



## **NetApp AI Pod mit NVIDIA DGX-Systemen**

NetApp artificial intelligence solutions

NetApp

February 12, 2026

# Inhalt

NetApp AI Pod mit NVIDIA DGX-Systemen .....	1
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Einführung .....	1
Zusammenfassung .....	1
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Hardwarekomponenten .....	2
NetApp AFF -Speichersysteme .....	2
NVIDIA DGX BasePOD .....	3
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Softwarekomponenten .....	5
NVIDIA -Software .....	6
NetApp Software .....	7
NVA-1173 NetApp AI Pod mit NVIDIA DGX H100-Systemen – Lösungsarchitektur .....	9
NetApp AI Pod mit DGX-Systemen .....	9
Netzwerkdesign .....	10
Speicherzugriffsübersicht für DGX H100-Systeme .....	11
Speichersystemdesign .....	11
Management-Plane-Server .....	12
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Bereitstellungsdetails .....	12
Speichernetzwerkkonfiguration .....	14
Speichersystemkonfiguration .....	16
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Leitfaden zur Lösungsvalidierung und Größenbestimmung .....	20
Lösungsvalidierung .....	20
Leitfaden zur Größenbestimmung von Speichersystemen .....	21
NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Fazit und weitere Informationen .....	21
Abschluss .....	22
Weitere Informationen .....	22
Danksagung .....	23

# NetApp AIPOD mit NVIDIA DGX-Systemen

## NVA-1173 NetApp AIPOD mit NVIDIA DGX-Systemen – Einführung

# POWERED BY



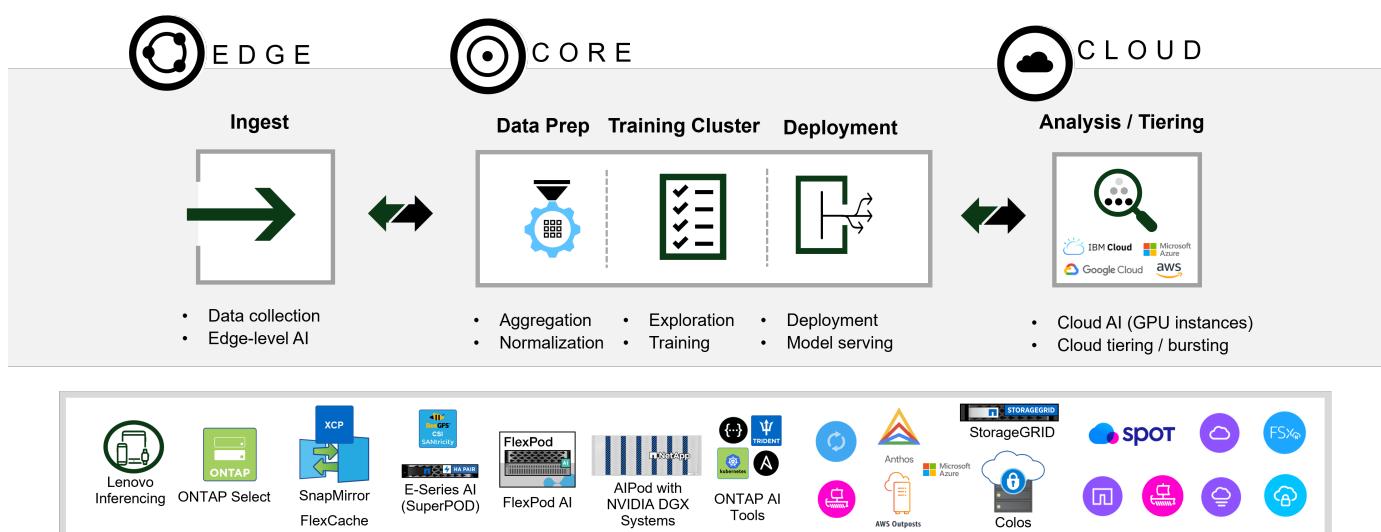
# nVIDIA.

NetApp Solution Engineering

### Zusammenfassung

Der NetApp® AIPOD mit NVIDIA DGX®-Systemen und NetApp Cloud-verbundenen Speichersystemen vereinfacht die Infrastrukturbereitstellung für Workloads im Bereich maschinelles Lernen (ML) und künstliche Intelligenz (KI), indem er Designkomplexität und Rätselraten eliminiert. AIPOD mit NVIDIA DGX-Systemen baut auf dem NVIDIA DGX BasePOD™-Design auf, um eine außergewöhnliche Rechenleistung für Workloads der nächsten Generation zu liefern, und fügt NetApp AFF Speichersysteme hinzu, die es Kunden ermöglichen, klein anzufangen und unterbrechungsfrei zu wachsen, während sie gleichzeitig Daten vom Rand über den Kern bis zur Cloud und zurück intelligent verwalten. NetApp AIPOD ist Teil des größeren Portfolios an NetApp KI-Lösungen, wie in der folgenden Abbildung dargestellt.

