



FlexPod NetApp con NVIDIA

NetApp Solutions

NetApp
August 14, 2024

Sommario

- NetApp ai con NVIDIA 1
 - NetApp FlexPod con sistemi NVIDIA DGX 1
 - NetApp ONTAP ai con sistemi NVIDIA DGX A100 1
 - NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum 1
 - NetApp FlexPod con sistemi NVIDIA DGX - Introduzione 1
 - NVA-1151-DESIGN: Guida alla progettazione dei sistemi NetApp ONTAP ai con NVIDIA DGX A100 11
 - NVA-1151-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100 12
 - NVA-1153-DESIGN: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum 12
 - NVA-1153-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum 12

NetApp ai con NVIDIA

Panoramica delle soluzioni di infrastruttura convergente ONTAP ai di NetApp e NVIDIA.

NetApp FlexPod con sistemi NVIDIA DGX

- ["Pod NetApp ai con sistemi NVIDIA DGX"](#)

NetApp ONTAP ai con sistemi NVIDIA DGX A100

- ["Guida alla progettazione"](#)
- ["Guida all'implementazione"](#)

NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum

- ["Guida alla progettazione"](#)
- ["Guida all'implementazione"](#)

NetApp FlexPod con sistemi NVIDIA DGX - Introduzione

Questa sezione fornisce un'introduzione all'FlexPod di NetApp con i sistemi NVIDIA DGX.

Progettazione delle soluzioni NetApp

NetApp™ AIPod con NVIDIA DGX™ sistemi e sistemi storage connessi al cloud di NetApp, semplifica le implementazioni delle infrastrutture per i workload di machine learning (ML) e intelligenza artificiale (ai), eliminando la complessità della progettazione e le congetture. Basandosi sul design di NVIDIA DGX BasePOD per offrire performance di calcolo eccezionali per i carichi di lavoro di prossima generazione, AIPod con sistemi NVIDIA DGX aggiunge sistemi storage NetApp AFF che consentono ai clienti di partire con poco e crescere senza interruzioni gestendo intelligente i dati dall'edge al core, al cloud e viceversa. NetApp FlexPod fa parte del portfolio più ampio di soluzioni ai di NetApp, mostrato nella figura sotto

Portfolio di soluzioni ai di NetApp immagine::aipod_nv_portfolio.png[]

Questo documento descrive i componenti chiave dell'architettura di riferimento di AIPod, le informazioni sulla connettività dei sistemi e le linee guida sul dimensionamento delle soluzioni. Questo documento è destinato ai tecnici di soluzioni per partner e NetApp e ai decision maker strategici per i clienti interessati all'implementazione di un'infrastruttura dalle performance elevate per i workload di analytics e ML/DL.

NetApp FlexPod con sistemi NVIDIA DGX - Introduzione

Questa sezione fornisce un'introduzione all'FlexPod di NetApp con i sistemi NVIDIA DGX.

Progettazione delle soluzioni NetApp

NetApp™ AIPod con NVIDIA DGX™ sistemi e sistemi storage connessi al cloud di NetApp, semplifica le

implementazioni delle infrastrutture per i workload di machine learning (ML) e intelligenza artificiale (ai), eliminando la complessità della progettazione e le congetture. Basandosi sul design di NVIDIA DGX BasePOD per offrire performance di calcolo eccezionali per i carichi di lavoro di prossima generazione, AIPod con sistemi NVIDIA DGX aggiunge sistemi storage NetApp AFF che consentono ai clienti di partire con poco e crescere senza interruzioni gestendo intelligente i dati dall'edge al core, al cloud e viceversa. NetApp FlexPod fa parte del portfolio più ampio di soluzioni ai di NetApp, mostrato nella figura sotto

Portfolio di soluzioni ai di NetApp immagine::aipod_nv_portfolio.png[]

Questo documento descrive i componenti chiave dell'architettura di riferimento di AIPod, le informazioni sulla connettività dei sistemi e le linee guida sul dimensionamento delle soluzioni. Questo documento è destinato ai tecnici di soluzioni per partner e NetApp e ai decision maker strategici per i clienti interessati all'implementazione di un'infrastruttura dalle performance elevate per i workload di analytics e ML/DL.

NetApp FlexPod con sistemi NVIDIA DGX - componenti hardware

Questa sezione si concentra sui componenti hardware per l'FlexPod con sistemi NVIDIA DGX NetApp.

Sistemi storage di NetApp AFF

I sistemi storage NetApp AFF, all'avanguardia, consentono ai reparti IT di soddisfare i requisiti di storage Enterprise con performance leader di settore, flessibilità superiore, integrazione del cloud e gestione dei dati Best-in-class. Progettati appositamente per la tecnologia flash, i sistemi AFF aiutano ad accelerare, gestire e proteggere i dati business-critical.

