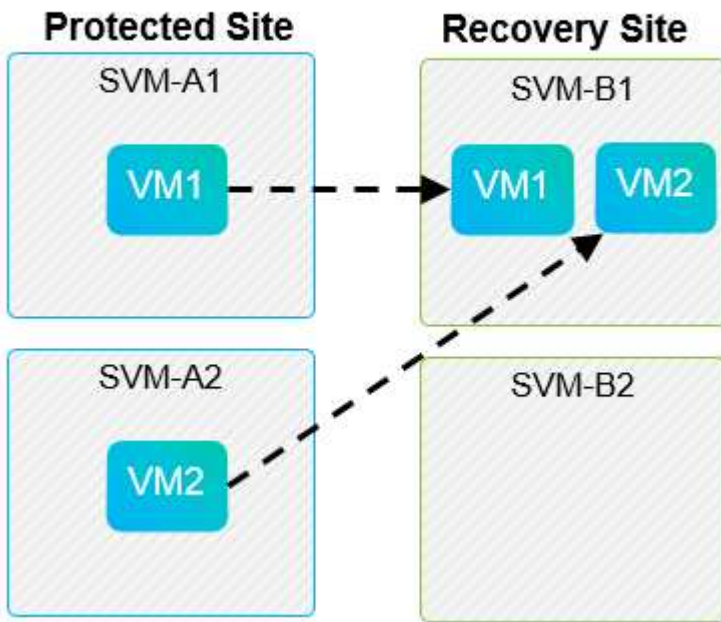
As figuras a seguir mostram os cenários de layout de relacionamento do SnapMirror que o SRM e o SRA suportam. Cada VM nos volumes replicados possui dados em apenas um array SRM (SVM) em cada local.

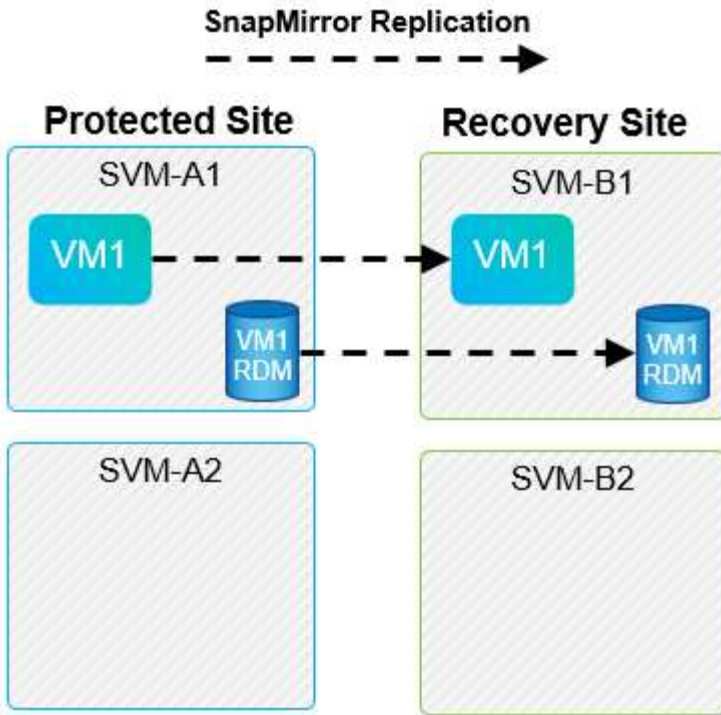


SnapMirror Replication
----->



SnapMirror Replication
----->





Layouts do Array Manager compatíveis

Quando você usa a replicação baseada em array (ABR) no SRM, os grupos de proteção são isolados a um único par de array, como mostrado na captura de tela a seguir. Neste cenário, SVM1 e SVM2 são percorridos com SVM3 e SVM4 no local de recuperação. No entanto, você pode selecionar apenas um dos dois pares de matrizes ao criar um grupo de proteção.

New Protection Group

- 1 Name and direction
- 2 Type
- 3 Datastore groups
- 4 Recovery plan
- 5 Ready to complete

Type ×

Select the type of protection group you want to create:

- Datastore groups (array-based replication)**
Protect all virtual machines which are on specific datastores.
- Individual VMs (vSphere Replication)**
Protect specific virtual machines, regardless of the datastores.
- Virtual Volumes (vVol replication)**
Protect virtual machines which are on replicated vVol storage.
- Storage policies (array-based replication)**
Protect virtual machines with specific storage policies.

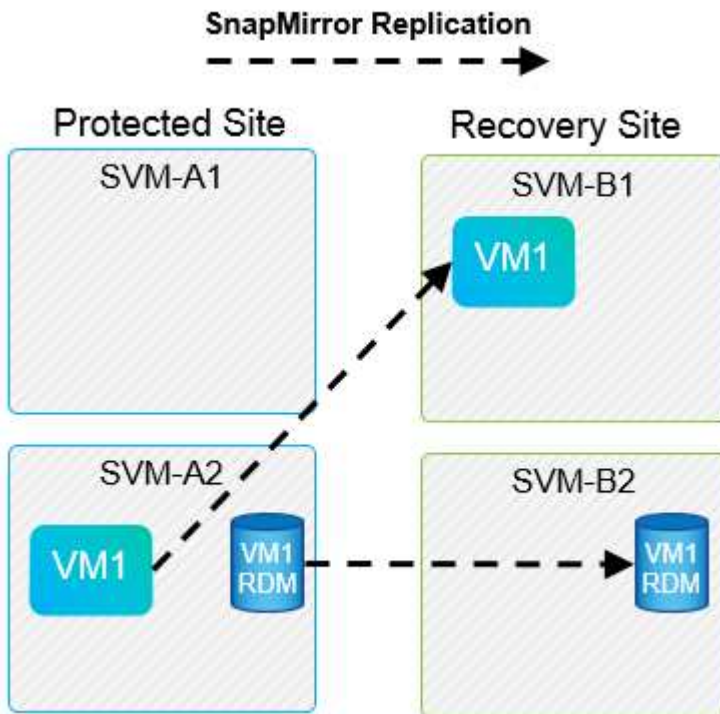
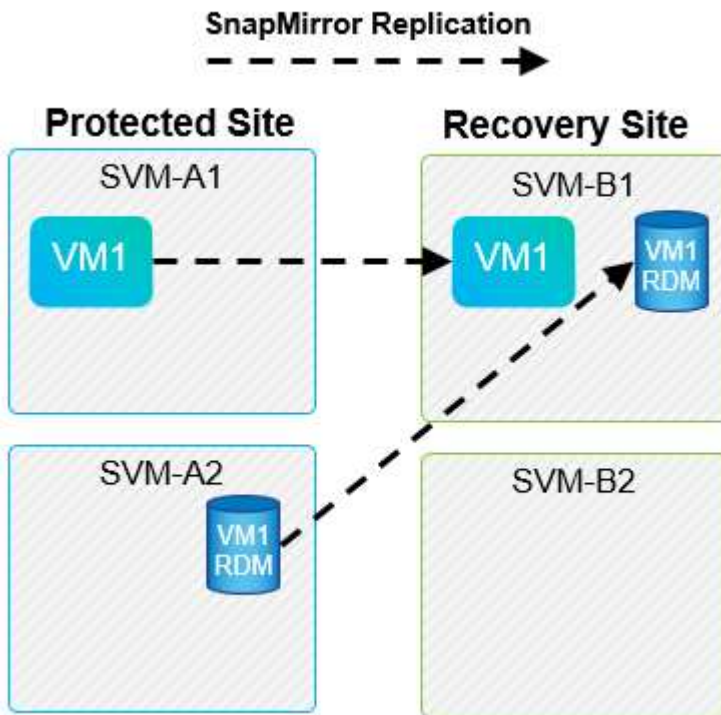
Select array pair

	Array Pair	Array Manager Pair
<input type="radio"/>	✓ cluster1:svm1 ↔ cluster2:svm2	vc1 array manager ↔ vc2 array manager
<input type="radio"/>	✓ cluster1:svm3 ↔ cluster2:svm4	vc1 trad datastores ↔ vc2 trad datastores

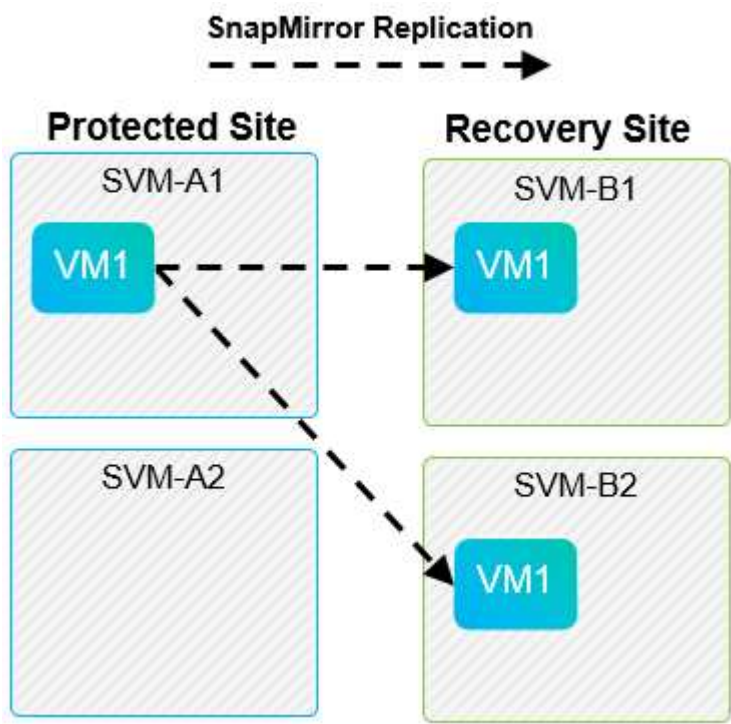
CANCEL
BACK
NEXT

Esquemas não suportados

Configurações não suportadas têm dados (VMDK ou RDM) em vários SVMs que são de propriedade de uma VM individual. Nos exemplos mostrados nas figuras a seguir, VM1 não pode ser configurado para proteção com SRM VM1 porque tem dados em dois SVMs.

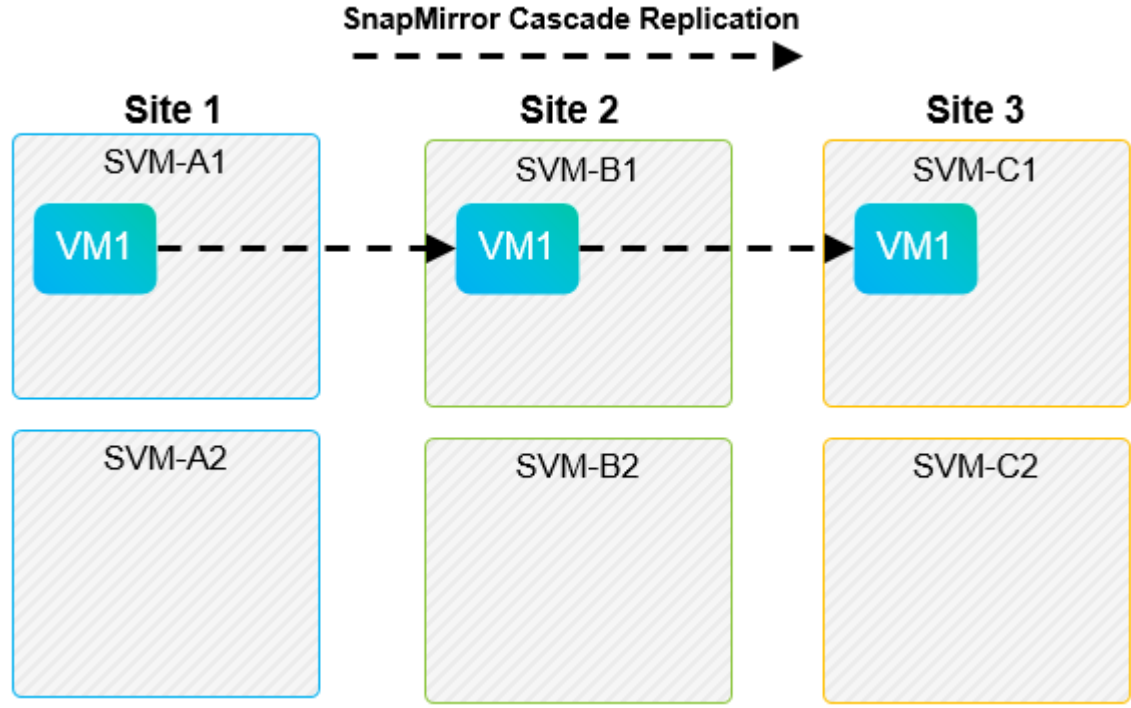


Qualquer relação de replicação na qual um volume de NetApp individual é replicado de uma SVM de origem para vários destinos no mesmo SVM ou em SVMs diferentes é chamada de fan-out do SnapMirror. Fan-out não é suportado com SRM. No exemplo mostrado na figura a seguir, VM1 não pode ser configurado para proteção no SRM porque ele é replicado com o SnapMirror para dois locais diferentes.



Cascata de SnapMirror

O SRM não oferece suporte a cascata de relacionamentos SnapMirror, nas quais um volume de origem é replicado para um volume de destino e esse volume de destino também é replicado com o SnapMirror para outro volume de destino. No cenário mostrado na figura a seguir, o SRM não pode ser usado para failover entre sites.



SnapMirror e SnapVault

O software NetApp SnapVault permite o backup baseado em disco de dados empresariais entre sistemas de

storage NetApp. O SnapVault e o SnapMirror podem coexistir no mesmo ambiente; no entanto, o SRM suporta o failover apenas das relações SnapMirror.



O NetApp SRA suporta o `mirror-vault` tipo de política.

SnapVault foi reconstruído a partir do zero para ONTAP 8,2. Embora antigos usuários do Data ONTAP 7-Mode devam encontrar semelhanças, grandes melhorias foram feitas nesta versão do SnapVault. Um grande avanço é a capacidade de preservar eficiências de storage de dados primários durante transferências SnapVault.

Uma mudança arquitetônica importante é que o in ONTAP 9 replica no nível de volume em vez de no nível de qtree, como é o caso do SnapVault 7-Mode SnapVault. Essa configuração significa que a origem de um relacionamento do SnapVault deve ser um volume e esse volume deve ser replicado para seu próprio volume no sistema secundário do SnapVault.

Em um ambiente em que o SnapVault é usado, os snapshots nomeados especificamente são criados no sistema de storage primário. Dependendo da configuração implementada, os instantâneos nomeados podem ser criados no sistema principal por um agendamento do SnapVault ou por um aplicativo como o NetApp Active IQ Unified Manager. Os instantâneos nomeados que são criados no sistema primário são replicados para o destino SnapMirror e, a partir daí, são abobadados para o destino SnapVault.

Um volume de origem pode ser criado em uma configuração em cascata na qual um volume é replicado para um destino SnapMirror no local de DR e, a partir daí, é abobadado para um destino SnapVault. Um volume de origem também pode ser criado em uma relação de fan-out em que um destino é um destino SnapMirror e o outro destino é um destino SnapVault. No entanto, o SRA não reconfigura automaticamente a relação do SnapVault para usar o volume de destino do SnapMirror como a origem do Vault quando ocorre failover ou reversão de replicação do SRM.

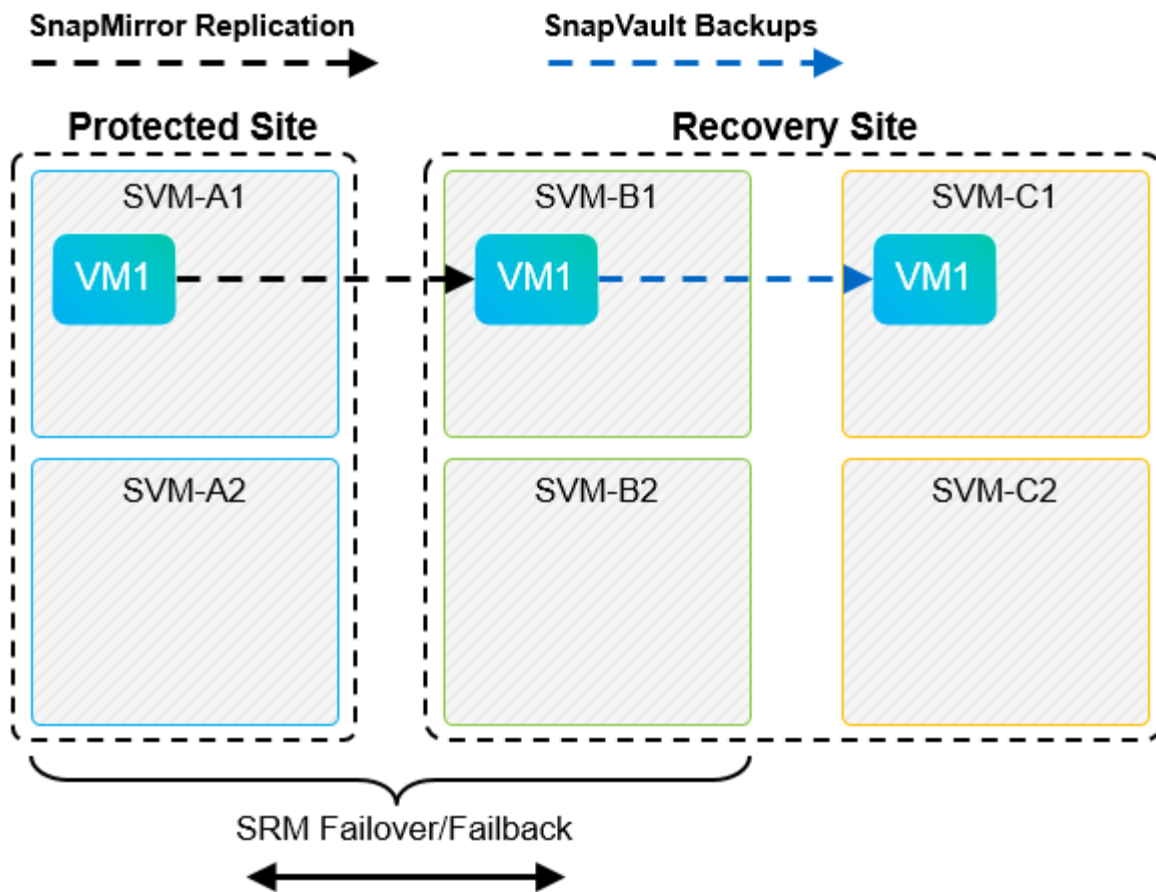
Para obter as informações mais recentes sobre o SnapMirror e o SnapVault for ONTAP 9, consulte ["TR-4015 Guia de práticas recomendadas de configuração do SnapMirror para ONTAP 9."](#)

Prática recomendada

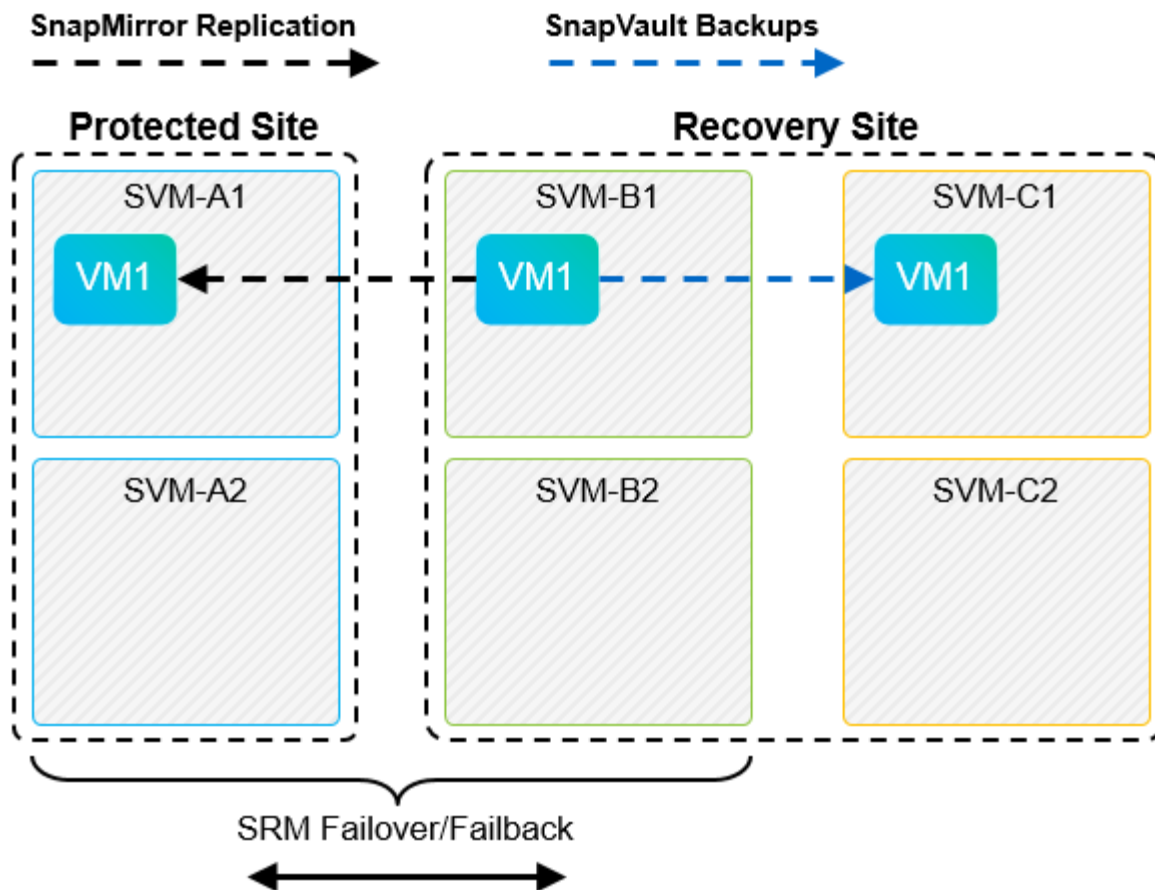
Se o SnapVault e o SRM forem usados no mesmo ambiente, a NetApp recomenda o uso de uma configuração em cascata SnapMirror to SnapVault na qual os backups do SnapVault normalmente são executados a partir do destino do SnapMirror no local de DR. Em caso de desastre, essa configuração torna o site primário inacessível. Manter o destino do SnapVault no local de recuperação permite que os backups do SnapVault sejam reconfigurados após o failover para que os backups do SnapVault possam continuar operando no local de recuperação.

Em um ambiente VMware, cada datastore tem um identificador exclusivo universal (UUID) e cada VM tem um ID de objeto gerenciado exclusivo (MOID). Essas IDs não são mantidas pelo SRM durante o failover ou failback. Como os UUIDs do datastore e os MOIDs de VM não são mantidos durante o failover pelo SRM, todos os aplicativos que dependem desses IDs devem ser reconfigurados após o failover do SRM. Um aplicativo de exemplo é o NetApp Active IQ Unified Manager, que coordena a replicação do SnapVault com o ambiente vSphere.

A figura a seguir mostra uma configuração em cascata SnapMirror to SnapVault. Se o destino do SnapVault estiver no local de DR ou em um local terciário que não seja afetado por uma interrupção no local primário, o ambiente poderá ser reconfigurado para permitir que os backups continuem após o failover.



A figura a seguir mostra a configuração depois que o SRM foi usado para reverter a replicação do SnapMirror de volta para o local principal. O ambiente também foi reconfigurado de modo que os backups do SnapVault estão ocorrendo a partir do que é agora a fonte SnapMirror. Esta configuração é uma configuração de fan-out do SnapMirror SnapVault.



Depois que o SRM executa o failback e uma segunda reversão das relações do SnapMirror, os dados de produção estão de volta ao local principal. Agora, esses dados estão protegidos da mesma maneira que antes do failover para o local de recuperação de desastres, por meio de backups SnapMirror e SnapVault.

Uso de Qtrees em ambientes do Site Recovery Manager

Qtrees são diretórios especiais que permitem a aplicação de cotas de sistema de arquivos para nas. O ONTAP 9 permite a criação de qtrees, e qtrees podem existir em volumes replicados com o SnapMirror. No entanto, o SnapMirror não permite replicação de qtrees individuais ou replicação em nível de qtree. Toda a replicação do SnapMirror está apenas no nível do volume. Por esta razão, o NetApp não recomenda o uso de qtrees com SRM.

Ambientes FC e iSCSI mistos

Com os protocolos SAN compatíveis (FC, FCoE e iSCSI), o ONTAP 9 fornece serviços LUN, ou seja, a capacidade de criar e mapear LUNs para hosts conectados. Como o cluster consiste em vários controladores, há vários caminhos lógicos gerenciados pela e/S multipath em qualquer LUN individual. O acesso de unidade lógica assimétrica (ALUA) é usado nos hosts para que o caminho otimizado para um LUN seja selecionado e seja ativado para transferência de dados. Se o caminho otimizado para qualquer LUN mudar (por exemplo, porque o volume que contém é movido), o ONTAP 9 reconhece e ajusta-se automaticamente para essa alteração sem interrupções. Se o caminho otimizado ficar indisponível, o ONTAP poderá alternar para qualquer outro caminho disponível sem interrupções.

O VMware SRM e o NetApp SRA suportam o uso do protocolo FC em um local e do protocolo iSCSI no outro local. No entanto, ele não dá suporte a uma combinação de armazenamentos de dados anexados a FC e armazenamentos de dados anexados a iSCSI no mesmo host ESXi ou em hosts diferentes no mesmo cluster. Esta configuração não é suportada com o SRM porque, durante o failover SRM ou failover de teste, o SRM

inclui todos os iniciadores FC e iSCSI nos hosts ESXi na solicitação.

Prática recomendada

O SRM e o SRA oferecem suporte a protocolos FC e iSCSI mistos entre os locais protegidos e de recuperação. No entanto, cada local deve ser configurado com apenas um protocolo, FC ou iSCSI, e não com ambos os protocolos no mesmo local. Se houver um requisito para que os protocolos FC e iSCSI sejam configurados no mesmo local, o NetApp recomenda que alguns hosts usem iSCSI e outros hosts usem FC. O NetApp também recomenda, neste caso, que os mapeamentos de recursos do SRM sejam configurados para que as VMs sejam configuradas para failover em um grupo de hosts ou outro.

Solução de problemas do SRM ao usar a replicação vVols

O fluxo de trabalho dentro do SRM é significativamente diferente ao usar a replicação vVols do que é usado com SRA e datastores tradicionais. Por exemplo, não há conceito de gerenciador de array. Como tal, `discoverarrays` e `discoverdevices` comandos nunca são vistos.

Ao solucionar problemas, é benéfico entender os novos fluxos de trabalho, listados abaixo:

1. `QueryReplicationPeer`: Descobre os acordos de replicação entre dois domínios de falha.
2. `QueryFaultDomain`: Descobre a hierarquia do domínio de falha.
3. `QueryReplicationGroup`: Descobre os grupos de replicação presentes nos domínios de origem ou destino.
4. `SyncReplicationGroup`: Sincroniza os dados entre origem e destino.
5. `QueryPointInTimeReplica`: Descobre as réplicas de ponto no tempo em um destino.
6. `TestFailoverReplicationGroupStart`: Inicia o failover de teste.
7. `TestFailoverReplicationGroupStop`: Termina o failover de teste.
8. `PromoteReplicationGroup`: Promove um grupo atualmente em teste para produção.
9. `PrepareFailoverReplicationGroup`: Prepara-se para uma recuperação de desastres.
10. `FailoverReplicationGroup`: Executa recuperação de desastres.
11. `ReverseReplicateGroup`: Inicia a replicação reversa.
12. `QueryMatchingContainer`: Localiza contentores (junto com hosts ou grupos de replicação) que podem satisfazer uma solicitação de provisionamento com uma determinada política.
13. `QueryResourceMetadata`: Descobre os metadados de todos os recursos do provedor VASA, a utilização de recursos pode ser retornada como uma resposta para a função `queryMatchingContainer`.

O erro mais comum visto ao configurar a replicação do vVols é uma falha ao descobrir as relações do SnapMirror. Isso ocorre porque os volumes e as relações SnapMirror são criadas fora do escopo das Ferramentas do ONTAP. Portanto, é uma prática recomendada sempre garantir que sua relação com o SnapMirror esteja totalmente inicializada e que você tenha executado uma redescoberta nas Ferramentas do ONTAP em ambos os sites antes de tentar criar um armazenamento de dados vVols replicado.

Informações adicionais

Para saber mais sobre as informações descritas neste documento, consulte os seguintes documentos e/ou sites:

- TR-4597: VMware vSphere for ONTAP "<https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vsphere-overview.html>"
- TR-4400: VMware vSphere Virtual volumes com ONTAP "<https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vvols-overview.html>"
- TR-4015 Guia de práticas recomendadas de configuração do SnapMirror para ONTAP 9 <https://www.netapp.com/media/17229-tr4015.pdf?v=127202175503P>
- Criador de usuários RBAC para ONTAP "<https://mysupport.netapp.com/site/tools/tool-eula/rbac>"
- Ferramentas do ONTAP para recursos do VMware vSphere "<https://mysupport.netapp.com/site/products/all/details/otv/docsandkb-tab>"
- Documentação do VMware Site Recovery Manager "<https://docs.vmware.com/en/Site-Recovery-Manager/index.html>"

Consulte o "[Ferramenta de Matriz de interoperabilidade \(IMT\)](#)" no site de suporte da NetApp para validar se as versões exatas de produtos e recursos descritas neste documento são compatíveis com o seu ambiente específico. O NetApp IMT define os componentes e versões do produto que podem ser usados para construir configurações compatíveis com o NetApp. Os resultados específicos dependem da instalação de cada cliente de acordo com as especificações publicadas.

Cluster de armazenamento do vSphere Metro com o ONTAP

Cluster de armazenamento do vSphere Metro com o ONTAP

O hypervisor vSphere líder do setor da VMware pode ser implantado como um cluster estendido chamado vSphere Metro Storage Cluster (vMSC).

As soluções vmsc são suportadas com o NetApp MetroCluster e o SnapMirror ativo Sync (anteriormente conhecido como SnapMirror Business Continuity, ou SMBC) e fornecem continuidade de negócios avançada se um ou mais domínios de falha sofrerem uma interrupção total. A resiliência a diferentes modos de falha depende de quais opções de configuração você escolher.

Soluções de disponibilidade contínua para ambientes vSphere

A arquitetura do ONTAP é uma plataforma de storage flexível e dimensionável que fornece serviços SAN (FCP, iSCSI e NVMe-of) e nas (NFS v3 e v4,1) para datastores. Os sistemas de storage NetApp AFF, ASA e FAS usam o sistema operacional ONTAP para oferecer protocolos adicionais para acesso ao storage convidado, como S3 e SMB.

A NetApp MetroCluster usa a função de HA (failover de controladora ou CFO) da NetApp para proteger contra falhas de controladora. Ele também inclui tecnologia SyncMirror local, failover de cluster em desastre (failover de controladora sob demanda ou CFOD), redundância de hardware e separação geográfica para alcançar altos níveis de disponibilidade. O SyncMirror espelha de forma síncrona os dados entre as duas metades da configuração do MetroCluster gravando dados em dois plexos: O Plex local (na gaveta local) fornecendo dados ativamente e o Plex remoto (na gaveta remota) normalmente não fornecendo dados. A redundância de hardware é implementada para todos os componentes MetroCluster, como controladores, armazenamento, cabos, switches (usados com Fabric MetroCluster) e adaptadores.

O NetApp SnapMirror ativo Sync fornece proteção granular do armazenamento de dados com os protocolos FCP e SAN iSCSI, permitindo que você proteja seletivamente apenas workloads de alta prioridade. Ele oferece acesso ativo-ativo simétrico em ambos os sites espelhados, ao contrário do NetApp MetroCluster, que é uma solução de espera ativa.

Para criar um cluster VMware HA/DRS em dois locais, os hosts ESXi são usados e gerenciados por um vCenter Server Appliance (VCSA). As redes de gerenciamento vSphere, vMotion e máquinas virtuais são conectadas por meio de uma rede redundante entre os dois sites. O vCenter Server que gerencia o cluster HA/DRS pode se conectar aos hosts ESXi em ambos os sites e deve ser configurado usando o vCenter HA.

```
https://docs.vmware.com/en/VMware-vSphere/8.0/vsphere-vcenter-esxi-management/GUID-F7818000-26E3-4E2A-93D2-FCDCE7114508.html["Como criar e configurar clusters no vSphere Client"]Consulte para configurar o vCenter HA.
```

Você também deve se referir "[Práticas recomendadas do VMware vSphere Metro Storage Cluster](#)" a .

O que é o vSphere Metro Storage Cluster?

O vSphere Metro Storage Cluster (vMSC) é uma configuração certificada que protege máquinas virtuais (VMs) e contêineres contra falhas. Isso é conseguido usando conceitos de armazenamento estendidos juntamente com clusters de hosts ESXi, que são distribuídos em diferentes domínios de falha, como racks, edifícios, campi ou até mesmo cidades. As tecnologias de storage de sincronização ativa NetApp MetroCluster e SnapMirror são usadas para fornecer proteção RPO igual a 0 ou próximo a 0, respectivamente, para os clusters de host. A configuração do vMSC foi projetada para garantir que os dados estejam sempre disponíveis, mesmo que um "site" físico ou lógico completo falhe. Um dispositivo de armazenamento que faça parte da configuração do vMSC deve ser certificado após passar por um processo de certificação vMSC bem-sucedido. Todos os dispositivos de armazenamento suportados podem ser encontrados no "[Guia de compatibilidade do VMware Storage](#)".

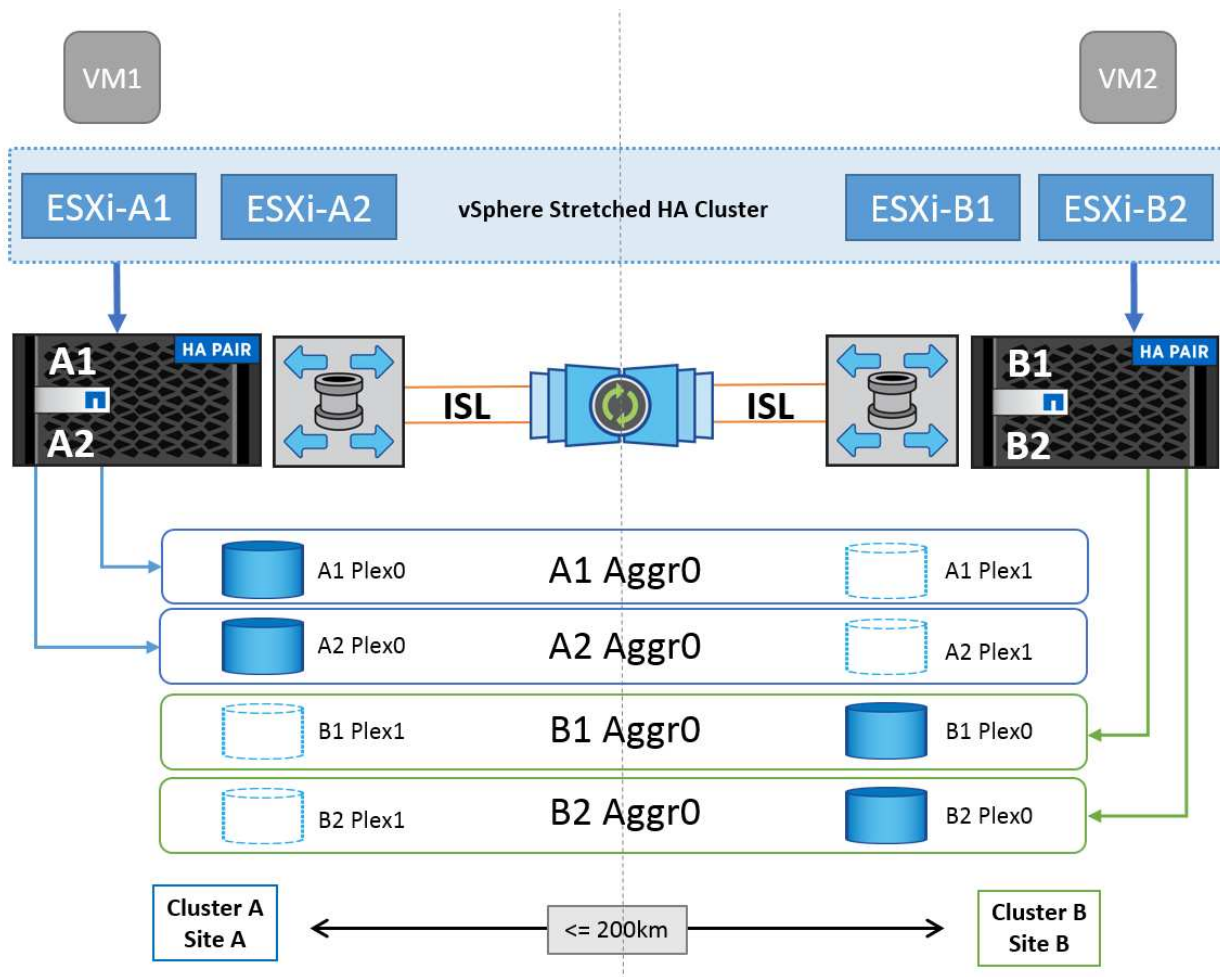
Se você quiser obter mais informações sobre as diretrizes de design do vSphere Metro Storage Cluster, consulte a seguinte documentação:

- "[Suporte ao VMware vSphere com o NetApp MetroCluster](#)"
- "[Suporte ao VMware vSphere com o NetApp SnapMirror Business Continuity](#)" (Agora conhecido como SnapMirror ativo Sync)

Dependendo das considerações de latência, o NetApp MetroCluster pode ser implantado em duas configurações diferentes para uso com o vSphere:

- Stretch MetroCluster
- Fabric MetroCluster

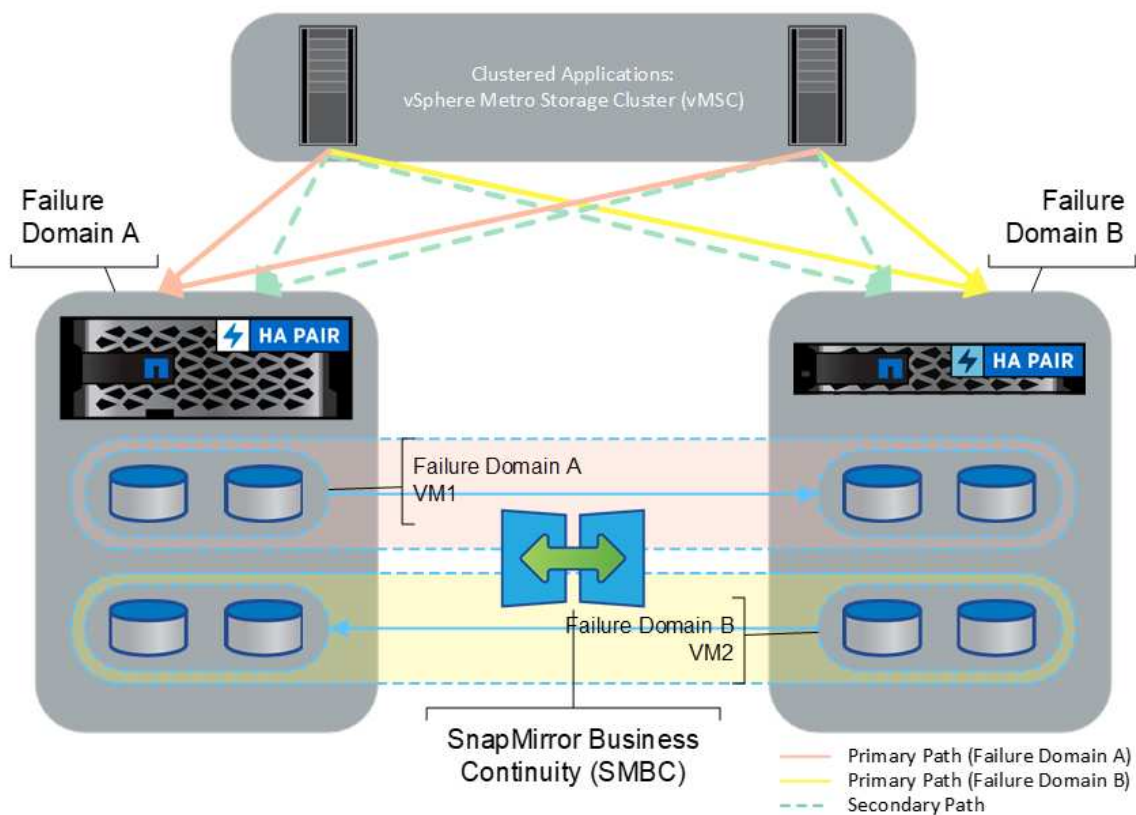
A seguir ilustra um diagrama de topologia de alto nível do Stretch MetroCluster.