Sistemi storage AFF A900

NetApp AFF A900, basato sul software per la gestione dei dati NetApp ONTAP, offre data Protection integrata, funzionalità anti-ransomware opzionali e le performance elevate e la resilienza necessarie per supportare i workload aziendali più critici. Elimina le interruzioni alle operazioni mission-critical, riduce al minimo il tuning delle performance e protegge i tuoi dati dagli attacchi ransomware. Offre:

- Prestazioni leader nel settore
- Sicurezza dei dati senza compromessi
- Aggiornamenti semplificati senza interruzioni

Sistema di archiviazione NetApp AFF A900 immagine::aipod_nv_A900.png[]

Performance leader di settore

AFF A900 gestisce con facilità i workload di prossima generazione come deep learning, ai e analytics ad alta velocità, oltre a database aziendali tradizionali come Oracle, SAP HANA, Microsoft SQL Server e applicazioni virtualizzate. Mantiene le applicazioni business-critical in funzione alla massima velocità con fino a 2,4M IOPS per coppia ha e latenza fino a 100µs, oltre ad aumentare le performance fino al 50% rispetto ai precedenti modelli NetApp. Grazie a NFS su RDMA, pNFS e Session Trunking, i clienti possono ottenere l'elevato livello di prestazioni di rete richiesto per le applicazioni di nuova generazione utilizzando l'infrastruttura di rete del data center esistente. I clienti possono anche scalare e crescere con un supporto multiprotocollo unificato per SAN, NAS e storage a oggetti, oltre a offrire la massima flessibilità con un software ONTAP unificato e singolo per la gestione dei dati on-premise o nel cloud. Inoltre, la salute del sistema può essere ottimizzata con analytics predittivi basati su ai forniti dal consulente digitale Active IQ (o consulente digitale) e da Cloud Insights.

Sicurezza dei dati senza compromessi

I sistemi AFF A900 offrono una suite completa di software per la data Protection integrata e coerente con l'applicazione di NetApp. Offre una data Protection integrata e soluzioni anti-ransomware all'avanguardia per prevenire e ripristinare in caso di attacchi. È possibile bloccare la scrittura su disco di file pericolosi e monitorare con facilità le anomalie dello storage per ottenere informazioni utili.

Upgrade semplificati senza interruzioni

AFF A900 è disponibile come aggiornamento in-chassis senza interruzioni per i clienti A700 esistenti. NetApp semplifica il refresh ed elimina le interruzioni delle operazioni mission-critical grazie alle nostre avanzate caratteristiche RASM (affidabilità, disponibilità, praticità e gestibilità). Inoltre, NetApp aumenta ulteriormente l'efficienza delle operazioni e semplifica le attività quotidiane dei team IT, perché il software ONTAP applica automaticamente gli aggiornamenti del firmware per tutti i componenti del sistema.

Per le implementazioni più grandi, i sistemi AFF A900 offrono le opzioni più elevate in termini di performance e capacità, mentre altri sistemi storage NetApp, come AFF A800, AFF C800, AFF A400, AFF C400 e AFF A250, offrono opzioni per implementazioni più piccole a prezzi più bassi.

NVIDIA DGX BasePOD

NVIDIA DGX BasePOD è una soluzione integrata che consiste di componenti hardware e software NVIDIA, soluzioni MLOps e storage di terze parti. Sfruttando le Best practice della progettazione di sistemi scale-out con i prodotti NVIDIA e le soluzioni validate dei partner, i clienti possono implementare una piattaforma efficiente e gestibile per lo sviluppo dell'intelligenza artificiale. La Figura 1 evidenzia i vari componenti di NVIDIA DGX BasePOD.

NVIDIA DGX BasePOD Solution immagine::aipod_nv_basepod_layers.png[]

SISTEMI NVIDIA DGX H100

Il sistema NVIDIA DGX H100™ è la potenza dell'AI che viene accelerata dalle rivoluzionarie performance della GPU NVIDIA H100 Tensor Core.

Immagine sistema NVIDIA DGX H100::aipod_nv_H100_3D.png[]

Le specifiche chiave del sistema DGX H100 sono:

- Otto GPU NVIDIA H100.
- 80 GB di memoria GPU per GPU, per un totale di 640GB.
- Quattro chip NVIDIA NVSwitch™.
- Due processori Intel® Xeon® Platinum 8480 a 56 core con supporto PCIe 5,0.
- 2 TB di memoria di sistema da DDR5 GB.
- Quattro porte OSFP che servono otto adattatori NVIDIA ConnectX-7 (InfiniBand/Ethernet) a una porta e due adattatori NVIDIA ConnectX-7 (InfiniBand/Ethernet) a due porte.
- Due unità da 1,92 TB M,2 NVMe per DGX OS, otto unità da 3,84 TB U,2 NVMe per storage/cache.
- 10,2 kW di potenza massima.

Le porte posteriori del vassoio CPU DGX H100 sono mostrate di seguito. Quattro delle porte OSFP servono otto adattatori ConnectX-7 per il fabric di calcolo InfiniBand. Ogni coppia di adattatori ConnectX-7 dual-port fornisce percorsi paralleli ai fabric di storage e gestione. La porta out-of-band viene utilizzata per l'accesso a BMC.

Immagine pannello posteriore NVIDIA DGX H100::aipod_nv_H100_REAR.png[]

Rete NVIDIA

Switch NVIDIA Quantum-2 QM9700

Switch InfiniBand NVIDIA Quantum-2 QM9700 immagine::aipod_nv_QM9700.png

Gli switch NVIDIA Quantum-2 QM9700 con connettività InfiniBand da 400GB GB/s alimentano il fabric di calcolo nelle configurazioni NVIDIA Quantum-2 InfiniBand BasePOD. Gli adattatori a porta singola ConnectX-7 vengono utilizzati per il fabric di calcolo InfiniBand. Ogni sistema NVIDIA DGX ha doppie connessioni a ogni switch QM9700, offrendo percorsi multipli a elevata larghezza di banda e bassa latenza tra i sistemi.

Switch NVIDIA Spectrum-3 SN4600

NVIDIA Spectrum-3 SN4600 switch image::aipod_nv_SN4600_hires_minimal.png

Gli switch NVIDIA Spectrum-3 SN4600 offrono un totale di 128 porte (64 per switch) per fornire connettività ridondante per la gestione in banda dei sistemi DGX BasePOD. Lo switch NVIDIA SN4600 è in grado di fornire velocità comprese tra 1 GbE e 200 GbE. Per le appliance di storage collegate tramite Ethernet, vengono utilizzati anche gli switch NVIDIA SN4600. Le porte degli adattatori NVIDIA DGX dual-port ConnectX-7 sono utilizzate sia per la gestione in banda che per la connettività dello storage.

Switch NVIDIA Spectrum SN2201 GbE

NVIDIA Spectrum SN2201 switch immagine::aipod_nv_SN2201.png

Gli switch NVIDIA Spectrum SN2201 offrono 48 porte per fornire connettività per la gestione fuori banda. La gestione out-of-band fornisce una connettività di gestione consolidata per tutti i componenti in DGX BasePOD.

Adattatore NVIDIA ConnectX-7

Scheda NVIDIA ConnectX-7 immagine::aipod_nv_CX7.png

L'adattatore NVIDIA ConnectX-7 è in grado di fornire un throughput di 25 GB/400G GB/100 GB/200 GB/50 GB. I sistemi NVIDIA DGX utilizzano gli adattatori ConnectX-7 a porta singola e doppia per garantire la flessibilità nelle implementazioni DGX BasePOD con InfiniBand da 400GB GB/s ed Ethernet 100/200GB.

NetApp FlexPod con sistemi NVIDIA DGX - componenti software

Questa sezione si concentra sui componenti software dell'FlexPod con sistemi NVIDIA DGX NetApp.

Software NVIDIA

Comando base NVIDIA

NVIDIA base Command™ è alla base di ogni DGX BasePOD, che consente alle organizzazioni di sfruttare il meglio dell'innovazione software NVIDIA. Le aziende possono liberare il pieno potenziale del proprio investimento con una piattaforma comprovata che include orchestrazione di livello Enterprise e gestione dei cluster, librerie che accelerano calcolo, infrastruttura di storage e rete e un sistema operativo ottimizzato per i carichi di lavoro ai.

Soluzione NVIDIA BaseCommand immagine::aipod_nv_BaseCommand_new.png

NVIDIA GPU CLOUD (NGC)

NVIDIA NGC™ fornisce un software per soddisfare le esigenze di data scientist, sviluppatori e ricercatori con vari livelli di esperienza nel campo dell'intelligenza artificiale. Il software ospitato su NGC viene sottoposto a scansioni contro un insieme aggregato di vulnerabilità ed esposizioni (CVE), crittografia e chiavi private comuni. È stata testata e progettata per scalare a più GPU e, in molti casi, a più nodi, garantendo agli utenti di massimizzare il proprio investimento nei sistemi DGX.

Immagine *NVIDIA GPU Cloud*::aipod_nv_ngc.png[]

NVIDIA ai Enterprise

NVIDIA ai Enterprise è la piattaforma software end-to-end che porta l'ai generativa nella portata di ogni azienda, offrendo il runtime più veloce ed efficiente per modelli di base ai generativi ottimizzati per l'esecuzione sulla piattaforma NVIDIA DGX. Con sicurezza, stabilità e gestibilità di livello produttivo, ottimizza lo sviluppo di soluzioni ai generative. NVIDIA ai Enterprise è inclusa con DGX BasePOD per gli sviluppatori aziendali per accedere a modelli preformati, framework ottimizzati, microservizi, librerie accelerate e supporto Enterprise.