<https://www.netapp.com/support-and-training/documentation/metrocluster/> ["Documentação do MetroCluster"] Consulte para obter informações específicas sobre design e implantação do MetroCluster.

O SnapMirror ativo Sync também pode ser implantado de duas maneiras diferentes.

- Assimétrico
- Simétrica (visualização privada no ONTAP 9.14,1)



Para obter mais informações sobre as vantagens do design e implantação simétricos ou assimétricos, consulte o ["Documentação de sincronização ativa do SnapMirror"](#).

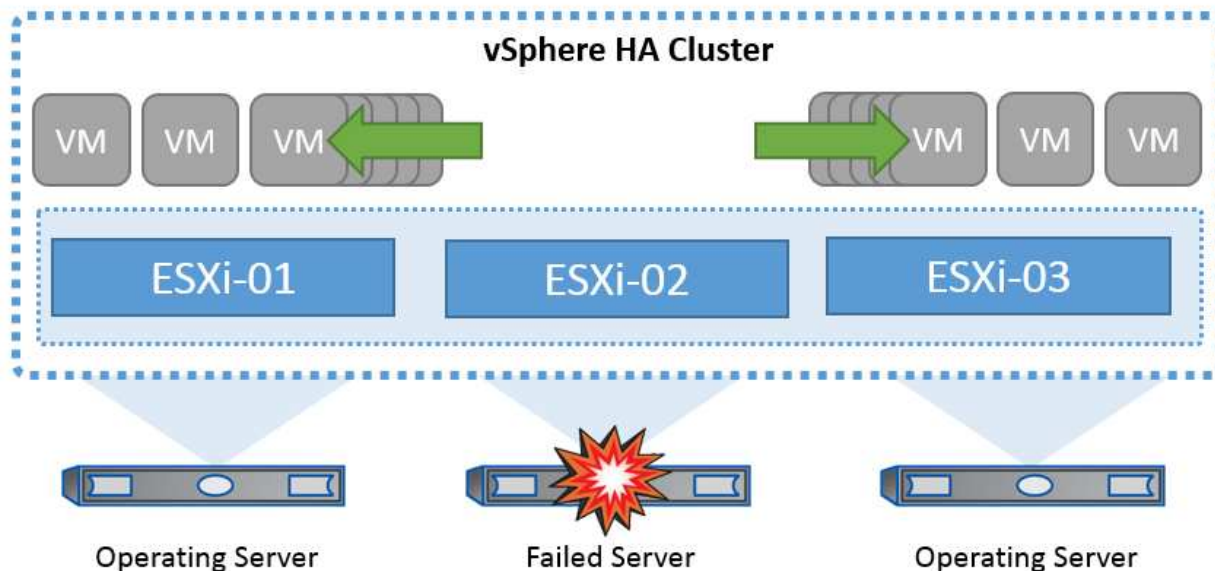
Visão geral da solução VMware vSphere

O vCenter Server Appliance (VCSA) é o poderoso sistema de gerenciamento centralizado e um painel único para o vSphere, que permite que os administradores operem os clusters ESXi com eficiência. Ele facilita as principais funções, como provisionamento de VM, operação vMotion, alta disponibilidade (HA), Agendador de recursos distribuídos (DRS), Tanzu Kubernetes Grid e muito mais. Ele é um componente essencial em ambientes de nuvem da VMware e deve ser projetado pensando na disponibilidade de serviços.

Alta disponibilidade do vSphere

A tecnologia de cluster da VMware agrupa os servidores ESXi em pools de recursos compartilhados para máquinas virtuais e fornece o vSphere High Availability (HA). O vSphere HA oferece alta disponibilidade e é fácil de usar para aplicativos executados em máquinas virtuais. Quando o recurso HA está ativado no cluster, cada servidor ESXi mantém a comunicação com outros hosts para que, se qualquer host ESXi ficar sem resposta ou isolado, o cluster HA possa negociar a recuperação das máquinas virtuais que estavam sendo executadas nesse host ESXi entre os hosts sobreviventes no cluster. No caso de uma falha do sistema operacional convidado, o vSphere HA reinicia a máquina virtual afetada no mesmo servidor físico. O vSphere HA possibilita reduzir o tempo de inatividade planejado, evitar tempo de inatividade não planejado e recuperar rapidamente de interrupções.

Cluster do vSphere HA recuperando VMs do servidor com falha.



É importante entender que o VMware vSphere não tem conhecimento da sincronização ativa do NetApp MetroCluster ou do SnapMirror e vê todos os hosts ESXi no cluster vSphere como hosts qualificados para operações de cluster de HA, dependendo das configurações de afinidade de host e grupo de VM.

Detecção de falha do host

Assim que o cluster de HA é criado, todos os hosts do cluster participam das eleições, e um dos hosts se torna um mestre. Cada escravo executa o batimento cardíaco da rede para o mestre, e o mestre, por sua vez, executa o batimento cardíaco da rede em todos os hosts escravos. O host mestre de um cluster do vSphere HA é responsável por detectar a falha de hosts escravos.

Dependendo do tipo de falha detectada, as máquinas virtuais que estão sendo executadas nos hosts podem precisar ser reexecutadas.

Em um cluster do vSphere HA, três tipos de falha de host são detectados:

- Falha - Um host pára de funcionar.
- Isolamento - Um host se torna isolado na rede.
- Partição - Um host perde a conectividade de rede com o host mestre.

O host principal monitora os hosts escravos no cluster. Esta comunicação é feita através da troca de batimentos cardíacos de rede a cada segundo. Quando o host mestre deixa de receber esses batimentos cardíacos de um host escravo, ele verifica a presença de host antes de declarar que o host falhou. A verificação de vivacidade que o host mestre executa é determinar se o host escravo está trocando heartbeats com um dos datastores. Além disso, o host principal verifica se o host responde aos pings ICMP enviados para seus endereços IP de gerenciamento para detectar se ele é meramente isolado de seu nó mestre ou completamente isolado da rede. Ele faz isso fazendo ping no gateway padrão. Um ou mais endereços de isolamento podem ser especificados manualmente para melhorar a confiabilidade da validação de isolamento.

Melhor prática

A NetApp recomenda especificar um mínimo de dois endereços de isolamento adicionais e que cada um desses endereços seja local. Isso aumentará a confiabilidade da validação de isolamento.

Resposta de isolamento do host

A resposta de isolamento é uma configuração no vSphere HA que determina a ação acionada em máquinas virtuais quando um host em um cluster do vSphere HA perde suas conexões de rede de gerenciamento, mas continua sendo executado. Há três opções para essa configuração, "Desativado", "Desligar e reiniciar VMs" e "Desligar e reiniciar VMs".

"Desligar" é melhor do que "Desligar", que não libera as alterações mais recentes em transações de disco ou commit. Se as máquinas virtuais não desligarem em 300 segundos, elas serão desligadas. Para alterar o tempo de espera, use a opção avançada `das.isolationshutdowntimeout`.

Antes que o HA inicie a resposta de isolamento, ele primeiro verifica se o agente mestre do vSphere HA possui o datastore que contém os arquivos de configuração da VM. Caso contrário, o host não acionará a resposta de isolamento, porque não há mestre para reiniciar as VMs. O host verificará periodicamente o estado do datastore para determinar se ele é reivindicado por um agente do vSphere HA que detém a função mestre.

Melhor prática

A NetApp recomenda definir a "resposta de isolamento do host" como Desativado.

Uma condição de split-brain pode ocorrer se um host ficar isolado ou particionado do host master do vSphere HA e o mestre não conseguir se comunicar por meio de datastores de heartbeat ou por ping. O mestre declara o host isolado morto e reinicia as VMs em outros hosts no cluster. Uma condição de split-brain agora existe porque existem duas instâncias da máquina virtual em execução, apenas uma das quais pode ler ou gravar os discos virtuais. As condições de split-brain agora podem ser evitadas configurando a VM Component Protection (VMCP).

Proteção de componentes VM (VMCP)

Um dos aprimoramentos de recursos do vSphere 6, relevantes para o HA, é o VMCP. O VMCP fornece proteção aprimorada contra todas as condições de APD (caminhos para baixo) e PDL (perda permanente de dispositivo) para bloco (FC, iSCSI, FCoE) e armazenamento de arquivos (NFS).

Perda permanente de dispositivo (PDL)

PDL é uma condição que ocorre quando um dispositivo de armazenamento falha permanentemente ou é removido administrativamente e não se espera que retorne. A matriz de armazenamento NetApp emite um código SCSI Sense para ESXi declarando que o dispositivo é perdido permanentemente. Na seção condições de falha e resposta da VM do vSphere HA, você pode configurar o que a resposta deve ser depois que uma condição de PDL é detetada.

Melhor prática

A NetApp recomenda definir a "resposta para datastore com PDL" para "**Desligar e reiniciar VMs**". Quando essa condição é detetada, uma VM será reinicializada instantaneamente em um host saudável dentro do cluster do vSphere HA.

Todos os caminhos para baixo (APD)

APD é uma condição que ocorre quando um dispositivo de armazenamento se torna inacessível para o host e nenhum caminho para o array está disponível. O ESXi considera isso um problema temporário com o dispositivo e espera que ele fique disponível novamente.

Quando uma condição APD é detetada, um temporizador é iniciado. Após 140 segundos, a condição APD é

oficialmente declarada e o dispositivo é marcado como APD time out. Quando os 140 segundos tiverem passado, o HA começará a contar o número de minutos especificado no atraso para o APD de failover da VM. Quando o tempo especificado tiver passado, o HA reiniciará as máquinas virtuais afetadas. Você pode configurar o VMCP para responder de forma diferente, se desejado (Desativado, Eventos de problemas ou Desligar e reiniciar VMs).

Melhor prática

O NetApp recomenda configurar a "resposta para datastore com APD" para "**Desligar e reiniciar VMs (conservative)**".

Conservador refere-se à probabilidade de que o HA seja capaz de reiniciar VMs. Quando definido como Conservador, o HA só reiniciará a VM afetada pelo APD se souber que outro host pode reiniciá-la. No caso de agressivo, o HA tentará reiniciar a VM, mesmo que não saiba o estado dos outros hosts. Isso pode fazer com que as VMs não sejam reiniciadas se não houver nenhum host com acesso ao datastore em que ele está localizado.

Se o status APD for resolvido e o acesso ao storage for restaurado antes que o tempo limite tenha passado, o HA não reiniciará desnecessariamente a máquina virtual, a menos que você a configure explicitamente para fazê-lo. Se uma resposta for desejada, mesmo quando o ambiente foi recuperado da condição APD, então Response for APD Recovery After APD Timeout deve ser configurado para Reset VMs.

Melhor prática

O NetApp recomenda configurar a resposta para recuperação do APD após o tempo limite do APD para Desativado.

Implementação do VMware DRS para NetApp MetroCluster

O VMware DRS é um recurso que agrega os recursos de host em um cluster e é usado principalmente para o balanceamento de carga em um cluster em uma infraestrutura virtual. O VMware DRS calcula principalmente os recursos de CPU e memória para realizar o balanceamento de carga em um cluster. Como o vSphere não tem conhecimento do clustering estendido, ele considera todos os hosts em ambos os locais quando o balanceamento de carga. Para evitar o tráfego entre locais, o NetApp recomenda configurar regras de afinidade do DRS para gerenciar uma separação lógica de VMs. Isso garantirá que, a menos que haja uma falha completa no local, o HA e o DRS usarão somente hosts locais.

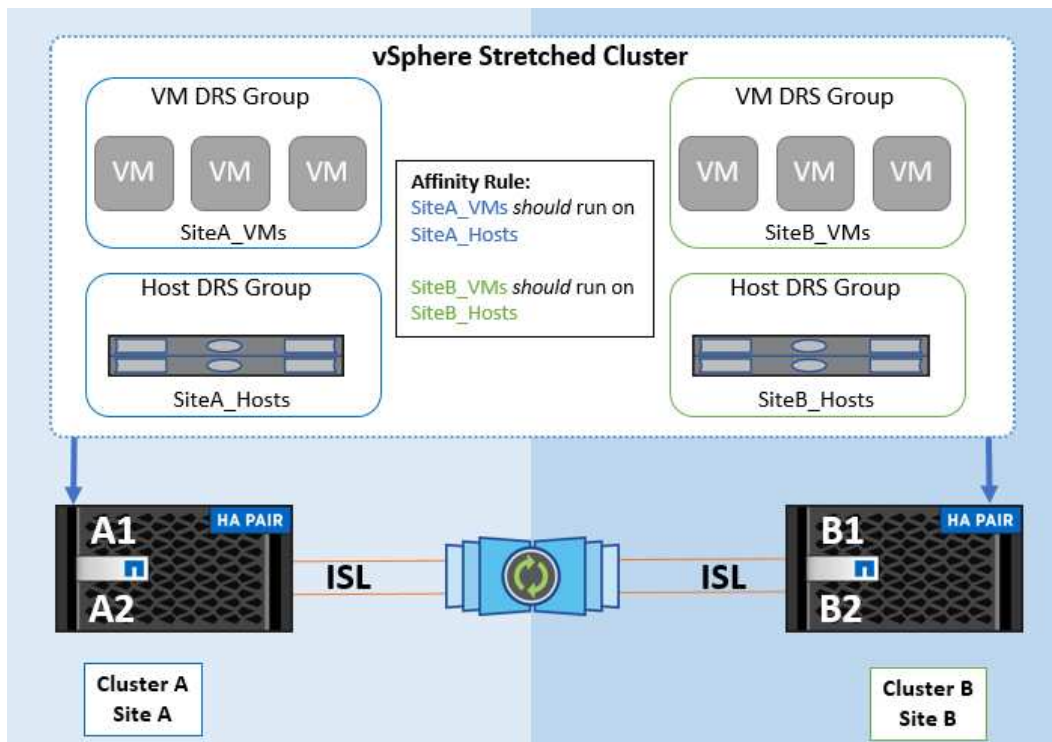
Se você criar uma regra de afinidade DRS para o cluster, poderá especificar como o vSphere aplica essa regra durante um failover de máquina virtual.

Existem dois tipos de regras que você pode especificar o comportamento de failover do vSphere HA:

- As regras de anti-afinidade da VM forçam as máquinas virtuais especificadas a permanecerem separadas durante as ações de failover.
- As regras de afinidade de host da VM colocam máquinas virtuais especificadas em um host específico ou em um membro de um grupo definido de hosts durante ações de failover.

Usando regras de afinidade de host de VM no VMware DRS, pode-se ter uma separação lógica entre o local A e o local B para que a VM seja executada no host no mesmo local do array configurado como o controlador de leitura/gravação primário para um determinado datastore. Além disso, as regras de afinidade de host da VM permitem que as máquinas virtuais permaneçam locais para o armazenamento, o que, por sua vez, verifica a conexão da máquina virtual em caso de falhas de rede entre os sites.

A seguir está um exemplo de grupos de hosts de VM e regras de afinidade.



Melhor prática

A NetApp recomenda a implementação de regras "devem" em vez de regras "obrigatórias" porque elas são violadas pelo vSphere HA em caso de falha. O uso de regras obrigatórias pode potencialmente levar a interrupções de serviço.

A disponibilidade dos serviços deve sempre prevalecer sobre o desempenho. No cenário em que um data center completo falha, as regras "obrigatórias" devem escolher hosts do grupo de afinidade de host da VM e, quando o data center não estiver disponível, as máquinas virtuais não serão reiniciadas.

Implementação do VMware Storage DRS com o NetApp MetroCluster

O recurso VMware Storage DRS permite a agregação de armazenamentos de dados em uma única unidade e equilibra discos de máquina virtual quando os limites de controle de e/S de armazenamento são excedidos.

O controle de e/S de armazenamento é habilitado por padrão nos clusters DRS habilitados para Storage DRS. O controle de e/S de armazenamento permite que um administrador controle a quantidade de e/S de armazenamento que é alocada a máquinas virtuais durante períodos de congestionamento de e/S, o que permite que máquinas virtuais mais importantes tenham preferência sobre máquinas virtuais menos importantes para alocação de recursos de e/S.

O Storage DRS usa o Storage vMotion para migrar as máquinas virtuais para diferentes datastores dentro de um cluster de datastore. Em um ambiente NetApp MetroCluster, a migração de uma máquina virtual precisa ser controlada nos datastores desse site. Por exemplo, a máquina virtual A, em execução em um host no local A, deve idealmente migrar dentro dos armazenamentos de dados do SVM no local A. se não o fizer, a máquina virtual continuará operando, mas com desempenho degradado, uma vez que a leitura/gravação do disco virtual será do local B através de links entre sites.

Melhor prática

A NetApp recomenda a criação de clusters de armazenamento de dados com relação à afinidade do local de storage; ou seja, armazenamentos de dados com afinidade de local para o local A não devem ser misturados

com clusters de armazenamento de dados com armazenamentos de dados com afinidade de local para o local B.

Sempre que uma máquina virtual é recém-provisionada ou migrada usando o Storage vMotion, a NetApp recomenda que todas as regras do VMware DRS específicas dessas máquinas virtuais sejam atualizadas manualmente, de acordo com isso. Isso verificará a afinidade da máquina virtual no nível do local para o host e o datastore e, assim, reduzirá a sobrecarga de rede e armazenamento.

Diretrizes de projeto e implementação do vMSC

Este documento descreve as diretrizes de design e implementação do vMSC com sistemas de storage ONTAP.

Configuração de armazenamento NetApp

As instruções de configuração para NetApp MetroCluster (referidas como uma configuração MCC) estão disponíveis em ["Documentação do MetroCluster"](#). As instruções para a sincronização ativa do SnapMirror também estão disponíveis em ["Visão geral da continuidade dos negócios da SnapMirror"](#).

Depois de configurar o MetroCluster, administrá-lo é como gerenciar um ambiente ONTAP tradicional. É possível configurar máquinas virtuais de storage (SVMs) usando várias ferramentas, como a interface de linha de comando (CLI), o System Manager ou o Ansible. Uma vez configurados os SVMs, crie interfaces lógicas (LIFs), volumes e números de unidades lógicas (LUNs) no cluster que será usado para operações normais. Esses objetos serão replicados automaticamente para o outro cluster usando a rede de peering de cluster.

Se não estiver usando o MetroCluster, você poderá usar o SnapMirror active Sync, que fornece proteção granular do armazenamento de dados e acesso ativo-ativo em vários clusters ONTAP em diferentes domínios de falha. O SnapMirror active Sync usa grupos de consistência para garantir a consistência da ordem de gravação entre um ou mais datastores e você pode criar vários grupos de consistência, dependendo dos requisitos do aplicativo e do armazenamento de dados. Os grupos de consistência são especialmente úteis para aplicativos que exigem sincronização de dados entre vários datastores. A sincronização ativa do SnapMirror também suporta RDMs (Mapeamentos de dispositivo brutos) e armazenamento conectado ao convidado com iniciadores iSCSI in-Guest. Você pode aprender mais sobre grupos de consistência em ["Visão geral dos grupos de consistência"](#).

Há alguma diferença no gerenciamento de uma configuração vmsc com a sincronização ativa do SnapMirror quando comparada a um MetroCluster. Primeiro, essa é uma configuração somente SAN, nenhum datastores NFS pode ser protegido com a sincronização ativa do SnapMirror. Em segundo lugar, você deve mapear ambas as cópias dos LUNs para os hosts ESXi para que eles acessem os datastores replicados em ambos os domínios de falha.

VMware vSphere HA

Crie um cluster do vSphere HA

A criação de um cluster do vSphere HA é um processo de várias etapas totalmente documentado em ["Como criar e configurar clusters no vSphere Client em docs.vmware.com"](#). Em suma, primeiro você deve criar um cluster vazio e, usando o vCenter, você deve adicionar hosts e especificar o vSphere HA e outras configurações do cluster.



Nada neste documento substitui ["Práticas recomendadas do VMware vSphere Metro Storage Cluster"](#)

Para configurar um cluster de HA, execute as seguintes etapas:

1. Conecte-se à IU do vCenter.
2. Em hosts e clusters, navegue até o data center onde você deseja criar seu cluster de HA.
3. Clique com o botão direito do rato no objeto do data center e selecione novo cluster. Em noções básicas, certifique-se de que você ativou o vSphere DRS e o vSphere HA. Conclua o assistente.

New Cluster

- 1 Basics
- 2 Image
- 3 Review

Basics

Name	MCC Cluster
Location	Raleigh
vSphere DRS	<input checked="" type="checkbox"/>
vSphere HA	<input checked="" type="checkbox"/>
vSAN	<input type="checkbox"/> Enable vSAN ESA

Manage all hosts in the cluster with a single image

Choose how to set up the cluster's image

- Compose a new image
- Import image from an existing host in the vCenter inventory
- Import image from a new host

Manage configuration at a cluster level

1. Selecione o cluster e vá para a guia configurar. Selecione vSphere HA e clique em Edit.
2. Em Monitoramento de host, selecione a opção Ativar monitoramento de host.

vSphere HA



Failures and responses | Admission Control | Heartbeat Datastores | Advanced Options

You can configure how vSphere HA responds to the failure conditions on this cluster. The following failure conditions are supported: host, host isolation, VM component protection (datastore with PDL and APD), VM and application.

Enable Host Monitoring

> Host Failure Response	Restart VMs ▾
> Response for Host Isolation	Disabled ▾
> Datastore with PDL	Power off and restart VMs ▾
> Datastore with APD	Power off and restart VMs - Conservative restart policy ▾
> VM Monitoring	Disabled ▾

CANCEL

OK

1. Enquanto ainda estiver na guia falhas e respostas, em Monitoramento de VM, selecione a opção somente Monitoramento de VM ou a opção Monitoramento de VM e aplicativo.

> Response for Host Isolation Disabled

> Datastore with PDL Power off and restart VMs

> Datastore with APD Power off and restart VMs - Conservative restart policy

▼ VM Monitoring

Enable heartbeat monitoring

VM monitoring resets individual VMs if their VMware tools heartbeats are not received within a set time. Application monitoring resets individual VMs if their in-guest heartbeats are not received within a set time.

Disabled

VM Monitoring Only

Turns on VMware tools heartbeats. When heartbeats are not received within a set time, the VM is reset.

VM and Application Monitoring

Turns on application heartbeats. When heartbeats are not received within a set time, the VM is reset.

1. Em Controle de admissão, defina a opção de controle de admissão HA para reserva de recursos de cluster; use 50% CPU/MEM.

vSphere HA

Failures and responses | Admission Control | Heartbeat Datastores | Advanced Options

Admission control is a policy used by vSphere HA to ensure failover capacity within a cluster. Raising the number of potential host failures will increase the availability constraints and capacity reserved.

Host failures cluster tolerates
 Maximum is one less than number of hosts in cluster.

Define host failover capacity by

Override calculated failover capacity.

Reserved failover CPU capacity: % CPU

Reserved failover Memory capacity: % Memory

Reserve Persistent Memory failover capacity ⓘ

Override calculated Persistent Memory failover capacity

CANCEL OK

1. Clique em "OK".
2. Selecione DRS e clique EM editar.
3. Defina o nível de automação para manual, a menos que seja necessário pelas suas aplicações.

vSphere DRS

Automation | Additional Options | Power Management | Advanced Options

Automation Level
 DRS generates both power-on placement recommendations, and migration recommendations for virtual machines. Recommendations need to be manually applied or ignored.

Migration Threshold ⓘ
 Conservative (Less Frequent vMotions) Aggressive (More Frequent vMotions)
 ⓘ DRS provides recommendations when workloads are moderately imbalanced. This threshold is suggested for environments with stable workloads. (Default)

Predictive DRS ⓘ Enable

Virtual Machine Automation ⓘ Enable

1. Ativar a proteção de componentes VM, "docs.vmware.com" consulte a .
2. As seguintes configurações adicionais do vSphere HA são recomendadas para vMSC com MCC:

Falha	Resposta
Falha do host	Reinicie as VMs
Isolamento de host	Desativado
Armazenamento de dados com perda permanente de dispositivo (PDL)	Desligue e reinicie as VMs
Datastore com todos os caminhos para baixo (APD)	Desligue e reinicie as VMs
Hóspede não é coração batendo	Repor as VMs
Política de reinicialização da VM	Determinado pela importância da VM
Resposta para isolamento do host	Encerre e reinicie as VMs
Resposta para datastore com PDL	Desligue e reinicie as VMs
Resposta para datastore com APD	Desligar e reiniciar as VMs (conservadoras)
Atraso para failover de VM para APD	3 minutos
Resposta para recuperação APD com tempo limite APD	Desativado
Sensibilidade de monitoramento da VM	Predefinição alta

Configurar datastores para Heartbearing

O vSphere HA usa datastores para monitorar hosts e máquinas virtuais quando a rede de gerenciamento falhou. Você pode configurar como o vCenter seleciona armazenamentos de dados Heartbeat. Para configurar armazenamentos de dados para batimentos cardíacos, execute as seguintes etapas:

1. Na seção Heartbearing do datastore, selecione usar datastores na Lista especificada e elogiar automaticamente, se necessário.
2. Selecione os datastores que você deseja que o vCenter use em ambos os sites e pressione OK.

vSphere HA

Failures and responses Admission Control **Heartbeat Datastores** Advanced Options

vSphere HA uses datastores to monitor hosts and virtual machines when the HA network has failed. vCenter Server selects 4 datastores for each host using the policy and datastore preferences specified below.

Heartbeat datastore selection policy:

- Automatically select datastores accessible from the hosts
- Use datastores only from the specified list
- Use datastores from the specified list and complement automatically if needed

Available heartbeat datastores

	Name ↑	Datastore Cluster	Hosts Mounting Datastore
<input checked="" type="checkbox"/>	d11	N/A	2
<input checked="" type="checkbox"/>	d12	N/A	2
<input checked="" type="checkbox"/>	d21	N/A	2
<input checked="" type="checkbox"/>	d22	N/A	2
<input type="checkbox"/>	d31	N/A	2
<input type="checkbox"/>	d32	N/A	2
<input type="checkbox"/>	d41	N/A	2
<input type="checkbox"/>	d42	N/A	2

11 items

Configurar opções avançadas

Detecção de falha do host

Os eventos de isolamento ocorrem quando os hosts dentro de um cluster de HA perdem a conectividade com a rede ou com outros hosts no cluster. Por padrão, o vSphere HA usará o gateway padrão para sua rede de gerenciamento como endereço de isolamento padrão. No entanto, você pode especificar endereços de isolamento adicionais para o host fazer ping para determinar se uma resposta de isolamento deve ser acionada. Adicione dois IPs de isolamento que podem fazer ping, um por local. Não utilize o IP do gateway. A configuração avançada do vSphere HA usada é `das.isolationaddress`. Você pode usar endereços IP do ONTAP ou Mediator para esse fim.

```
https://core.vmware.com/resource/vmware-vsphere-metro-storage-cluster-recommended-practices#sec2-sub5["core.vmware.com"]Consulte para obter mais informações__._
```


vSphere HA

Failures and responses Admission Control Heartbeat Datastores **Advanced Options**

You can set advanced options that affect the behavior of your vSphere HA cluster.

+ Add ✕ Delete

Option	Value
das.IgnoreRedundantNetWarning	true
das.Isolationaddress0	10.61.99.100
das.Isolationaddress1	10.61.99.110
das.heartbeatDsPerHost	4

4 items

CANCEL OK

Adicionar uma configuração avançada chamada `das.heartbeatDsPerHost` pode aumentar o número de datastores de heartbeat. Use quatro datastores de heartbeat (HB DSS) - dois por local. Use a opção "Selecionar a partir da lista, mas elogio". Isso é necessário porque, se um local falhar, você ainda precisará de dois DSS HB. No entanto, eles não precisam ser protegidos com MCC ou sincronização ativa do SnapMirror.

```
https://core.vmware.com/resource/vmware-vsphere-metro-storage-cluster-recommended-practices#sec2-sub5["core.vmware.com"]Consulte para obter mais informações____.____
```

Afinidade do VMware DRS para NetApp MetroCluster

Nesta seção, criamos grupos DRS para VMs e hosts para cada site/cluster no ambiente MetroCluster. Em seguida, configuramos regras VM/Host para alinhar a afinidade do host da VM com os recursos de armazenamento local. Por exemplo, as VMs do Site A pertencem ao grupo VM `sitea_vms` e os hosts do Site A pertencem ao grupo de hosts `sitea_hosts`. Em seguida, nas regras VM/Host, declaramos que o `sitea_vms` deve ser executado em hosts no `sitea_hosts`.

- O NetApp recomenda altamente a especificação **deve ser executada em hosts no Grupo** em vez da especificação **deve ser executada em hosts no Grupo**. No caso de uma falha de host de um local, as VMs do local A precisam ser reiniciadas em hosts no local B por meio do vSphere HA, mas a última especificação não permite que o HA reinicie VMs no local B porque é uma regra geral. A especificação anterior é uma regra suave e será violada em caso de HA, permitindo assim disponibilidade em vez de desempenho.

***[NOTA]**

- Você pode criar um alarme baseado em eventos que é acionado quando uma máquina virtual viola uma regra de afinidade VM-Host. No vSphere Client, adicione um novo alarme para a máquina virtual e selecione "VM está violando a regra de afinidade VM-Host" como o gatilho do evento. Para obter mais informações sobre como criar e editar alarmes, ["Monitoramento e desempenho do vSphere"](#) consulte a documentação.

Crie grupos de hosts DRS

Para criar grupos de hosts DRS específicos ao local A e local B, execute as seguintes etapas:

1. No cliente da Web vSphere, clique com o botão direito do Mouse no cluster no inventário e selecione Configurações.
2. Clique em VM/Host Groups.
3. Clique em Adicionar.
4. Digite o nome do grupo (por exemplo, sitea_hosts).
5. No menu tipo, selecione Grupo anfitrião.
6. Clique em Adicionar e selecione os hosts desejados no site A e clique em OK.
7. Repita estas etapas para adicionar outro grupo de hosts para o local B.
8. Clique em OK.

Crie grupos de VM DRS

Para criar grupos de VM DRS específicos para o local A e o local B, execute as seguintes etapas:

1. No cliente da Web vSphere, clique com o botão direito do Mouse no cluster no inventário e selecione Configurações.
2. Clique em VM/Host Groups.
3. Clique em Adicionar.
4. Digite o nome do grupo (por exemplo, sitea_vms).
5. No menu tipo, selecione Grupo VM.
6. Clique em Adicionar e selecione as VMs desejadas no local A e clique em OK.
7. Repita estas etapas para adicionar outro grupo de hosts para o local B.
8. Clique em OK.

Criar regras de host de VM

Para criar regras de afinidade do DRS específicas ao local A e ao local B, execute as seguintes etapas:

1. No cliente da Web vSphere, clique com o botão direito do Mouse no cluster no inventário e selecione Configurações.
2. Clique em VM/Host Rules.
3. Clique em Adicionar.
4. Digite o nome da regra (por exemplo, sitea_Affinity).
5. Verifique se a opção Ativar regra está marcada.
6. No menu tipo, selecione máquinas virtuais para hosts.
7. Selecione o grupo VM (por exemplo, sitea_vms).
8. Selecione o grupo Host (por exemplo, sitea_hosts).
9. Repita estas etapas para adicionar outra VM/regra de host para o local B.
10. Clique em OK.

Create VM/Host Rule | Cluster-01 ×

Name	sitea_affinity <input checked="" type="checkbox"/> Enable rule.
Type	Virtual Machines to Hosts ▼

Virtual machines that are members of the Cluster VM Group sitea_vms should run on host group sitea_hosts.

VM Group:

sitea_vms ▼
Should run on hosts in group ▼

Host Group:

sitea_hosts ▼

CANCEL	OK
--------	----

VMware vSphere Storage DRS for NetApp MetroCluster

Crie clusters de armazenamento de dados

Para configurar um cluster de datastore para cada site, execute as seguintes etapas:

1. Usando o cliente da Web vSphere, navegue até o data center em que o cluster HA reside em Storage.
2. Clique com o botão direito do rato no objeto do data center e selecione armazenamento > novo cluster do datastore.
3. Selecione a opção Ativar o Storage DRS e clique em Avançar.
4. Defina todas as opções como sem automação (modo manual) e clique em Avançar.

Melhor prática

- O NetApp recomenda que o DRS de armazenamento seja configurado no modo manual, para que o administrador possa decidir e controlar quando as migrações precisam acontecer.



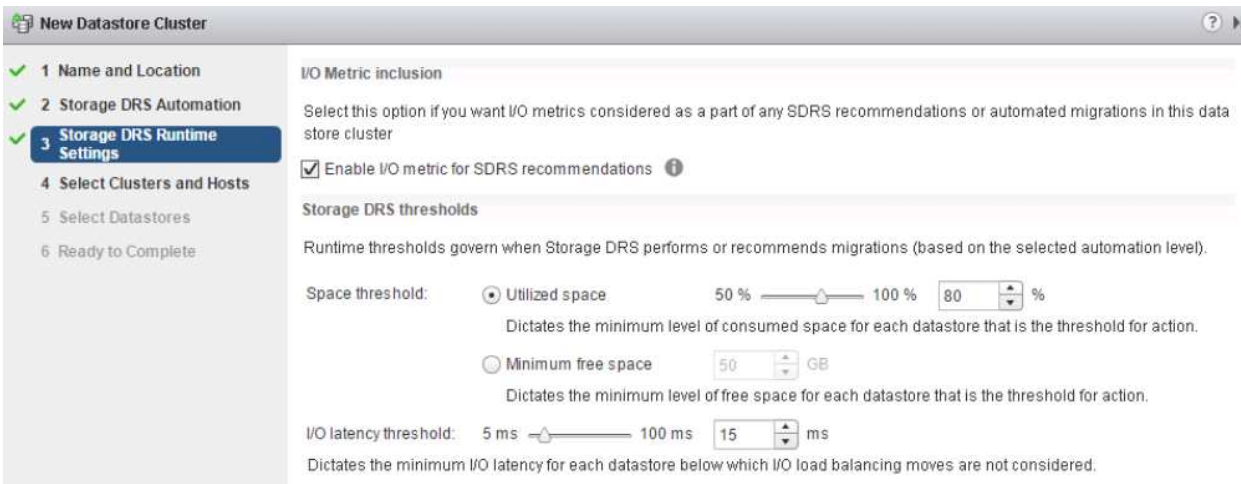
Storage DRS automation

Cluster automation level

No Automation (Manual Mode)
vCenter Server will make migration recommendations for virtual machine storage, but will not perform automatic migrations.

Fully Automated
Files will be migrated automatically to optimize resource usage.

1. Verifique se a caixa de verificação Ativar métricas de e/S para recomendações de SDRS está marcada; as configurações de métrica podem ser deixadas com os valores padrão.



New Datastore Cluster

1 Name and Location

2 Storage DRS Automation

3 **Storage DRS Runtime Settings**

4 Select Clusters and Hosts

5 Select Datastores

6 Ready to Complete

I/O Metric inclusion

Select this option if you want I/O metrics considered as a part of any SDRS recommendations or automated migrations in this datastore cluster

Enable I/O metric for SDRS recommendations

Storage DRS thresholds

Runtime thresholds govern when Storage DRS performs or recommends migrations (based on the selected automation level).

Space threshold: Utilized space 50 % — 100 % 80 %
Dictates the minimum level of consumed space for each datastore that is the threshold for action.

Minimum free space 50 GB
Dictates the minimum level of free space for each datastore that is the threshold for action.

I/O latency threshold: 5 ms — 100 ms 15 ms
Dictates the minimum I/O latency for each datastore below which I/O load balancing moves are not considered.

1. Selecione o cluster HA e clique em Next (seguinte).



New Datastore Cluster

1 Name and Location

2 Storage DRS Automation

3 Storage DRS Runtime Settings

4 **Select Clusters and Hosts**

5 Select Datastores

6 Ready to Complete

Select all hosts and clusters that require connectivity to the datastores in the datastore cluster.

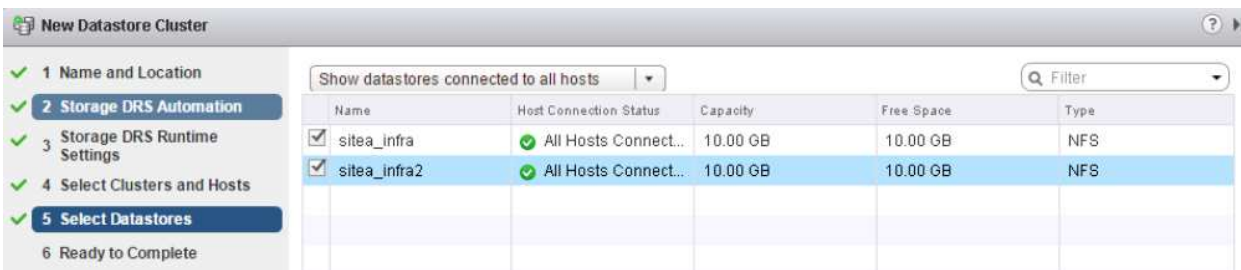
Filter (1) Selected Objects

Clusters Standalone Hosts

Filter

Name
<input checked="" type="checkbox"/> MCC HA Cluster

1. Selecione os datastores pertencentes ao site A e clique em Avançar.



New Datastore Cluster

1 Name and Location

2 Storage DRS Automation

3 Storage DRS Runtime Settings

4 Select Clusters and Hosts

5 **Select Datastores**

6 Ready to Complete

Show datastores connected to all hosts

Filter

Name	Host Connection Status	Capacity	Free Space	Type
<input checked="" type="checkbox"/> sitea_infra	All Hosts Connect...	10.00 GB	10.00 GB	NFS
<input checked="" type="checkbox"/> sitea_infra2	All Hosts Connect...	10.00 GB	10.00 GB	NFS

1. Reveja as opções e clique em concluir.
2. Repita essas etapas para criar o cluster do datastore do site B e verifique se somente os datastores do site B estão selecionados.

Disponibilidade do vCenter Server

Os dispositivos do vCenter Server (VCSAs) devem ser protegidos com o vCenter HA. O vCenter HA permite implantar dois VCSAs em um par de HA ativo-passivo. Um em cada domínio de falha. Você pode ler mais sobre o vCenter HA no "docs.vmware.com".

Resiliência para eventos planejados e não planejados

O NetApp MetroCluster e o SnapMirror ativo Sync são ferramentas poderosas que melhoram a alta disponibilidade e as operações ininterruptas do hardware NetApp e do software ONTAP.

Essas ferramentas fornecem proteção em todo o local para todo o ambiente de storage, garantindo que seus dados estejam sempre disponíveis. Se você estiver usando servidores autônomos, clusters de servidores de alta disponibilidade, contentores Docker ou servidores virtualizados, a tecnologia NetApp mantém perfeitamente a disponibilidade de armazenamento em caso de uma interrupção total devido à perda de energia, resfriamento ou conectividade de rede, desligamento de storage array ou erro operacional.

O MetroCluster e o SnapMirror ativo Sync oferecem três métodos básicos para a continuidade dos dados em caso de eventos planejados ou não planejados:

- Componentes redundantes para proteção contra falha de componente único
- Takeover de HA local para eventos que afetam um único controlador
- Proteção completa do local – retomada rápida do serviço movendo o armazenamento e o acesso do cliente do cluster de origem para o cluster de destino

Isso significa que as operações continuam sem problemas em caso de falha de um único componente e retornam automaticamente à operação redundante quando o componente com falha é substituído.

Todos os clusters do ONTAP, exceto clusters de nó único (normalmente versões definidas por software, como o ONTAP Select, por exemplo), têm recursos de HA incorporados, chamados de takeover e giveback. Cada controlador no cluster é emparelhado com outro controlador, formando um par de HA. Esses pares garantem que cada nó esteja conectado localmente ao storage.

O takeover é um processo automatizado no qual um nó assume o storage do outro para manter os serviços de dados. Giveback é o processo reverso que restaura a operação normal. O takeover pode ser planejado, como ao executar a manutenção de hardware ou atualizações de ONTAP, ou não planejado, resultante de um pânico de nó ou falha de hardware.

Durante um takeover, as interfaces lógicas de storage conectadas à rede (LIFs nas) nas configurações do MetroCluster fazem failover automático. No entanto, os LIFs de rede de Área de armazenamento (SAN LIFs) não falham; eles continuarão a usar o caminho direto para os LUNs (Logical Unit Numbers).

Para obter mais informações sobre a aquisição de HA e a giveback, consulte o "[Visão geral do gerenciamento do par HA](#)". Vale a pena notar que essa funcionalidade não é específica para a sincronização ativa do MetroCluster ou do SnapMirror.

O switchover do local com o MetroCluster ocorre quando um local está off-line ou como uma atividade planejada para manutenção em todo o local. O local restante assume a propriedade dos recursos de storage (discos e agregados) do cluster off-line, e os SVMs no site com falha são colocados on-line e reiniciados no local de desastre, preservando sua identidade completa para acesso ao cliente e ao host.