Software NetApp

NetApp ONTAP

ONTAP 9, l'ultima generazione di software per la gestione dello storage NetApp, consente alle aziende di modernizzare l'infrastruttura e passare a un data center predisposto per il cloud. Sfruttando le funzionalità di gestione dei dati leader del settore, ONTAP consente la gestione e la protezione dei dati con un singolo set di strumenti, indipendentemente dalla posizione dei dati. Puoi anche spostare liberamente i dati ovunque siano necessari: Edge, core o cloud. ONTAP 9 include numerose funzionalità che semplificano la gestione dei dati, accelerano e proteggono i dati critici e abilitano le funzionalità dell'infrastruttura di nuova generazione nelle architetture di cloud ibrido.

Accelera e proteggi i dati

ONTAP offre livelli superiori di performance e protezione dei dati ed estende queste funzionalità nei seguenti modi:

- Performance e latenza ridotta. ONTAP offre il massimo throughput possibile con la latenza più bassa possibile, incluso il supporto per NVIDIA GPUDirect Storage (GDS) utilizzando NFS su RDMA, Parallel NFS (pNFS) e NFS session trunking.
- Protezione dei dati. ONTAP offre funzionalità di data Protection integrate e la più solida garanzia anti-ransomware del settore con una gestione comune su tutte le piattaforme.
- NetApp Volume Encryption (NVE). ONTAP offre crittografia nativa a livello di volume con supporto per la gestione delle chiavi sia integrata che esterna.
- Multi-tenancy dello storage e autenticazione multifattore. ONTAP consente la condivisione delle risorse dell'infrastruttura con i massimi livelli di sicurezza.

Semplifica la gestione dei dati

La gestione dei dati è fondamentale per le operazioni IT aziendali e per i data scientist, in modo che le risorse appropriate vengano utilizzate per le applicazioni ai e per la formazione dei set di dati ai/ML. Le seguenti informazioni aggiuntive sulle tecnologie NetApp non rientrano nell'ambito di questa convalida, ma potrebbero essere rilevanti a seconda dell'implementazione.

Il software per la gestione dei dati ONTAP include le seguenti funzionalità per ottimizzare e semplificare le operazioni e ridurre il costo totale delle operazioni:

- Snapshot e cloni favoriscono la collaborazione, la sperimentazione parallela e una governance dei dati migliorata per i flussi di lavoro ML/DL.
- SnapMirror consente lo spostamento perfetto dei dati negli ambienti di cloud ibrido e multisito, fornendo i dati dove e quando è necessario.
- Compaction dei dati inline e deduplica estesa. La compattazione dei dati riduce lo spazio sprecato all'interno dei blocchi di storage e la deduplica aumenta significativamente la capacità effettiva. Ciò vale per i dati memorizzati localmente e per i dati a più livelli nel cloud.
- Qualità del servizio (AQoS) minima, massima e adattativa. I controlli granulari della qualità del servizio (QoS) aiutano a mantenere i livelli di performance per le applicazioni critiche in ambienti altamente condivisi.
- I NetApp FlexGroup consentono la distribuzione dei dati in tutti i nodi del cluster storage, offrendo capacità elevata e performance più elevate per set di dati estremamente grandi.
- NetApp FabricPool. Offre il tiering automatico dei dati cold per le opzioni di cloud storage pubblico e privato, tra cui Amazon Web Services (AWS), Azure e la soluzione di storage NetApp StorageGRID. Per ulteriori informazioni su FabricPool, vedere "[TR-4598: Best practice FabricPool](#)".
- NetApp FlexCache. Offre funzionalità di caching remoto dei volumi che semplificano la distribuzione dei file, riducono la latenza WAN e riducono i costi della larghezza di banda WAN. FlexCache consente lo sviluppo distribuito di prodotti su più siti, oltre a un accesso accelerato ai set di dati aziendali da ubicazioni remote.

Infrastruttura a prova di futuro

ONTAP aiuta a soddisfare le esigenze di business esigenti e in continua evoluzione con le seguenti funzionalità:

- Scalabilità perfetta e operazioni senza interruzioni. ONTAP supporta l'aggiunta online di capacità ai controller esistenti e ai cluster scale-out. I clienti possono eseguire l'upgrade alle tecnologie più recenti, come NVMe e 32GB FC, senza costose migrazioni dei dati o interruzioni.
- Connessione al cloud. ONTAP è il software per la gestione dello storage più connesso al cloud, con opzioni per lo storage software-defined (ONTAP Select) e le istanze native del cloud (NetApp Cloud Volumes Service) in tutti i cloud pubblici.
- Integrazione con le applicazioni emergenti. ONTAP offre servizi dati di livello Enterprise per piattaforme e applicazioni di prossima generazione, come veicoli autonomi, città intelligenti e industria 4.0, utilizzando la stessa infrastruttura che supporta le applicazioni aziendali esistenti.

NetApp DataOps Toolkit

Il NetApp DataOps Toolkit è uno strumento basato su Python che semplifica la gestione degli spazi di lavoro di sviluppo/formazione e dei server di inferenza supportati dallo storage NetApp scale-out dalle performance elevate. DataOps Toolkit può funzionare come utility standalone ed è ancora più efficace negli ambienti Kubernetes sfruttando NetApp Astra Trident per automatizzare le operazioni di storage. Le funzionalità principali includono:

- Provisioning rapido di nuove aree di lavoro JupyterLab ad alta capacità supportate da storage NetApp scale-out dalle performance elevate.
- Provisioning rapido delle nuove istanze di NVIDIA Triton Inference Server supportate dallo storage NetApp di livello Enterprise.
- Clonazione quasi istantanea di spazi di lavoro JupyterLab ad alta capacità per consentire la sperimentazione o l'iterazione rapida.

- Istantanee quasi istantanee di spazi di lavoro JupyterLab ad alta capacità per backup e/o tracciabilità/baseline.
- Provisioning, cloning e Snapshot quasi istantanei di volumi di dati a elevata capacità e performance elevate.

NetApp Astra Trident

Astra Trident è un orchestrator di storage open-source completamente supportato per container e distribuzioni Kubernetes, incluso Anthos. Trident funziona con l'intero portfolio storage di NetApp, inclusi NetApp ONTAP, e supporta anche le connessioni NFS, NVMe/TCP e iSCSI. Trident accelera il workflow DevOps consentendo agli utenti finali di eseguire il provisioning e gestire lo storage dai sistemi storage NetApp senza richiedere l'intervento di un amministratore dello storage.

NetApp FlexPod con sistemi NVIDIA DGX - architettura della soluzione

Questa sezione si concentra sull'architettura per l'FlexPod di NetApp con sistemi NVIDIA DGX.

Pod NetApp ai con sistemi DGX H100

Questa architettura di riferimento sfrutta fabric separati per l'interconnessione del cluster di calcolo e l'accesso allo storage, con connettività InfiniBand (IB) a 400GB GB/s tra i nodi di calcolo. Il disegno qui sotto mostra la topologia della soluzione globale di NetApp FlexPod con sistemi DGX H100.

Topologia soluzione Alpod NetApp immagine::aipod_nv_a900topo.png[]

Configurazione di rete

In questa configurazione il cluster fabric di elaborazione utilizza una coppia di switch IB da QM9700 400GB GB/s, connessi insieme per high Availability. Ogni sistema DGX H100 è collegato agli switch utilizzando otto connessioni, con porte con numero pari collegate a uno switch e porte con numero dispari collegate all'altro switch.

Per l'accesso al sistema di storage, la gestione in banda e l'accesso client, viene utilizzata una coppia di switch Ethernet SN4600. Gli switch sono collegati con collegamenti inter-switch e configurati con più VLAN per isolare i vari tipi di traffico. Per implementazioni di dimensioni maggiori, la rete Ethernet può essere estesa a una configurazione leaf-spine aggiungendo ulteriori coppie di switch per gli switch spine e altre schede, se necessario.

Oltre all'interconnessione di elaborazione e alle reti Ethernet ad alta velocità, tutti i dispositivi fisici sono collegati anche a uno o più switch Ethernet SN2201 per la gestione fuori banda. Per ulteriori dettagli sulla connettività del sistema DGX H100, fare riferimento al "[Documentazione NVIDIA BasePOD](#)".

Configurazione del client per l'accesso allo storage

Ogni sistema DGX H100 viene fornito con due adattatori ConnectX-7 a due porte per il traffico di gestione e storage e, per questa soluzione, entrambe le porte su ogni scheda sono collegate allo stesso switch. Una porta di ogni scheda viene quindi configurata in un legame LACP MLAG con una porta collegata a ogni switch, e VLAN per la gestione in banda, l'accesso client e l'accesso di storage a livello utente sono ospitate su questo legame.

L'altra porta su ciascuna scheda viene utilizzata per la connettività ai sistemi di storage AFF A900 e può essere utilizzata in diverse configurazioni a seconda dei requisiti del carico di lavoro. Per le configurazioni che utilizzano NFS su RDMA per supportare NVIDIA Magnum io GPUDirect Storage, le porte sono configurate in

un legame attivo/passivo, in quanto RDMA non è supportato su nessun altro tipo di legame. Per le distribuzioni che non richiedono RDMA, le interfacce di storage possono anche essere configurate con il bonding LACP in modo da offrire elevata disponibilità e larghezza di banda aggiuntiva. Con o senza RDMA, i client possono montare il sistema di storage utilizzando NFS v4,1 pNFS e Session trunking per consentire l'accesso parallelo a tutti i nodi di storage nel cluster.

Configurazione del sistema storage

Ogni sistema storage AFF A900 è connesso tramite quattro porte 100 GbE da ciascun controller. Due porte da ciascun controller sono utilizzate per l'accesso ai dati dei carichi di lavoro dai sistemi DGX e due porte da ciascun controller sono configurate come un gruppo di interfacce LACP per supportare l'accesso dai server di piani di gestione per gli artefatti di gestione dei cluster e le home directory dell'utente. L'accesso ai dati dal sistema storage viene fornito tramite NFS, con una Storage Virtual Machine (SVM) dedicata all'accesso ai workload e una SVM separata dedicata agli utilizzi della gestione del cluster.