Com a sincronização ativa do SnapMirror, uma vez que ambas as cópias são usadas ativamente simultaneamente, seus hosts existentes continuarão operando. O Mediador NetApp é necessário para garantir

que o failover do site ocorra corretamente.

Cenários de falha para vMSC com MCC

As seções a seguir descrevem os resultados esperados de vários cenários de falha com sistemas vMSC e NetApp MetroCluster.

Falha de caminho de storage único

Nesse cenário, se componentes como a porta HBA, a porta de rede, a porta do switch de dados front-end ou um cabo FC ou Ethernet falharem, esse caminho específico para o dispositivo de armazenamento será marcado como morto pelo host ESXi. Se vários caminhos forem configurados para o dispositivo de storage fornecendo resiliência na porta HBA/rede/switch, o ESXi executará idealmente um switchover de caminho. Durante esse período, as máquinas virtuais permanecem em execução sem serem afetadas, pois a disponibilidade para o armazenamento é tratada fornecendo vários caminhos para o dispositivo de armazenamento.



Não há nenhuma mudança no comportamento do MetroCluster neste cenário, e todos os datastores continuam intactos de seus respectivos sites.

Melhor prática

Em ambientes em que os volumes NFS/iSCSI são usados, a NetApp recomenda ter pelo menos dois uplinks de rede configurados para a porta NFS vmkernel no vSwitch padrão e o mesmo no grupo de portas em que a interface NFS vmkernel é mapeada para o vSwitch distribuído. O agrupamento de NIC pode ser configurado em ativo-ativo ou ativo-standby.

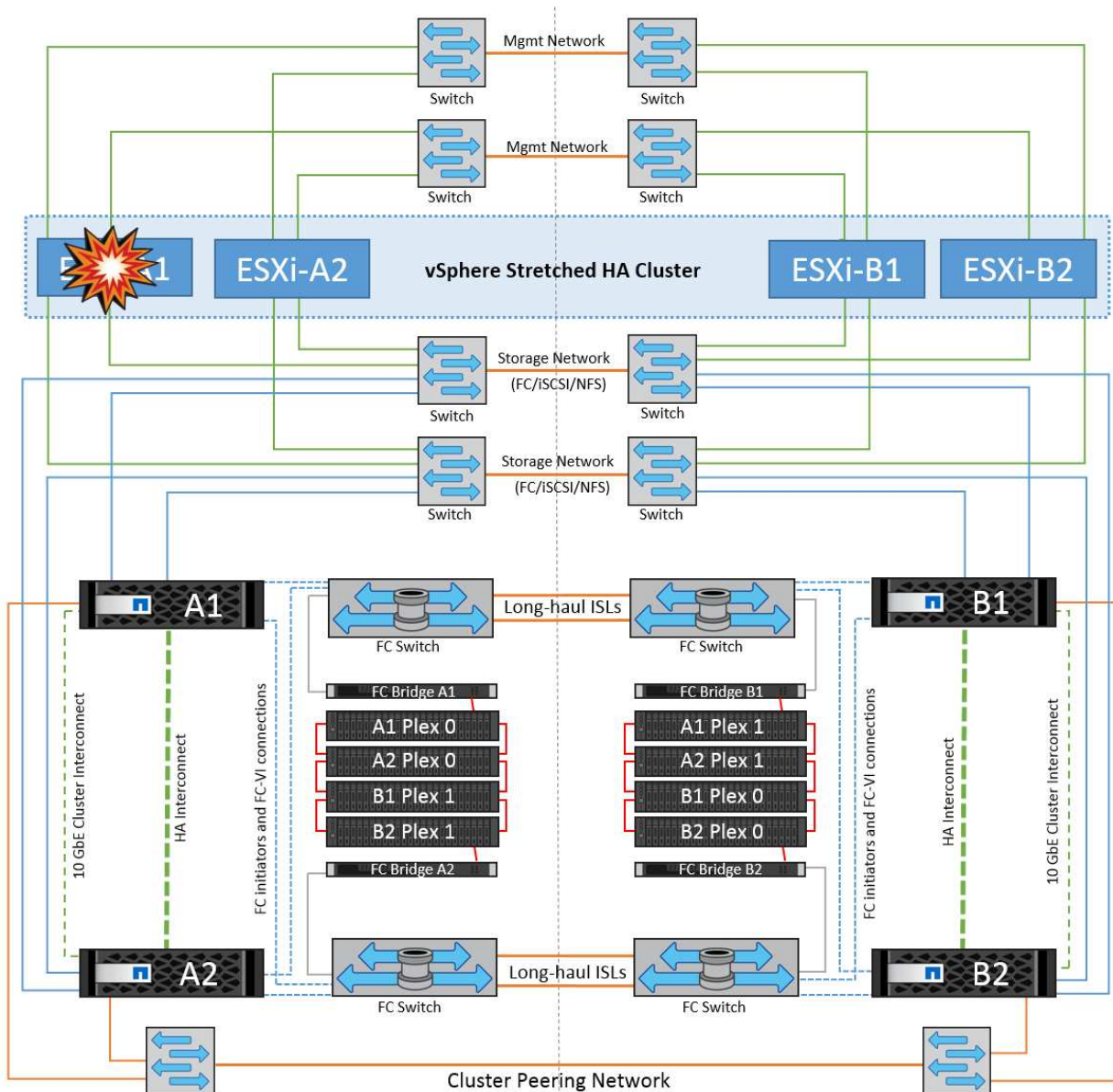
Além disso, para iSCSI LUNs, multipathing deve ser configurado vinculando as interfaces vmkernel aos adaptadores de rede iSCSI. Para obter mais informações, consulte a documentação de armazenamento do vSphere.

Melhor prática

Em ambientes em que LUNs de Fibre Channel são usados, a NetApp recomenda ter pelo menos dois HBAs, o que garante resiliência no nível de HBA/porta. O NetApp também recomenda um iniciador único para o zoneamento de destino único como a melhor prática para configurar o zoneamento.

O VSC (Virtual Storage Console) deve ser usado para definir políticas de multipathing porque define políticas para todos os dispositivos de armazenamento NetApp novos e existentes.

Falha única do host ESXi



Nesse cenário, se houver uma falha do host ESXi, o nó mestre no cluster do VMware HA detecta a falha do host, já que ele não recebe mais batimentos cardíacos da rede. Para determinar se o host está realmente inativo ou apenas uma partição de rede, o nó mestre monitora os batimentos cardíacos do datastore e, se eles estiverem ausentes, ele executa uma verificação final fazendo ping nos endereços IP de gerenciamento do host com falha. Se todas essas verificações forem negativas, o nó principal declara que este host é um host com falha e todas as máquinas virtuais que estavam sendo executadas nesse host com falha são reinicializadas no host sobrevivente no cluster.

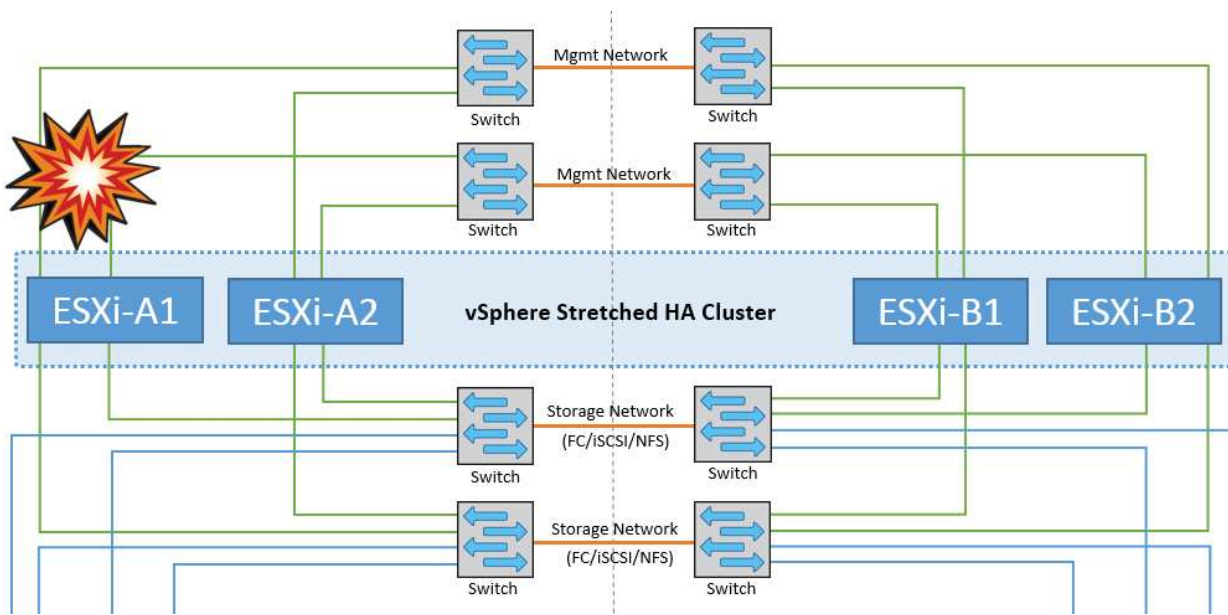
Se as regras de afinidade de host e VM DRS tiverem sido configuradas (as VMs no grupo VM sitea_vms devem executar hosts no grupo host sitea_hosts), o mestre HA primeiro verifica se há recursos disponíveis no local A. Se não houver hosts disponíveis no local A, o mestre tentará reiniciar as VMs nos hosts no local B.

É possível que as máquinas virtuais sejam iniciadas nos hosts ESXi no outro site se houver uma restrição de recursos no site local. No entanto, as regras de afinidade de host e VM DRS definidas corrigirão se alguma regra for violada migrando as máquinas virtuais de volta para qualquer host ESXi sobrevivente no site local. Nos casos em que o DRS é definido como manual, o NetApp recomenda chamar o DRS e aplicar as recomendações para corrigir o posicionamento da máquina virtual.

Não há nenhuma mudança no comportamento do MetroCluster neste cenário e todos os datastores continuam

intactos de seus respectivos sites.

Isolamento do host ESXi

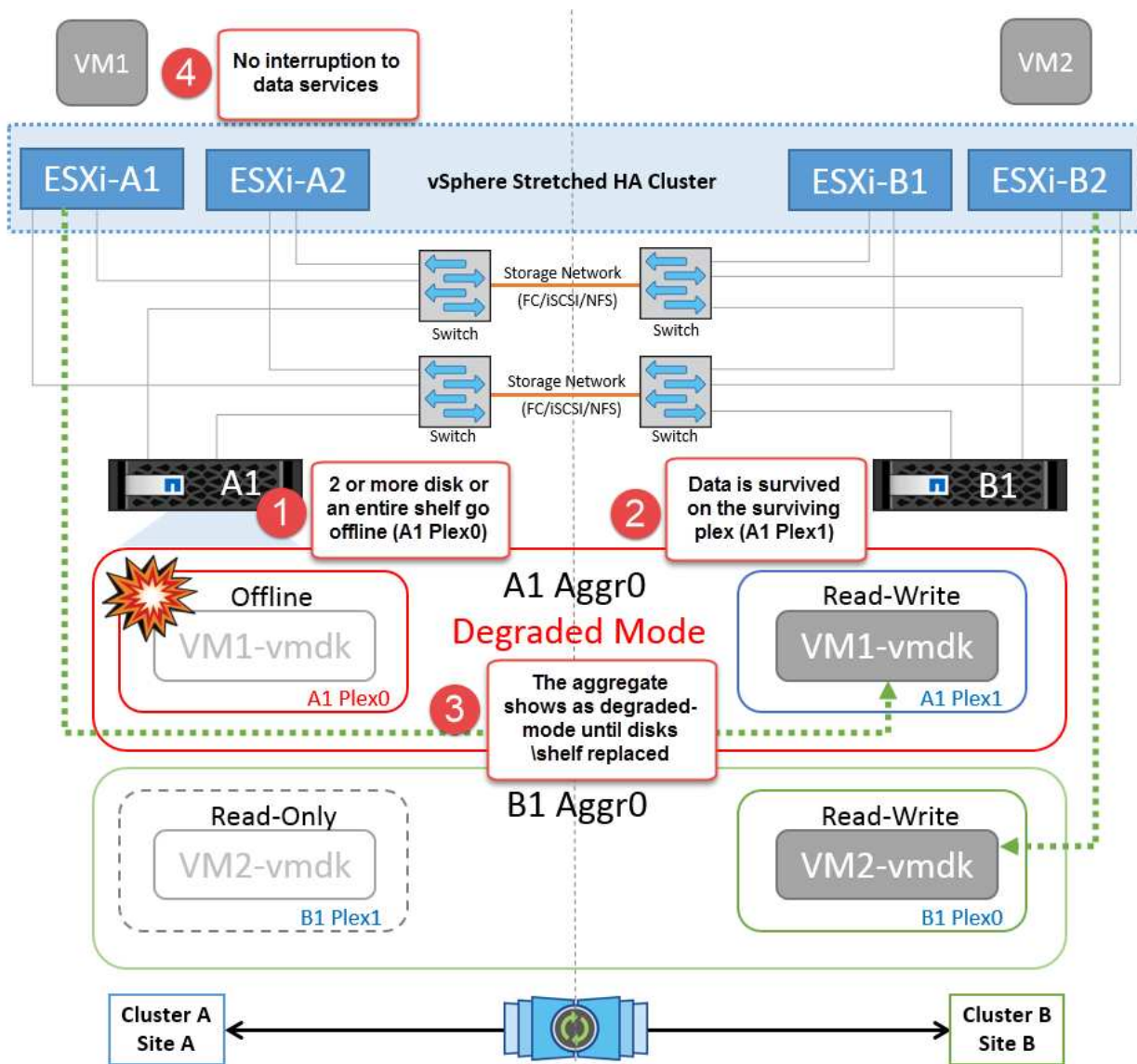


Nesse cenário, se a rede de gerenciamento do host ESXi estiver inativa, o nó mestre no cluster HA não receberá nenhum heartbeats e, portanto, esse host fica isolado na rede. Para determinar se ele falhou ou está isolado apenas, o nó principal começa a monitorar o batimento cardíaco do datastore. Se estiver presente, o host é declarado isolado pelo nó mestre. Dependendo da resposta de isolamento configurada, o host pode optar por desligar, desligar as máquinas virtuais ou até mesmo deixar as máquinas virtuais ligadas. O intervalo padrão para a resposta de isolamento é de 30 segundos.

Não há nenhuma mudança no comportamento do MetroCluster neste cenário e todos os datastores continuam intactos de seus respectivos sites.

Falha no compartimento de disco

Nesse cenário, há uma falha de mais de dois discos ou de uma gaveta inteira. Os dados são fornecidos do Plex sobrevivente sem interrupção para os serviços de dados. A falha do disco pode afetar um Plex local ou remoto. Os agregados serão apresentados como modo degradado porque apenas um Plex está ativo. Depois que os discos com falha forem substituídos, os agregados afetados serão ressincronizados automaticamente para reconstruir os dados. Após a ressincronização, os agregados retornarão automaticamente ao modo espelhado normal. Se mais de dois discos dentro de um único grupo RAID falharem, o Plex terá de ser reconstruído do zero.

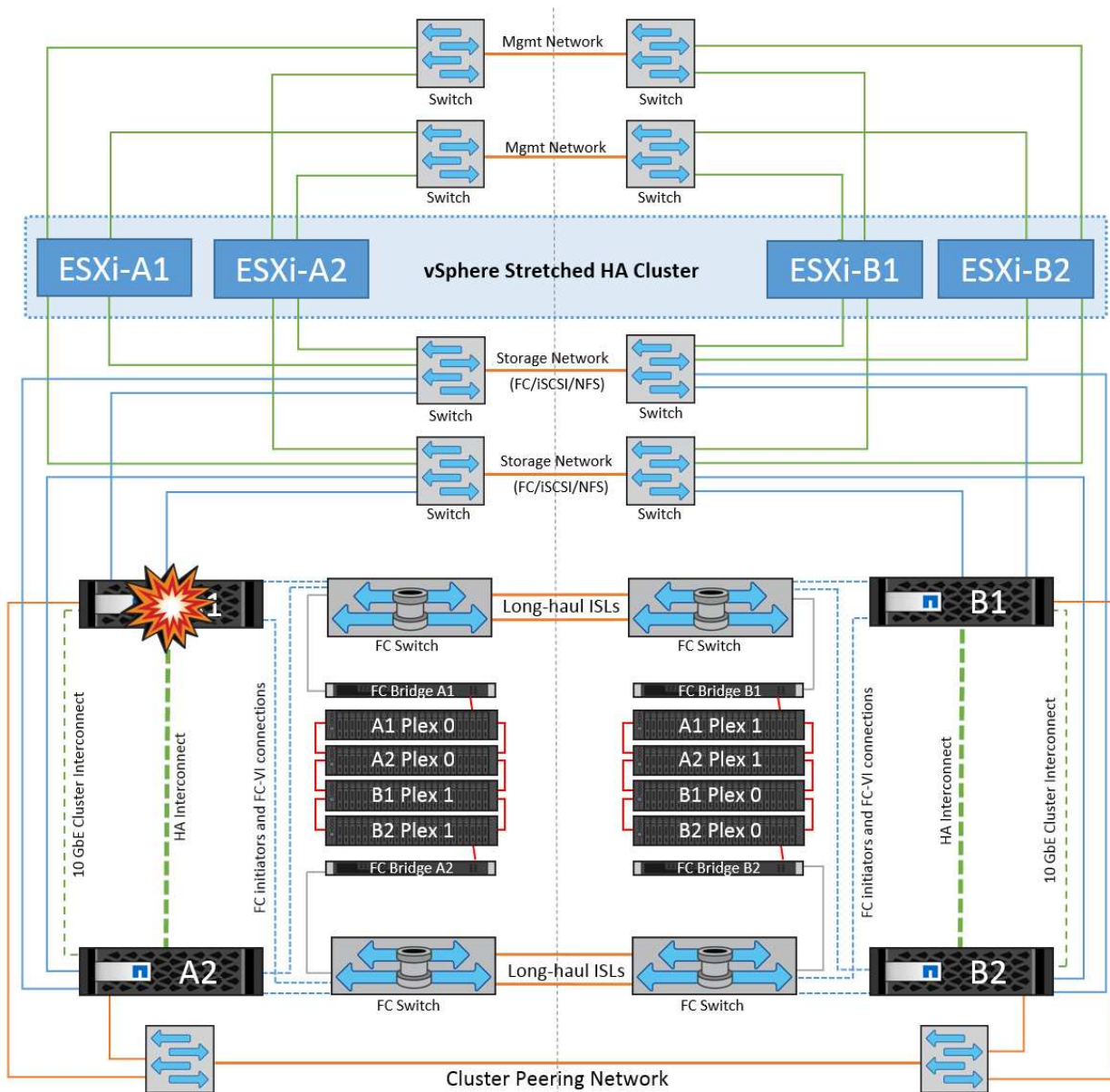


*[NOTA]

- Durante esse período, não há impactos nas operações de e/S da máquina virtual, mas há desempenho degradado porque os dados estão sendo acessados do compartimento de disco remoto por meio de links ISL.

Falha no controlador de storage único

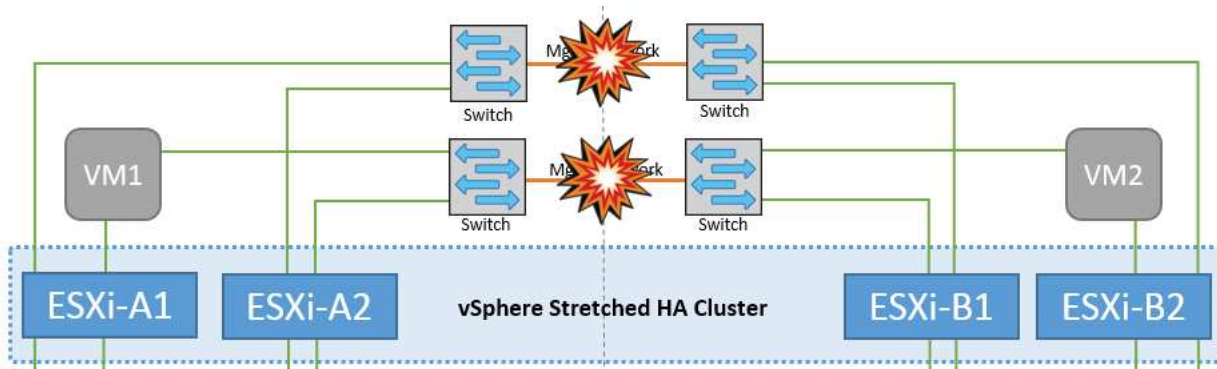
Nesse cenário, um dos dois controladores de storage falha em um local. Como há um par de HA em cada local, uma falha de um nó aciona o failover automaticamente para o outro nó. Por exemplo, se o nó A1 falhar, o storage e os workloads serão transferidos automaticamente para o nó A2. As máquinas virtuais não serão afetadas porque todos os plexos permanecem disponíveis. Os segundo nós do local (B1 e B2) não são afetados. Além disso, o vSphere HA não tomará nenhuma ação porque o nó mestre no cluster ainda estará recebendo os batimentos cardíacos da rede.