Il carico di lavoro SVM è configurato con un totale di otto interfacce logiche (LIF), con due LIF per ciascuna porta fisica. Questa configurazione offre la massima larghezza di banda oltre a mezzi per eseguire il failover di ciascuna LIF in un'altra porta dello stesso controller, in modo che entrambi i controller rimangano attivi in caso di guasto alla rete. Questa configurazione supporta anche NFS su RDMA per abilitare l'accesso allo storage GPUDirect. Viene effettuato il provisioning della capacità dello storage sotto forma di un singolo volume FlexGroup di grandi dimensioni che comprende tutti gli storage controller del cluster, con 16 volumi costituenti su ciascun controller. Questo FlexGroup è accessibile da una qualsiasi LIF nella SVM e utilizzando NFSv4,1 con pNFS e trunking di sessione, i client stabiliscono connessioni a ogni LIF nella SVM, permettendo l'accesso in parallelo ai dati locali in ogni nodo storage per performance notevolmente migliorate. Anche la SVM del carico di lavoro e ciascuna LIF dati sono configurate per l'accesso al protocollo RDMA. Per ulteriori informazioni sulla configurazione RDMA per ONTAP, fare riferimento alla "[Documentazione ONTAP](#)".

La SVM di gestione richiede un singolo LIF, ospitato nei gruppi di interfacce a 2 porte configurati su ciascun controller. Nella SVM di gestione viene eseguito il provisioning di altri volumi FlexGroup per alloggiare elementi legati alla gestione del cluster, come le immagini dei nodi del cluster, i dati storici del monitoraggio del sistema e le home directory degli utenti finali. Il disegno qui sotto mostra la configurazione logica del sistema di archiviazione.

Configurazione logica cluster di archiviazione NetApp A900 immagine::aipod_nv_A900logical.png[]

Server piani di gestione

Questa architettura di riferimento include anche cinque server basati su CPU per gli utilizzi dei piani di gestione. Due di questi sistemi sono utilizzati come nodi principali per NVIDIA base Command Manager per l'implementazione e la gestione del cluster. Gli altri tre sistemi sono utilizzati per fornire servizi cluster aggiuntivi come nodi master Kubernetes o nodi di login per implementazioni che utilizzano Slurm per la pianificazione dei processi. Le implementazioni che utilizzano Kubernetes possono sfruttare il driver NetApp Astra Trident CSI per fornire provisioning automatizzato e servizi dati con storage persistente per carichi di lavoro di gestione e ai sul sistema storage AFF A900.

Ciascun server è fisicamente collegato sia agli switch IB che agli switch Ethernet per consentire l'implementazione e la gestione del cluster e configurato con mount NFS al sistema storage tramite la SVM di gestione per lo storage degli artefatti di gestione dei cluster, come descritto in precedenza.

NetApp FlexPod con sistemi NVIDIA DGX - Guida alla convalida della soluzione e al dimensionamento

Questa sezione è incentrata sulla convalida della soluzione e sulle linee guida sul dimensionamento per l'FlexPod di NetApp con sistemi NVIDIA DGX.

Convalida della soluzione

La configurazione dello storage in questa soluzione è stata convalidata utilizzando una serie di carichi di lavoro sintetici utilizzando il tool open-source FIO. Questi test includono modelli di i/o in lettura e scrittura ideati per simulare il carico di lavoro dello storage generato dai sistemi DGX che eseguono lavori di training di deep learning. La configurazione dello storage è stata convalidata utilizzando un cluster di server CPU a 2 socket che eseguono contemporaneamente i carichi di lavoro FIO per simulare un cluster di sistemi DGX. Ciascun client è stato configurato con la stessa configurazione di rete descritta in precedenza, con l'aggiunta dei seguenti dettagli.

Per questa convalida sono state utilizzate le seguenti opzioni di montaggio:

vers=4,1	Abilita pNFS per l'accesso parallelo a più nodi storage
proto=rdma	Consente di impostare il protocollo di trasferimento su RDMA anziché sul TCP predefinito
porta=20049	Specificare la porta corretta per il servizio NFS RDMA
max_connect=16	Abilita il trunking di sessione NFS per aggregare la larghezza di banda delle porte di storage
write=desideroso	migliora le prestazioni di scrittura delle scritture bufferizzate
rsize=262144,wsiz=262144	Imposta le dimensioni di trasferimento i/o su 256k

Inoltre, i client sono stati configurati con un valore NFS max_session_slot pari a 1024. Quando la soluzione è stata collaudata con NFS su RDMA, le porte di storage network sono state configurate con un legame attivo/passivo. Per questa convalida sono stati utilizzati i seguenti parametri di legame:

mode=backup attivo	imposta il legame in modalità attiva/passiva
primary (primario) = <interface name>	le interfacce primarie per tutti i client sono state distribuite tra gli switch
intervallo-monitor-mii=100	specifica l'intervallo di monitoraggio di 100ms
fail-over-mac-policy=attivo	Specifica che l'indirizzo MAC del collegamento attivo è il MAC del collegamento. Ciò è necessario per il corretto funzionamento di RDMA sull'interfaccia collegata.