Se o failover fizer parte de um desastre contínuo (nó A1 faz failover para A2) e houver uma falha subsequente de A2 ou a falha completa do local A, o switchover após um desastre pode ocorrer no local B.

Avarias na ligação InterSwitch

Falha de ligação InterSwitch na rede de gestão

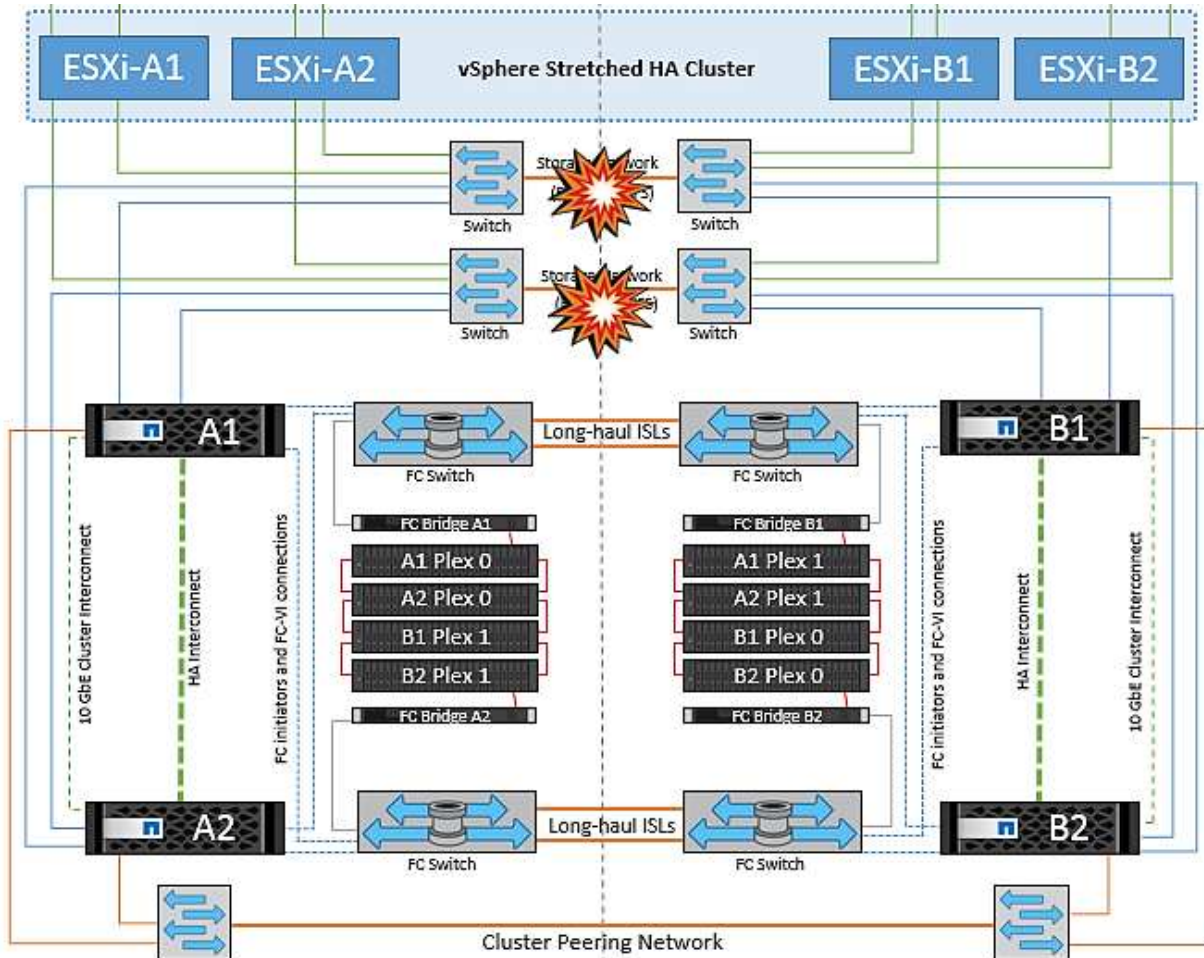


Nesse cenário, se os links ISL na rede de gerenciamento de host front-end falharem, os hosts ESXi no local A não poderão se comunicar com hosts ESXi no local B. Isso levará a uma partição de rede porque os hosts ESXi em um determinado local não poderão enviar os batimentos cardíacos da rede para o nó mestre no cluster HA. Como tal, haverá dois segmentos de rede por causa da partição e haverá um nó mestre em cada segmento que protegerá as VMs de falhas de host dentro do site específico.



Durante esse período, as máquinas virtuais permanecem em execução e não há alteração no comportamento do MetroCluster nesse cenário. Todos os armazenamentos de dados continuam intactos de seus respectivos sites.

Falha na ligação InterSwitch na rede de armazenamento

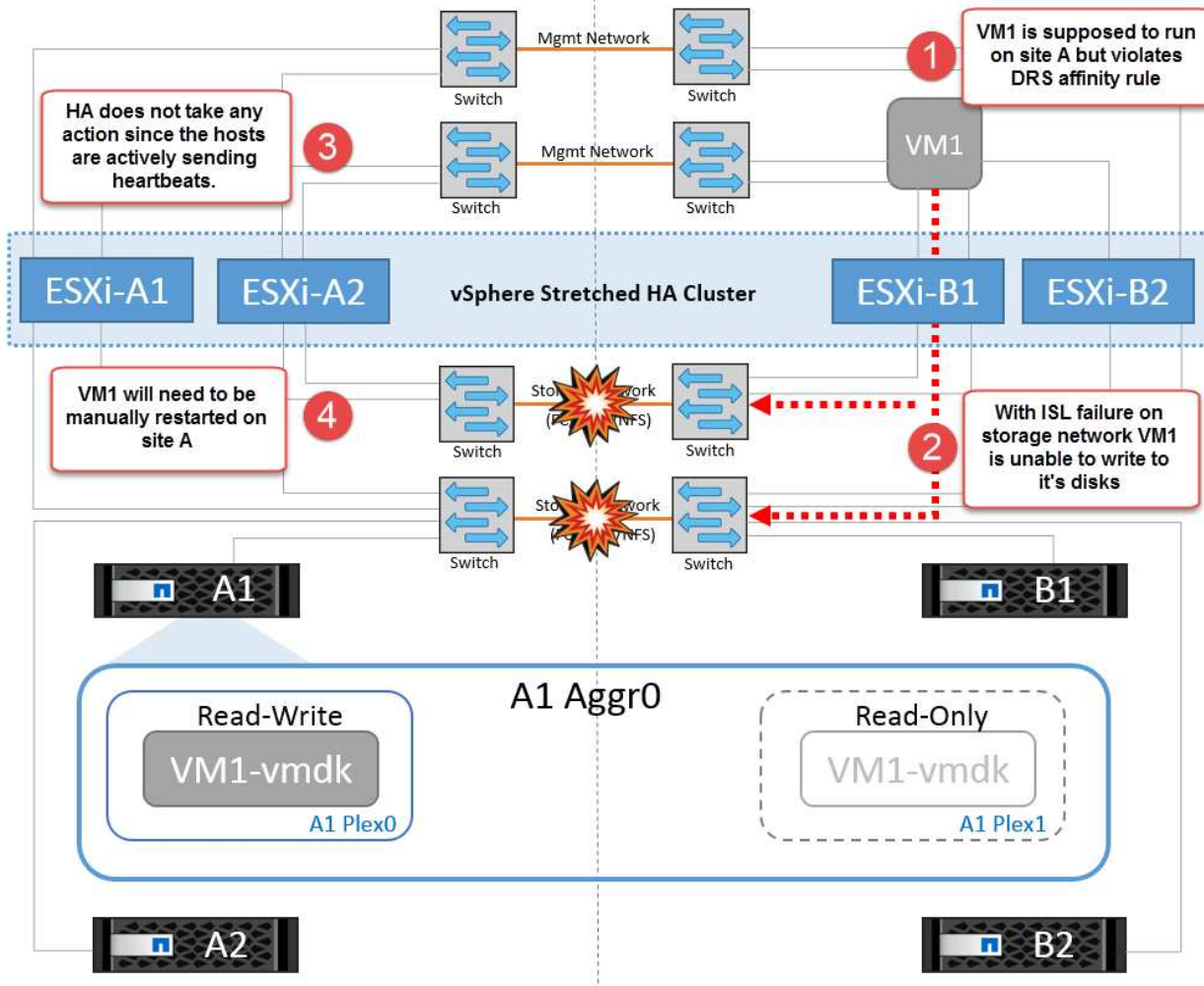


Nesse cenário, se os links ISL na rede de armazenamento de back-end falharem, os hosts no local A perderão acesso aos volumes de armazenamento ou LUNs do cluster B no local B e vice-versa. As regras do VMware DRS são definidas para que a afinidade do local de armazenamento de host facilite a execução das máquinas virtuais sem impactos no local.

Durante esse período, as máquinas virtuais permanecem em execução em seus respectivos sites e não há alteração no comportamento do MetroCluster nesse cenário. Todos os armazenamentos de dados continuam intactos de seus respectivos sites.

Se, por algum motivo, a regra de afinidade foi violada (por exemplo, VM1, que deveria ser executado a partir do site A, onde seus discos residem em nós de cluster local A, está sendo executado em um host no local B), o disco da máquina virtual será acessado remotamente por meio de links ISL. Devido à falha do link ISL, o VM1 em execução no local B não seria capaz de gravar em seus discos porque os caminhos para o volume

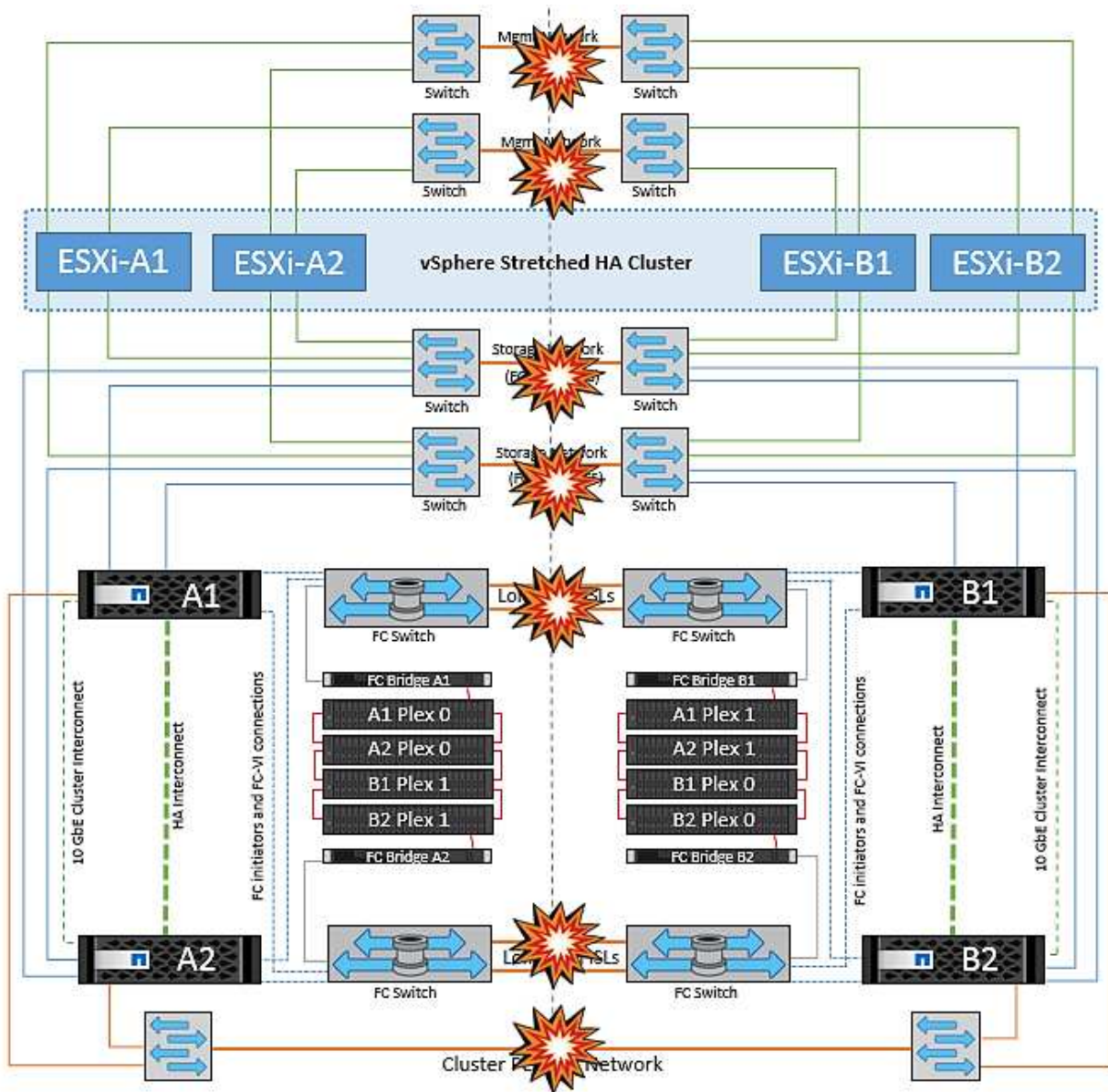
de armazenamento estão inativos e essa máquina virtual específica está inativa. Nessas situações, o VMware HA não toma nenhuma ação, uma vez que os hosts estão enviando batimentos cardíacos ativamente. Essas máquinas virtuais precisam ser manualmente desligadas e ligadas em seus respectivos sites. A figura a seguir ilustra uma VM que viola uma regra de afinidade DRS.



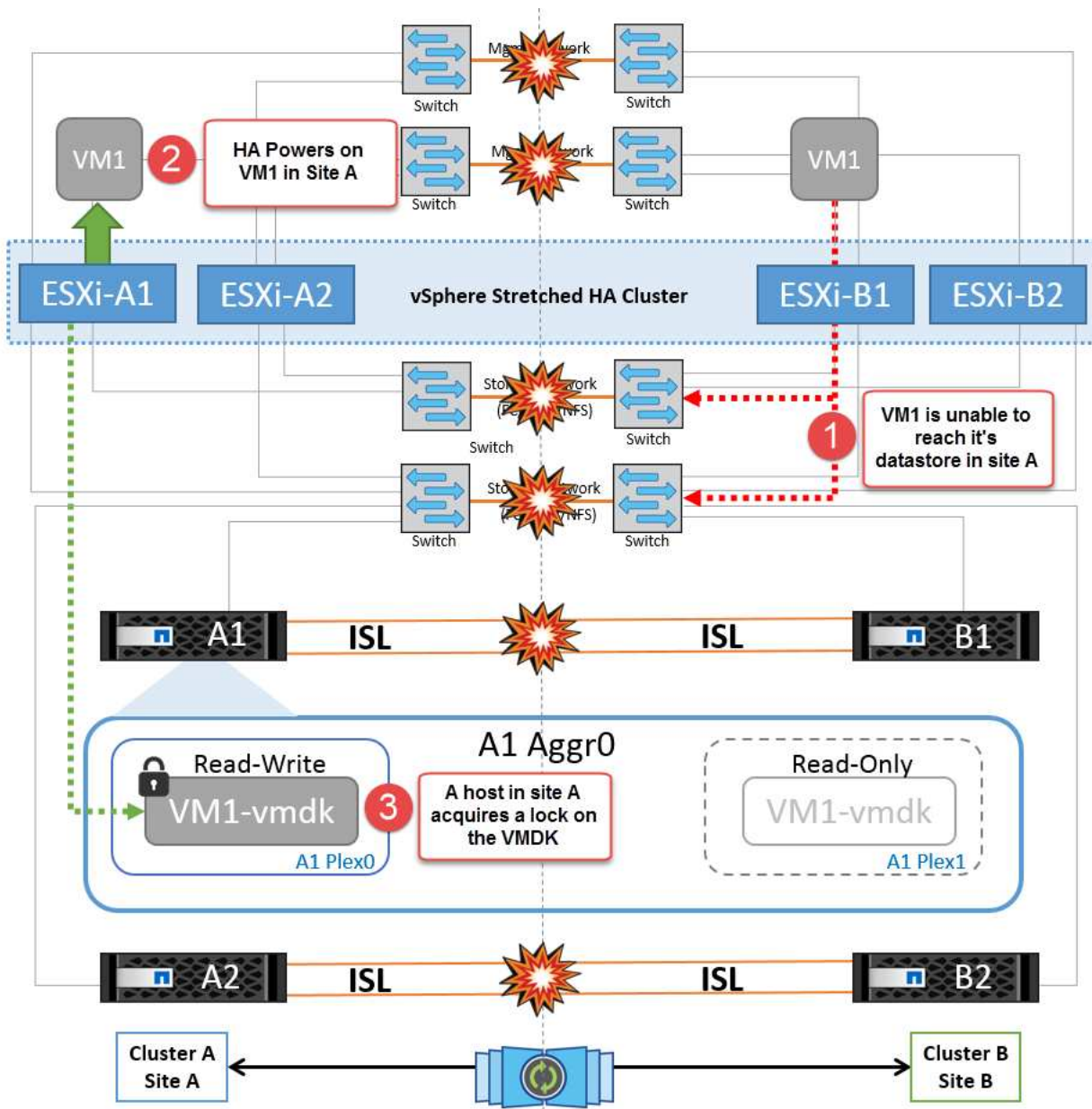
Todas as falhas do InterSwitch ou completa partição do data center

Neste cenário, todos os links ISL entre os sites estão inativos e ambos os sites são isolados uns dos outros. Como discutido em cenários anteriores, como falha de ISL na rede de gerenciamento e na rede de armazenamento, as máquinas virtuais não são afetadas em falha completa de ISL.

Depois que os hosts ESXi forem particionados entre sites, o agente do vSphere HA verificará os batimentos cardíacos do datastore e, em cada site, os hosts ESXi locais poderão atualizar os batimentos cardíacos do datastore para o respectivo volume/LUN de leitura e gravação. Os hosts no local A assumirão que os outros hosts ESXi no local B falharam porque não há heartbeats de rede/datastore. O vSphere HA no local A tentará reiniciar as máquinas virtuais do local B, o que acabará falhando porque os datastores do local B não estarão acessíveis devido a falha do ISL de armazenamento. Uma situação semelhante é repetida no local B..