Il sistema storage è stato configurato come descritto con due coppie ha A900 (4 controller) con due shelf di dischi NS224 da 24 1,9TB dischi NVMe collegati a ciascuna coppia ha. Come indicato nella sezione "architettura", la capacità dello storage di tutti i controller è stata combinata con un volume FlexGroup, distribuendo i dati di tutti i client in tutti i controller del cluster.

Guida al dimensionamento del sistema storage

NetApp ha completato con successo la certificazione DGX BasePOD, e le due coppie ha A900 testate possono supportare facilmente un cluster di otto sistemi DGX H100. Per implementazioni più estese con requisiti di performance dello storage più elevati, è possibile aggiungere sistemi AFF al cluster NetApp ONTAP fino a 12 coppie ha (24 nodi) in un singolo cluster. Utilizzando la tecnologia FlexGroup descritta in questa soluzione, un cluster a 24 nodi può offrire oltre 40 PB e un throughput fino a 300 Gbps in un singolo namespace. Altri sistemi storage NetApp come AFF A400, A250 e C800 offrono performance e/o capacità superiori per implementazioni più piccole a prezzi inferiori. Poiché ONTAP 9 supporta cluster di modelli misti, i clienti possono partire con un impatto iniziale minore e aggiungere al cluster più sistemi storage o più grandi man mano che crescono i requisiti di capacità e performance. La tabella seguente mostra una stima

approssimativa del numero di GPU A100 e H100 supportate su ciascun modello AFF.

Guida al dimensionamento del sistema storage NetApp

		Throughput ²	Raw capacity (typical / max)	Connectivity	# NVIDIA A100 GPUs supported ³	# NVIDIA H100 GPUs supported ⁴
NetApp® AFF A900	1 HA pair ¹	28GB/s	182TB / 14.7PB	100 GbE	1 - 64	1-32
	12 HA pairs	336GB/s	2.1PB / 176.4PB		768	384
AFF A800	1 HA pair	25GB/s	368TB / 3.6PB	100 GbE	1 - 64	1-32
	12 HA pairs	300GB/s	4.4PB / 43.2PB		768	384
AFF C800	1 HA pair	21GB/s	368TB / 3.6PB	100 GbE	1-48	1-24
	12 HA pairs	252GB/s	4.4PB / 43.2PB		576	288
AFF A400	1 HA pair	11GB/s	182TB / 14.7PB	40/100 GbE	1 - 32	1-16
	12 HA pairs	132GB/s	2.1PB / 176.4PB		384	192
AFF C400	1 HA pair	8GB/s	182TB / 14.7PB	40/100 GbE	1 - 16	1-8
	12 HA pairs	128GB/s	2.1PB / 176.4PB		192	96
AFF A250	1 HA pair	7.4GB/s	91.2TB / 4.4PB	25 GbE 40/100GbE	1 - 16	1-8
	4 HA pairs	29.6GB/s	364.8TB / 17.6PB		64	32
AFF C250	1 HA pair	5 GB/s	91.2TB / 4.4PB	25 GbE 40/100GbE	1-8	1-4
	4 HA pairs	20 GB/s	364.8TB / 17.6PB		32	8

1 – 1 AFF = 1 HA pair = 2 Nodes. 12 HA pairs = 24 nodes
2 – 100% sequential read

3 – Based on workload testing in NVA-1153
4 – Based on BasePOD validation test results

NetApp FlexPod con sistemi NVIDIA DGX - conclusioni e informazioni aggiuntive

Questa sezione include riferimenti per informazioni aggiuntive per NetApp APod con sistemi NVIDIA DGX.

Conclusione

L'architettura DGX BasePOD è una piattaforma di deep learning di nuova generazione che richiede capacità di storage e di gestione dei dati altrettanto avanzate. Combinando DGX BasePOD con i sistemi NetApp AFF, l'architettura dei sistemi NetApp AIPod con DGX può essere implementata su quasi ogni scala fino a 48 sistemi DGX H100 su un cluster AFF A900 a 24 nodi. Combinato con l'integrazione superiore del cloud e le capacità software-defined di NetApp ONTAP, AFF offre una gamma completa di pipeline di dati che spazia dall'edge al core fino al cloud per progetti DL di successo.