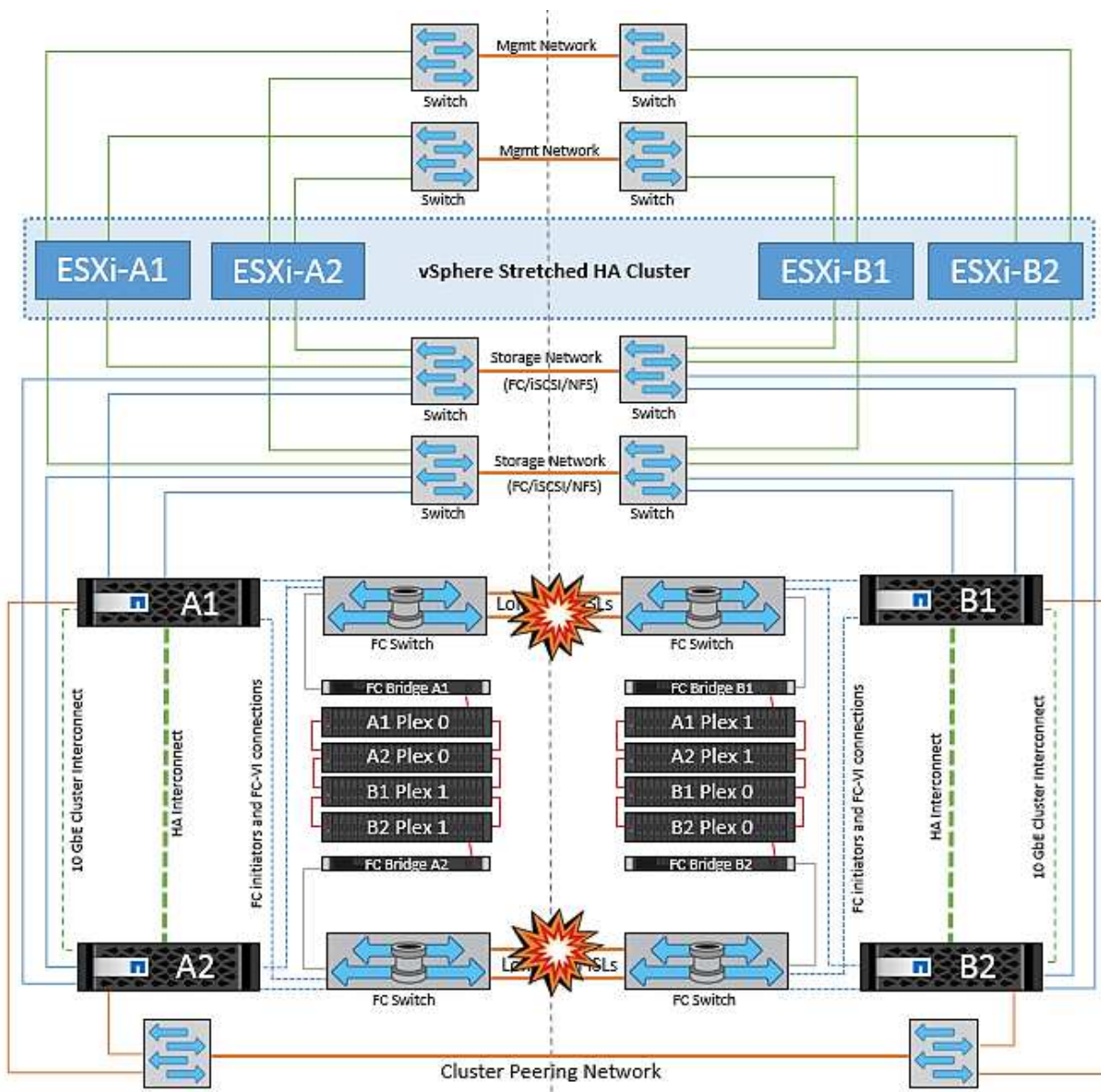
A NetApp recomenda determinar se alguma máquina virtual violou as regras do DRS. Todas as máquinas virtuais executadas a partir de um site remoto ficarão inativas, uma vez que não poderão acessar o datastore, e o vSphere HA reiniciará essa máquina virtual no site local. Depois que os links ISL estiverem novamente online, a máquina virtual que estava sendo executada no local remoto será morta, uma vez que não pode haver duas instâncias de máquinas virtuais executando com os mesmos endereços MAC.



Falha de ligação InterSwitch em ambas as malhas no NetApp MetroCluster

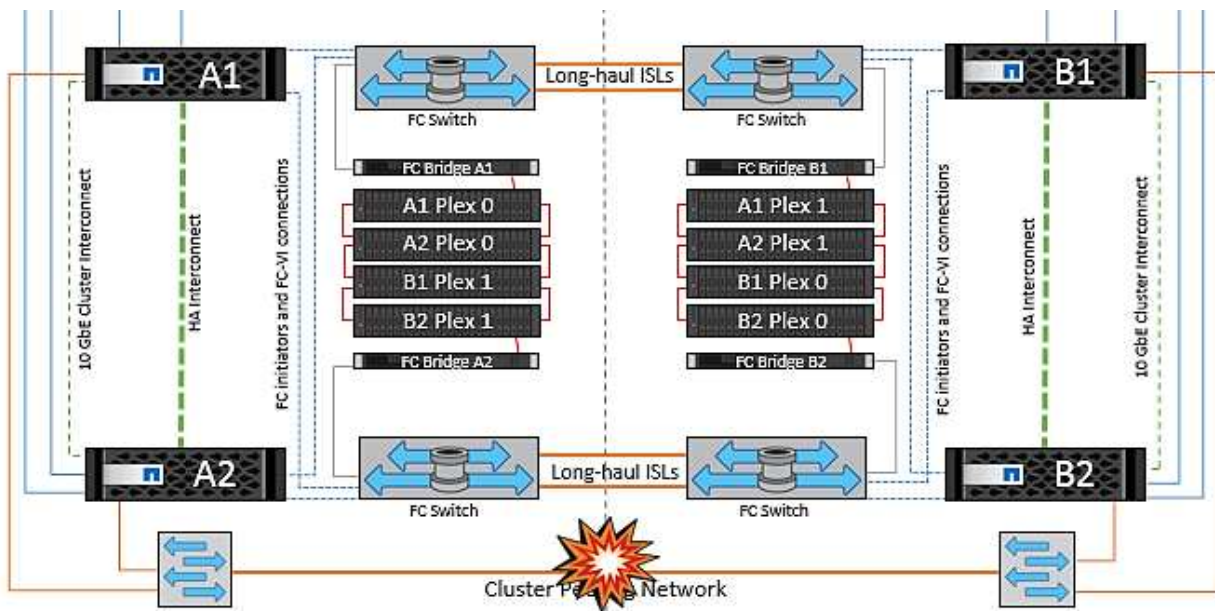
Em um cenário de falha de um ou mais ISLs, o tráfego continua através dos links restantes. Se todos os ISLs em ambas as malhas falharem, de modo que não haja nenhum link entre os locais para armazenamento e replicação do NVRAM, cada controladora continuará fornecendo seus dados locais. Na restauração de um mínimo de um ISL, a resincronização de todos os plexos acontecerá automaticamente.

Quaisquer gravações que ocorram depois de todos os ISLs estarem inativos não serão espelhadas para o outro site. Um switchover em caso de desastre, enquanto a configuração estiver nesse estado, incorreria, portanto, na perda dos dados que não haviam sido sincronizados. Neste caso, a intervenção manual é necessária para a recuperação após a mudança. Se for provável que nenhum ISLs esteja disponível por um período prolongado, um administrador pode optar por encerrar todos os serviços de dados para evitar o risco de perda de dados se for necessário um switchover em caso de desastre. A execução dessa ação deve ser ponderada contra a probabilidade de um desastre exigir mudança antes de pelo menos uma ISL ficar disponível. Alternativamente, se os ISLs estiverem falhando em um cenário em cascata, um administrador pode acionar um switchover planejado para um dos sites antes que todos os links tenham falhado.



Falha na ligação do cluster de peered

Em um cenário de falha de enlace de cluster com peered, como os ISLs de malha ainda estão ativos, os serviços de dados (leituras e gravações) continuam em ambos os locais para ambos os plexos. Qualquer alteração na configuração de cluster (por exemplo, adicionar um novo SVM, provisionar um volume ou LUN em um SVM existente) não pode ser propagado para o outro local. Estes são mantidos nos volumes de metadados locais do CRS e propagados automaticamente para o outro cluster após a restauração do link do cluster de peered. Se for necessário um switchover forçado antes que o link do cluster com peering possa ser restaurado, as alterações pendentes na configuração do cluster serão reproduzidas automaticamente da cópia replicada remota dos volumes de metadados no local que sobreviveu como parte do processo de switchover.



Falha no local completo

Em um local completo Um cenário de falha, os hosts ESXi no local B não receberão o heartbeat da rede dos hosts ESXi no local A porque estão inoperantes. O mestre de HA no local B verificará se os batimentos cardíacos do armazenamento de dados não estão presentes, declarará que os hosts no local A estão com falha e tentará reiniciar o local. Uma máquina virtual no local B. durante esse período, o administrador de storage executa um switchover para retomar os serviços dos nós com falha no local sobrevivente, o que restaurará todos os serviços de armazenamento do Local A no local B. após o local A volumes ou LUNs estarem disponíveis no local B, o agente de HA tentará reiniciar o local B.

Se a tentativa do agente mestre do vSphere HA de reiniciar uma VM (que envolve registrá-la e ligá-la) falhar, a reinicialização será novamente tentada após um atraso. O atraso entre reinicializações pode ser configurado até um máximo de 30 minutos. O vSphere HA tenta reiniciar para um número máximo de tentativas (seis tentativas por padrão).



O mestre de HA não inicia as tentativas de reinicialização até que o gerente de colocação encontre um armazenamento adequado, portanto, no caso de uma falha completa no local, isso seria depois que o switchover foi executado.

Se o local A tiver sido substituído, uma falha subsequente de um dos nós do local B sobreviventes pode ser tratada de forma otimizada pelo failover para o nó sobrevivente. Neste caso, o trabalho de quatro nós agora está sendo executado por apenas um nó. A recuperação neste caso consistiria em realizar um giveback para o nó local. Em seguida, quando o local A é restaurado, uma operação de switchover é executada para restaurar a operação de estado estável da configuração.

uma vulnerabilidade descoberta recentemente relatada a qualquer momento.

- **Verificação de vulnerabilidades.** O objetivo da verificação de vulnerabilidades é detetar vulnerabilidades de segurança comuns e conhecidas nos produtos NetApp antes de serem lançadas aos clientes.
- * **Teste de penetração.*** O teste de penetração é o processo de avaliação de um sistema, aplicativo da Web ou rede para encontrar vulnerabilidades de segurança que possam ser exploradas por um invasor. Os testes de penetração (testes de caneta) na NetApp são conduzidos por um grupo de empresas terceirizadas aprovadas e confiáveis. Seu escopo de teste inclui o lançamento de ataques contra um aplicativo ou software semelhante a intrusos hostis ou hackers usando métodos ou ferramentas de exploração sofisticados.

Recursos de segurança do produto

As ferramentas do ONTAP para VMware vSphere incluem os seguintes recursos de segurança em cada versão.

- * **Login banner.*** O SSH é desativado por padrão e só permite logins únicos se ativado a partir do console da VM. O banner de login a seguir é exibido depois que o usuário insere um nome de usuário no prompt de login:

AVISO: o acesso não autorizado a este sistema é proibido e será processado por lei. Ao acessar este sistema, você concorda que suas ações podem ser monitoradas se houver suspeita de uso não autorizado.

Depois que o usuário concluir o login através do canal SSH, o seguinte texto é exibido:

```
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.
```

- **Controle de acesso baseado em função (RBAC).** Dois tipos de controles RBAC estão associados às ferramentas do ONTAP:
 - Privileges nativo do vCenter Server
 - Privileges específico do plug-in do vCenter. Para obter detalhes, ["este link"](#) consulte .
- **Canais de comunicação criptografados.** Toda a comunicação externa acontece por HTTPS usando a versão 1,2 do TLS.
- * **Exposição mínima do porto.*** Apenas as portas necessárias estão abertas no firewall.

A tabela a seguir descreve os detalhes da porta aberta.

N.o da porta TCP v4/V6	Direção	Função
8143	de entrada	Conexões HTTPS para API REST
8043	de entrada	Conexões HTTPS

N.o da porta TCP v4/V6	Direção	Função
9060	de entrada	Conexões HTTPS usadas para SOAP em conexões https esta porta deve ser aberta para permitir que um cliente se conecte ao servidor API de ferramentas ONTAP.
22	de entrada	SSH (Desativado por padrão)
9080	de entrada	Conexões HTTPS - VP e SRA - conexões internas somente de loopback
9083	de entrada	Conexões HTTPS - VP e SRA usados para SOAP em conexões https
1162	de entrada	Pacotes de intercetção SNMP VP
1527	apenas interno	Porta de banco de dados Derby, apenas entre este computador e ele mesmo, conexões externas não aceitas — somente conexões internas
443	bidirecional	Usado para conexões com clusters ONTAP

- **Suporte para certificados assinados pela autoridade de certificação (CA).** As ferramentas do ONTAP para VMware vSphere oferecem suporte a certificados assinados pela CA. Consulte este ["artigo da kb"](#) documento para obter mais informações.
- **Registro de auditoria.** Os pacotes de suporte podem ser baixados e são extremamente detalhados. As ferramentas do ONTAP registram todas as atividades de login e logout do usuário em um arquivo de log separado. As chamadas de API VASA são registradas em um log de auditoria VASA dedicado (local cxf.log).
- **Políticas de senha.** As seguintes políticas de senha são seguidas:
 - As senhas não são registradas em nenhum arquivo de log.
 - As senhas não são comunicadas em texto simples.
 - As senhas são configuradas durante o próprio processo de instalação.
 - O histórico de senhas é um parâmetro configurável.
 - A idade mínima da senha é definida para 24 horas.
 - O preenchimento automático dos campos de senha está desativado.
 - As ferramentas do ONTAP criptografam todas as informações de credenciais armazenadas usando hash SHA256.

Plug-in do SnapCenter VMware vSphere

O plug-in do NetApp SnapCenter para a engenharia de software VMware vSphere usa as seguintes atividades de desenvolvimento seguro:

- **Modelagem de ameaças.** O objetivo da modelagem de ameaças é descobrir falhas de segurança em um recurso, componente ou produto no início do ciclo de vida do desenvolvimento de software. Um modelo de ameaça é uma representação estruturada de todas as informações que afetam a segurança de um aplicativo. Em essência, é uma visão da aplicação e do seu ambiente através da lente da segurança.
- **Teste dinâmico de segurança de aplicativos (DAST).** Tecnologias projetadas para detectar condições vulneráveis em aplicativos em seu estado de execução. O DAST testa as interfaces HTTP e HTML expostas de aplicações web-enable.
- **Moeda de código de terceiros.** Como parte do desenvolvimento de software e do uso de software de código aberto (OSS), é importante abordar vulnerabilidades de segurança que podem estar associadas ao OSS que foi incorporado ao seu produto. Este é um esforço contínuo, uma vez que a versão do componente OSS pode ter uma vulnerabilidade recentemente descoberta relatada a qualquer momento.
- **Verificação de vulnerabilidades.** O objetivo da verificação de vulnerabilidades é detectar vulnerabilidades de segurança comuns e conhecidas nos produtos NetApp antes de serem lançadas aos clientes.
- * Teste de penetração.* O teste de penetração é o processo de avaliação de um sistema, aplicativo da Web ou rede para encontrar vulnerabilidades de segurança que possam ser exploradas por um invasor. Os testes de penetração (testes de caneta) na NetApp são conduzidos por um grupo de empresas terceirizadas aprovadas e confiáveis. Seu escopo de teste inclui o lançamento de ataques contra um aplicativo ou software como intrusos hostis ou hackers usando métodos ou ferramentas de exploração sofisticados.
- **Atividade de resposta a incidentes de Segurança do produto.** Vulnerabilidades de segurança são descobertas interna e externamente para a empresa e podem representar um sério risco para a reputação da NetApp se não forem abordadas em tempo hábil. Para facilitar esse processo, uma equipe de resposta a incidentes de Segurança do produto (PSIRT) relata e rastreia as vulnerabilidades.

Recursos de segurança do produto

O plug-in do NetApp SnapCenter para VMware vSphere inclui os seguintes recursos de segurança em cada versão:

- **Acesso restrito ao shell.** O SSH é desativado por padrão, e logins únicos só são permitidos se estiverem ativados a partir do console da VM.
- **Aviso de acesso no banner de login.** O banner de login a seguir é exibido depois que o usuário insere um nome de usuário no prompt de login:

AVISO: o acesso não autorizado a este sistema é proibido e será processado por lei. Ao acessar este sistema, você concorda que suas ações podem ser monitoradas se houver suspeita de uso não autorizado.

Depois que o usuário concluir o login pelo canal SSH, a seguinte saída é exibida:

```
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.
```

- **Controle de acesso baseado em função (RBAC).** Dois tipos de controles RBAC estão associados às ferramentas do ONTAP:

- Privileges nativo do vCenter Server.
- Privileges específico do plug-in do VMware vCenter. Para obter mais informações, "[Controle de acesso baseado em função \(RBAC\)](#)" consulte .
- **Canais de comunicação criptografados.** Toda a comunicação externa acontece por HTTPS usando TLS.
- * Exposição mínima do porto.* Apenas as portas necessárias estão abertas no firewall.

A tabela a seguir fornece os detalhes da porta aberta.

Número da porta TCP v4/V6	Função
8144	Conexões HTTPS para API REST
8080	Conexões HTTPS para GUI OVA
22	SSH (desativado por padrão)
3306	MySQL (somente conexões internas; conexões externas desativadas por padrão)
443	Nginx (serviços de proteção de dados)

- **Suporte para certificados assinados pela Autoridade de Certificação (CA).** O plug-in do SnapCenter para VMware vSphere oferece suporte ao recurso de certificados assinados pela CA. "[Como criar e/ou importar um certificado SSL para o plug-in SnapCenter para VMware vSphere \(SCV\)](#)" Consulte .
- **Políticas de senha.** As seguintes políticas de senha estão em vigor:
 - As senhas não são registradas em nenhum arquivo de log.
 - As senhas não são comunicadas em texto simples.
 - As senhas são configuradas durante o próprio processo de instalação.
 - Todas as informações de credenciais são armazenadas usando hash SHA256.
- **Imagem base do sistema operacional.** O produto é fornecido com o sistema operacional base Debian para OVA com acesso restrito e acesso shell desativado. Isso reduz a ocupação física do ataque. Todos os sistemas operacionais baseados em versões do SnapCenter são atualizados com os patches de segurança mais recentes disponíveis para cobertura máxima de segurança.

A NetApp desenvolve recursos de software e patches de segurança em relação ao plug-in do SnapCenter para o dispositivo VMware vSphere e, em seguida, os lança aos clientes como uma plataforma de software agrupada. Como esses dispositivos incluem dependências específicas do sistema sub-operacional Linux, bem como nosso software proprietário, a NetApp recomenda que você não faça alterações no sistema sub-operacional porque isso tem um alto potencial para afetar o dispositivo NetApp. Isto pode afetar a capacidade do NetApp de suportar o aparelho. A NetApp recomenda testar e implantar nossa versão de código mais recente para dispositivos porque eles são lançados para corrigir quaisquer problemas relacionados à segurança.

Guia de fortalecimento da segurança para as ferramentas do ONTAP para VMware vSphere

Guia de fortalecimento da segurança para as ferramentas do ONTAP para VMware vSphere

O guia de proteção de segurança para as ferramentas do ONTAP para VMware vSphere fornece um conjunto abrangente de instruções para configurar as configurações mais seguras.

Estes guias aplicam-se aos aplicativos e ao SO convidado do próprio aparelho.

Verificando a integridade das ferramentas do ONTAP para os pacotes de instalação do VMware vSphere

Existem dois métodos disponíveis para os clientes verificarem a integridade de seus pacotes de instalação de ferramentas ONTAP.

1. Verificando as somas de verificação
2. Verificando a assinatura

As somas de verificação são fornecidas nas páginas de download dos pacotes de instalação do OTV. Os usuários devem verificar as somas de verificação dos pacotes baixados em relação à soma de verificação fornecida na página de download.

Verificando a assinatura das ferramentas ONTAP OVA

O pacote de instalação do vApp é entregue na forma de um tarball. Este tarball contém certificados intermediários e raiz para o dispositivo virtual, juntamente com um arquivo README e um pacote OVA. O arquivo README orienta os usuários sobre como verificar a integridade do pacote vApp OVA.

Os clientes também devem fazer o upload do certificado raiz e intermediário fornecido no vCenter versão 7.0U3E e superior. Para versões do vCenter entre 7.0.1 e 7,0.U3E, a funcionalidade de verificação de certificado não é suportada pela VMware. Os clientes não precisam carregar nenhum certificado para o vCenter versões 6.x.

Carregar o certificado raiz confiável para o vCenter

1. Faça login com o VMware vSphere Client no vCenter Server.
2. Especifique o nome de usuário e a senha para o administrador ou outro membro do grupo Administradores de logon único do vCenter. Se você especificou um domínio diferente durante a instalação, faça login como administrador.
3. Navegue até a interface do usuário Gerenciamento de certificados: a. no menu inicial, selecione Administração. b. em certificados, clique em Gerenciamento de certificados.
4. Se o sistema solicitar, insira as credenciais do vCenter Server.
5. Em certificados raiz confiáveis, clique em Adicionar.
6. Clique em Procurar e selecione a localização do ficheiro .pem do certificado (OTV_OVA_INTER_ROOT_CERT_Chain.pem).
7. Clique em Adicionar. O certificado é adicionado à loja.

["Adicione um certificado raiz confiável ao armazenamento de certificados"](#) Consulte para obter mais informações. Ao implantar um vApp (usando o arquivo OVA), a assinatura digital do pacote vApp pode ser verificada na página 'Detalhes de revisão'. Se o pacote vApp baixado for original, a coluna 'Publisher' exibe

'certificado confiável' (como na captura de tela a seguir).