Ulteriori informazioni

Per ulteriori informazioni sulle informazioni descritte in questo documento, fare riferimento ai seguenti documenti e/o siti Web:

- Software per la gestione dei dati NetApp ONTAP: Libreria di informazioni ONTAP

["https://docs.netapp.com/us-en/ontap-family/"](https://docs.netapp.com/us-en/ontap-family/)

- Sistemi storage NetApp AFF A900 -

["https://www.netapp.com/data-storage/aff-a-series/aff-a900/"](https://www.netapp.com/data-storage/aff-a-series/aff-a900/)

- Informazioni RDMA NetApp ONTAP-
["https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html"](https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html)
- NetApp DataOps Toolkit
["https://github.com/NetApp/netapp-dataops-toolkit"](https://github.com/NetApp/netapp-dataops-toolkit)
- NetApp Astra Trident
["Panoramica"](#)
- Blog sullo storage GPUDirect di NetApp -
["https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/"](https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/)
- NVIDIA DGX BasePOD
["https://www.nvidia.com/en-us/data-center/dgx-basepod/"](https://www.nvidia.com/en-us/data-center/dgx-basepod/)
- SISTEMI NVIDIA DGX H100
["https://www.nvidia.com/en-us/data-center/dgx-h100/"](https://www.nvidia.com/en-us/data-center/dgx-h100/)
- Rete NVIDIA
["https://www.nvidia.com/en-us/networking/"](https://www.nvidia.com/en-us/networking/)
- NVIDIA Magnum io GPUDirect Storage
["https://docs.nvidia.com/gpudirect-storage"](https://docs.nvidia.com/gpudirect-storage)
- Comando base NVIDIA
["https://www.nvidia.com/en-us/data-center/base-command/"](https://www.nvidia.com/en-us/data-center/base-command/)
- NVIDIA base Command Manager
["https://www.nvidia.com/en-us/data-center/base-command/manager"](https://www.nvidia.com/en-us/data-center/base-command/manager)
- NVIDIA ai Enterprise
["https://www.nvidia.com/en-us/data-center/products/ai-enterprise/"](https://www.nvidia.com/en-us/data-center/products/ai-enterprise/)

Ringraziamenti

Questo documento è il lavoro dei team di progettazione delle soluzioni NetApp e ONTAP: David Arnette, Olga Kornievskaia, Dustin Fischer, Srikanth Kaligotla, Mohit Kumar e Rajeev Badrinath. Gli autori vogliono anche ringraziare NVIDIA e il team tecnico NVIDIA DGX BasePOD per il loro costante supporto.

NVA-1151-DESIGN: Guida alla progettazione dei sistemi NetApp ONTAP ai con NVIDIA DGX A100

David Arnette e Sung-Han Lin, NetApp

NVA-1151-DESIGN descrive un'architettura verificata NetApp per l'apprendimento automatico e i carichi di lavoro di intelligenza artificiale che utilizzano i sistemi storage NetApp AFF A800, i sistemi NVIDIA DGX A100 e gli switch di rete NVIDIA Mellanox. Include anche i risultati dei test di benchmark per l'architettura implementata.

["NVA-1151-DESIGN: Guida alla progettazione dei sistemi NetApp ONTAP ai con NVIDIA DGX A100"](#)

NVA-1151-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100

David Arnette, NetApp

NVA-1151-DEPLOY include istruzioni per l'implementazione del sistema storage per un'architettura verificata NetApp (NVA) per i carichi di lavoro di apprendimento automatico (ML) e intelligenza artificiale (ai) utilizzando i sistemi storage NetApp AFF A800, i sistemi NVIDIA DGX A100 e gli switch di rete NVIDIA Mellanox. Include inoltre istruzioni per l'esecuzione dei test di benchmark di convalida al termine dell'implementazione.

["NVA-1151-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100"](#)

NVA-1153-DESIGN: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum

David Arnette e Sung-Han Lin, NetApp

NVA-1153-DESIGN descrive un'architettura verificata NetApp per i carichi di lavoro di apprendimento automatico (ML) e intelligenza artificiale (ai) che utilizzano i sistemi storage NetApp AFF A800, i sistemi NVIDIA DGX A100 e gli switch Ethernet NVIDIA Mellanox Spectrum SN3700V 200 GB. Questo design è dotato di RDMA over Converged Ethernet (RoCE) per il fabric di interconnessione del cluster di calcolo per offrire ai clienti un'architettura completamente basata su ethernet per carichi di lavoro dalle performance elevate. Questo documento include anche i risultati dei test di benchmark per l'architettura implementata.

["NVA-1153-DESIGN: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum"](#)

NVA-1153-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum

David Arnette, NetApp

NVA-1153-DEPLOY include istruzioni di implementazione del sistema storage per un'architettura verificata NetApp per i carichi di lavoro di apprendimento automatico (ML) e intelligenza artificiale (ai) che utilizzano i sistemi storage NetApp AFF A800, i sistemi NVIDIA DGX A100 e gli switch Ethernet NVIDIA Mellanox Spectrum SN3700V da 200

GB. Include inoltre istruzioni per l'esecuzione dei test di benchmark di convalida al termine dell'implementazione.

"NVA-1153-DEPLOY: NetApp ONTAP ai con sistemi NVIDIA DGX A100 e switch Ethernet Mellanox Spectrum"

Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALIZZABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.