Deploy OVF Template

- ✓ 1 Select an OVF template
- ✓ 2 Select a name and folder
- ✓ 3 Select a compute resource
- 4 Review details**
- 5 License agreements
- 6 Select storage
- 7 Select networks
- 8 Customize template
- 9 Ready to complete

Review details
Verify the template details.

Publisher	Entrust Code Signing CA - OVCS2 (Trusted certificate)
Product	Virtual Appliance - NetApp Inc. ONTAP tools for VMware vSphere
Version	See appliance for version
Vendor	NetApp Inc.
Description	Virtual Appliance - NetApp Inc. ONTAP tools for VMware vSphere for netapp storage systems. For more information or support please visit https://www.netapp.com/
Download size	2.2 GB
Size on disk	3.9 GB (thin provisioned) 53.0 GB (thick provisioned)

Activate
Go to Sys

CANCEL BACK NEXT

Verificando a assinatura das ferramentas ONTAP ISO e SRA tar.gz

A NetApp compartilha seu certificado de assinatura de código com os clientes na página de download do produto, juntamente com os arquivos zip do produto para OTV-ISO e SRA.tgz.

A partir do certificado de assinatura de código, os usuários podem extrair a chave pública como abaixo:

```
#> openssl x509 -in <code-sign-cert, pem file> -pubkey -noout > <public-key name>
```

Em seguida, a chave pública deve ser usada para verificar a assinatura para iso e tgz produto zip como abaixo:

```
#> openssl dgst -sha256 -verify <public-key> -signature <signature-file> <binary-name>
```

Exemplo:

```
#> 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-vmware-9.12-upgrade-iso.sig netapp-ontap-tools-for-vmware-
vmware-9.12-upgrade.iso
Verified OK => response
```

Portas e protocolos

Aqui estão listadas as portas e protocolos necessários que permitem a comunicação entre as ferramentas do ONTAP para o servidor VMware vSphere e outras entidades, como sistemas de storage gerenciados, servidores e outros componentes.

Portas de entrada e saída necessárias para o OTV

Observe a tabela abaixo que lista as portas de entrada e saída necessárias para o funcionamento adequado das ferramentas do ONTAP. É importante garantir que apenas as portas mencionadas na tabela estejam abertas para conexões de máquinas remotas, enquanto todas as outras portas devem ser bloqueadas para conexões de máquinas remotas. Isso ajudará a garantir a segurança e a segurança do seu sistema.

A tabela a seguir descreve os detalhes da porta aberta.

Número da porta TCP v4/V6	Direção	Função
8143	de entrada	Conexões HTTPS para API REST
8043	de entrada	Conexões HTTPS
9060	de entrada	Esta porta deve ser aberta para permitir que um cliente se conecte ao servidor de API do ONTAP Tools.
22	de entrada	SSH (Desativado por padrão)
9080	de entrada	Conexões HTTPS - VP e SRA - conexões internas somente de loopback
9083	de entrada	Conexões HTTPS - VP e SRA usados para SOAP em conexões HTTPS
1162	de entrada	Pacotes de intercetação SNMP VP
8443	de entrada	Plug-in remoto
1527	apenas interno	Porta de banco de dados Derby, apenas entre este computador e ele mesmo, conexões externas não aceitas - somente conexões internas
8150	apenas interno	O serviço de integridade de log é executado na porta
443	bidirecional	Usado para conexões com clusters ONTAP

Controlar o acesso remoto ao banco de dados Derby

Os administradores podem acessar o banco de dados derby com os seguintes comandos. Ele pode ser acessado através da VM local das ferramentas do ONTAP, bem como de um servidor remoto com as

seguintes etapas:

```
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>';
```

*[.Underline

```
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=██████████';  
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 |
```

Ferramentas do ONTAP para pontos de acesso do VMware vSphere (usuários)

A instalação das Ferramentas do ONTAP para VMware vSphere cria e usa três tipos de usuários:

1. Usuário do sistema: A conta de usuário raiz
2. Usuário do aplicativo: O usuário administrador, usuário de manutenção e contas de usuário de banco de dados
3. Usuário de suporte: A conta de usuário diag

1. Utilizador do sistema

O usuário System(root) é criado pela instalação de ferramentas ONTAP no sistema operacional subjacente (Debian).

- Um usuário padrão do sistema "root" é criado no Debian pela instalação de ferramentas do ONTAP. Seu padrão é desabilitado e pode ser habilitado em uma base ad-hoc através do console 'aint'.

2. Utilizador da aplicação

O usuário do aplicativo é nomeado como um usuário local nas ferramentas do ONTAP. Estes são usuários criados no aplicativo de ferramentas do ONTAP. A tabela abaixo lista os tipos de usuários de aplicativos:

Usuário	Descrição
Usuário Administrador	Ele é criado durante a instalação das ferramentas do ONTAP e o usuário fornece as credenciais ao implantar as ferramentas do ONTAP. Os usuários têm a opção de alterar a 'senha' no console 'não'. A senha expirará em 90 dias e espera-se que os usuários alterem a mesma.

Usuário	Descrição
Utilizador de manutenção	Ele é criado durante a instalação das ferramentas do ONTAP e o usuário fornece as credenciais ao implantar as ferramentas do ONTAP. Os usuários têm a opção de alterar a 'senha' no console 'não'. Este é um usuário de manutenção e é criado para executar as operações do console de manutenção.
Utilizador da base de dados	Ele é criado durante a instalação das ferramentas do ONTAP e o usuário fornece as credenciais ao implantar as ferramentas do ONTAP. Os usuários têm a opção de alterar a 'senha' no console 'não'. A senha expirará em 90 dias e espera-se que os usuários alterem a mesma.

3. Usuário do suporte (usuário diag)

Durante a instalação das ferramentas do ONTAP, um usuário de suporte é criado. Esse usuário pode ser usado para acessar as ferramentas do ONTAP em caso de qualquer problema ou falha no servidor e para coletar logs. Por padrão, esse usuário está desativado, mas pode ser habilitado em uma base adhoc através do console 'mint'. É importante notar que este utilizador será automaticamente desativado após um determinado período de tempo.

TLS mútuo (autenticação baseada em certificado)

As versões 9,7 e posteriores do ONTAP suportam comunicação TLS mútua. A partir das Ferramentas do ONTAP para VMware e vSphere 9,12, o TLS mútuo é usado para comunicação com clusters recém-adicionados (dependendo da versão do ONTAP).

ONTAP

Para todos os sistemas de armazenamento adicionados anteriormente: Durante uma atualização, todos os sistemas de armazenamento de dados adicionados serão auto-confiáveis e os mecanismos de autenticação baseados em certificado serão configurados.

Como na captura de tela abaixo, a página de configuração do cluster mostrará o status de TLS mútuo (autenticação baseada em certificado), configurado para cada cluster.

Name	Type	IP Address	ONTAP Release	Status	Capacity	NFS VAAI	Supported Protocols
CL_sti2l-vsim-ucs50im_1678878260	Cluster	10.234.85.142	9.12.0	Normal	20.42%		

Cluster Add

Durante o fluxo de trabalho de adição de cluster, se o cluster que está sendo adicionado suportar MTLS, o MTLS será configurado por padrão. O usuário não precisa fazer nenhuma configuração para isso. A captura

de tela abaixo mostra a tela apresentada ao usuário durante a adição do cluster.

Add Storage System

i 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: _____

Username: _____

Password: _____

Port: 443 _____

Advanced options ▲

ONTAP Cluster Certificate: Automatically fetch Manually upload

CANCEL ADD

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

10.234.85.52

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

Authorize Cluster Certificate

Host 10.234.85.142 has identified itself with a self-signed certificate.

[Hide certificate](#)

Certificate Information

This certificate identifies the 10.234.85.142 host.

Issued By

Name (CN or DN): C1_sti21-vsimsim-ucs581m_1678878260

Issued To

Name (CN or DN): C1_sti21-vsimsim-ucs581m_1678878260

Validity

Issued On: 03/15/2023 11:16:06

Expires On: 03/14/2024 11:16:06

Fingerprint Information

SHA-1 Fingerprint: 2C:38:E3:5C:4B:F3:5D:3F:39:C8:CE:4A:8
2:C1:A6:EE:34:53:A0:F3

SHA-256 Fingerprint: 05:0F:FE:CD:B0:C6:FC:6F:EB:8A:FC:86:F
7:E3:EF:D4:8D:CA:02:92:9B:E1:A4:70:84:
52:F8:76:98:64:FA:23

Do you want to trust this certificate?

NO

YES

Cluster Edit (Editar cluster)

Durante a operação de edição de cluster, existem dois cenários:

- Se o certificado ONTAP expirar, o usuário terá que obter o novo certificado e enviá-lo.
- Se o certificado OTV expirar, o usuário pode regenerá-lo marcando a caixa de seleção.
 - *Gerar um novo certificado de cliente para o ONTAP.*

Modify Storage System

Settings Provisioning Options

IP address or hostname: ▼

Port:

Username:

Password:

Upload Certificate (Optional) [BROWSE](#)

Skip monitoring of this storage system

Generate a new client certificate for ONTAP

CANCEL

OK



Certificado HTTPS de ferramentas ONTAP

Por padrão, as ferramentas do ONTAP usam um certificado autoassinado criado automaticamente durante a instalação para proteger o acesso HTTPS à IU da Web. As ferramentas do ONTAP oferecem os seguintes recursos:

1. Regenere o certificado HTTPS

Durante a instalação das ferramentas do ONTAP, um certificado de CA HTTPS é instalado e o certificado é armazenado no keystore. O usuário tem a opção de regenerar o certificado HTTPS por meio do console de manutenção.

As opções acima podem ser acessadas no console *mant* navegando até '*Application Configuration*' → '*Re-Generate certificates*'.

Banner de login

O banner de login a seguir é exibido depois que o usuário insere um nome de usuário no

prompt de login. Observe que o SSH está desativado por padrão e só permite logins únicos quando ativado a partir do console da 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.
```

Depois que o usuário concluir o login através do canal SSH, o seguinte texto é exibido:

```
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.
```

Tempo limite de inatividade

Para impedir o acesso não autorizado, é configurado um tempo limite de inatividade, que faz logout automaticamente de usuários inativos por um determinado período durante o uso de recursos autorizados. Isso garante que somente usuários autorizados possam acessar os recursos e ajuda a manter a segurança.

- Por padrão, as sessões do vSphere Client são encerradas após 120 minutos de tempo ocioso, exigindo que o usuário faça login novamente para continuar usando o cliente. Você pode alterar o valor do tempo limite editando o arquivo `webclient.properties`. Você pode configurar o tempo limite do vSphere Client "[Configure o valor de tempo limite do vSphere Client](#)".
- As ferramentas do ONTAP têm um tempo de logout de sessão da web-cli de 30 minutos.

Máximo de solicitações simultâneas por usuário (Network security protection/ataque dos)

Por padrão, o número máximo de solicitações simultâneas por usuário é 48. O usuário raiz nas ferramentas do ONTAP pode alterar esse valor dependendo dos requisitos de seu ambiente. **Este valor não deve ser definido para um valor muito alto, pois isso fornece um mecanismo contra ataques de negação de serviço (dos).**

Os usuários podem alterar o número máximo de sessões simultâneas e outros parâmetros suportados no arquivo `/opt/NetApp/vscserver/etc/dosfilterParams.json`.

Podemos configurar o filtro seguindo os seguintes parâmetros:

- **delayMS**: O atraso em milissegundos dado a todas as solicitações acima do limite de taxa antes de serem consideradas. Dê -1 apenas para rejeitar o pedido.
- **estrangulems**: Quanto tempo para sincronizar a espera pelo semáforo.
- **maxRequestMs**: Quanto tempo para permitir que essa solicitação seja executada.
- **ipWhitelist**: Uma lista separada por vírgulas de endereços IP que não serão limitados por taxa. (Isso pode ser vCenter, ESXi e SRA IPs)
- **maxRequestsPerSec**: O número máximo de solicitações de uma conexão por segundo.

Valores padrão no arquivo `dosfilterParams`:

```
{"delayMs": "-1",  
"throttleMs": "1800000",  
"maxRequestMs": "300000",  
"ipWhitelist": "10.224.58.52",  
"maxRequestsPerSec": "48"}
```

Configuração do NTP (Network Time Protocol)

Às vezes, problemas de segurança podem ocorrer devido a discrepâncias nas configurações de tempo de rede. É importante garantir que todos os dispositivos dentro de uma rede tenham configurações de tempo precisas para evitar esses problemas.

Virtual Appliance

Você pode configurar o(s) servidor(s) NTP a partir do console de manutenção no dispositivo virtual. Os usuários podem adicionar os detalhes do servidor NTP em *Configuração do sistema* > *Adicionar novo servidor NTP* opção

Por padrão, o serviço para NTP é `ntpd`. Este é um serviço legado e não funciona bem para máquinas virtuais em certos casos.

Debian

No Debian, o usuário pode acessar o arquivo `/etc/NTP.conf` para obter detalhes do servidor `ntp`.

Políticas de senha

Os usuários que implantarem ferramentas do ONTAP pela primeira vez ou atualizarem para a versão 9,12 ou posterior precisarão seguir a política de senha forte para o administrador e os usuários do banco de dados. Durante o processo de implantação, novos usuários serão solicitados a inserir suas senhas. Para usuários `brownfield` atualizando para a versão 9,12 ou posterior, a opção de seguir a política de senha forte estará disponível no console de manutenção.

- Uma vez que o usuário faça login no console de manutenção, as senhas serão verificadas em relação ao conjunto de regras complexas e, se for encontrado que não foi seguido, o usuário será solicitado a

redefinir o mesmo.

- A validade padrão da senha é de 90 dias e após 75 dias o usuário começará a receber a notificação para alterar a senha.
- É necessário definir uma nova senha em cada ciclo, o sistema não receberá a última senha como a nova senha.
- Sempre que um usuário fizer login no console de manutenção, ele verificará as políticas de senha, como as capturas de tela abaixo, antes de carregar o Menu Principal:

```
Maintenance Console : "MetApp ONTAP tools for VMware vSphere"  
Discovered interfaces: eth0 (ENABLED)  
validating password policies
```

- Se não for encontrado seguindo a política de senha ou sua configuração de atualização a partir das ferramentas ONTAP 9,11 ou antes. Em seguida, o utilizador verá o seguinte ecrã para repor a palavra-passe:

```
Your Administrator and Database password is expired or does not match password policy:  
1 ) Change 'administrator' user password  
2 ) Change database password  
x ) Exit  
Enter your choice: _
```

- Se o usuário tentar definir senha fraca ou fornecer a última senha novamente, o usuário verá o seguinte erro:

```
Changing password for administrator.  
User: administrator  
Enter new password:  
Retype new password:  
  
Password doesn't matches the password policy.  
For security reasons, it is recommended to use a password that is of eight to thirty characters and  
contains a minimum of one upper, one lower, one digit, and one special character.  
  
Enter new password:  
Retype new password:  
Check if new decoder works ?  
New decoder worked successfully  
08-02/23 13:36:53 Your new password must be different  
  
Error updating sra credential file  
  
Press ENTER to continue._
```

Avisos legais

Avisos legais fornecem acesso a declarações de direitos autorais, marcas registradas, patentes e muito mais.

Direitos de autor

["https://www.netapp.com/company/legal/copyright/"](https://www.netapp.com/company/legal/copyright/)

Marcas comerciais

NetApp, o logotipo DA NetApp e as marcas listadas na página de marcas comerciais da NetApp são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.

["https://www.netapp.com/company/legal/trademarks/"](https://www.netapp.com/company/legal/trademarks/)

Patentes

Uma lista atual de patentes de propriedade da NetApp pode ser encontrada em:

<https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf>

Política de privacidade

["https://www.netapp.com/company/legal/privacy-policy/"](https://www.netapp.com/company/legal/privacy-policy/)

Código aberto

Os arquivos de aviso fornecem informações sobre direitos autorais de terceiros e licenças usadas no software NetApp.

ONTAP

["Aviso para ONTAP 9.13,1"](#) ["Aviso para ONTAP 9.12,1"](#) ["Aviso para ONTAP 9.12,0"](#) ["Aviso para ONTAP 9.11,1"](#) ["Aviso para ONTAP 9.10,1"](#) ["Aviso para ONTAP 9.10,0"](#) ["Aviso para ONTAP 9.9,1"](#) ["Aviso para ONTAP 9.8"](#) ["Aviso para ONTAP 9.7"](#) ["Aviso para ONTAP 9.6"](#) ["Aviso para ONTAP 9.5"](#) ["Aviso para ONTAP 9.4"](#) ["Aviso para ONTAP 9.3"](#) ["Aviso para ONTAP 9.2"](#) ["Aviso para ONTAP 9.1"](#)

Mediador ONTAP para IP MCC

["9.9.1 Aviso para o Mediador ONTAP para IP MCC"](#) ["9,8 Aviso para o Mediador ONTAP para IP MCC"](#) ["9,7 Aviso para o Mediador ONTAP para IP MCC"](#)

Informações sobre direitos autorais

Copyright © 2024 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

O software derivado do material da NetApp protegido por direitos autorais está sujeito à seguinte licença e isenção de responsabilidade:

ESTE SOFTWARE É FORNECIDO PELA NETAPP "NO PRESENTE ESTADO" E SEM QUAISQUER GARANTIAS EXPRESSAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO E ADEQUAÇÃO A UM DETERMINADO PROPÓSITO, CONFORME A ISENÇÃO DE RESPONSABILIDADE DESTES DOCUMENTOS. EM HIPÓTESE ALGUMA A NETAPP SERÁ RESPONSÁVEL POR QUALQUER DANO DIRETO, INDIRETO, INCIDENTAL, ESPECIAL, EXEMPLAR OU CONSEQUENCIAL (INCLUINDO, SEM LIMITAÇÕES, AQUISIÇÃO DE PRODUTOS OU SERVIÇOS SOBRESSALIENTES; PERDA DE USO, DADOS OU LUCROS; OU INTERRUPTÃO DOS NEGÓCIOS), INDEPENDENTEMENTE DA CAUSA E DO PRINCÍPIO DE RESPONSABILIDADE, SEJA EM CONTRATO, POR RESPONSABILIDADE OBJETIVA OU PREJUÍZO (INCLUINDO NEGLIGÊNCIA OU DE OUTRO MODO), RESULTANTE DO USO DESTES SOFTWARES, MESMO SE ADVERTIDA DA RESPONSABILIDADE DE TAL DANO.

A NetApp reserva-se o direito de alterar quaisquer produtos descritos neste documento, a qualquer momento e sem aviso. A NetApp não assume nenhuma responsabilidade nem obrigação decorrentes do uso dos produtos descritos neste documento, exceto conforme expressamente acordado por escrito pela NetApp. O uso ou a compra deste produto não representam uma licença sob quaisquer direitos de patente, direitos de marca comercial ou quaisquer outros direitos de propriedade intelectual da NetApp.

O produto descrito neste manual pode estar protegido por uma ou mais patentes dos EUA, patentes estrangeiras ou pedidos pendentes.

LEGENDA DE DIREITOS LIMITADOS: o uso, a duplicação ou a divulgação pelo governo estão sujeitos a restrições conforme estabelecido no subparágrafo (b)(3) dos Direitos em Dados Técnicos - Itens Não Comerciais no DFARS 252.227-7013 (fevereiro de 2014) e no FAR 52.227- 19 (dezembro de 2007).

Os dados aqui contidos pertencem a um produto comercial e/ou serviço comercial (conforme definido no FAR 2.101) e são de propriedade da NetApp, Inc. Todos os dados técnicos e software de computador da NetApp fornecidos sob este Contrato são de natureza comercial e desenvolvidos exclusivamente com despesas privadas. O Governo dos EUA tem uma licença mundial limitada, irrevogável, não exclusiva, intransferível e não sublicenciável para usar os Dados que estão relacionados apenas com o suporte e para cumprir os contratos governamentais desse país que determinam o fornecimento de tais Dados. Salvo disposição em contrário no presente documento, não é permitido usar, divulgar, reproduzir, modificar, executar ou exibir os dados sem a aprovação prévia por escrito da NetApp, Inc. Os direitos de licença pertencentes ao governo dos Estados Unidos para o Departamento de Defesa estão limitados aos direitos identificados na cláusula 252.227-7015(b) (fevereiro de 2014) do DFARS.

Informações sobre marcas comerciais

NETAPP, o logotipo NETAPP e as marcas listadas em <http://www.netapp.com/TM> são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.