### NetApp AI-Lösungsportfolio



Dieses Dokument beschreibt die Hauptkomponenten der AIPOD -Referenzarchitektur, Informationen zur Systemkonnektivität und -konfiguration, Ergebnisse der Validierungstests und Leitlinien zur Größenbestimmung der Lösung. Dieses Dokument richtet sich an Lösungingenieure von NetApp und

Partnern sowie an strategische Entscheidungsträger bei Kunden, die an der Bereitstellung einer Hochleistungsinfrastruktur für ML/DL- und Analyse-Workloads interessiert sind.

## NVA-1173 NetApp AIPod mit NVIDIA DGX-Systemen – Hardwarekomponenten

Dieser Abschnitt konzentriert sich auf die Hardwarekomponenten für den NetApp AIPod mit NVIDIA DGX-Systemen.

### NetApp AFF -Speichersysteme

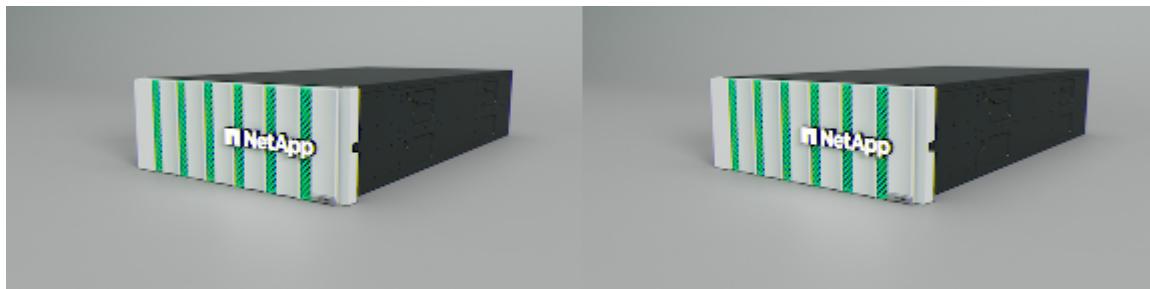
Mit den hochmodernen AFF -Speichersystemen von NetApp können IT-Abteilungen die Speicheranforderungen von Unternehmen mit branchenführender Leistung, überragender Flexibilität, Cloud-Integration und erstklassigem Datenmanagement erfüllen. AFF -Systeme wurden speziell für Flash entwickelt und helfen bei der Beschleunigung, Verwaltung und dem Schutz geschäftskritischer Daten.

#### AFF A90 Speichersysteme

Der NetApp AFF A90 mit der Datenverwaltungssoftware NetApp ONTAP bietet integrierten Datenschutz, optionale Anti-Ransomware-Funktionen sowie die hohe Leistung und Ausfallsicherheit, die zur Unterstützung der kritischsten Geschäfts-Workloads erforderlich sind. Es verhindert Störungen unternehmenskritischer Vorgänge, minimiert die Leistungsoptimierung und schützt Ihre Daten vor Ransomware-Angriffen. Es bietet:

- Branchenführende Leistung
- Kompromisslose Datensicherheit
- Vereinfachte, unterbrechungsfreie Upgrades

*NetApp AFF A90 Speichersystem*



#### Branchenführende Leistung

Die AFF A90 bewältigt problemlos Workloads der nächsten Generation wie Deep Learning, KI und Hochgeschwindigkeitsanalysen sowie traditionelle Unternehmensdatenbanken wie Oracle, SAP HANA, Microsoft SQL Server und virtualisierte Anwendungen. Es sorgt dafür, dass geschäftskritische Anwendungen mit bis zu 2,4 Millionen IOPS pro HA-Paar und einer Latenz von nur 100 µs mit Höchstgeschwindigkeit ausgeführt werden – und steigert die Leistung im Vergleich zu früheren NetApp Modellen um bis zu 50 %. Mit NFS über RDMA, pNFS und Session Trunking können Kunden mithilfe der vorhandenen Netzwerkinfrastruktur ihres Rechenzentrums die für Anwendungen der nächsten Generation erforderliche hohe Netzwerkleistung

erreichen. Kunden können außerdem mit einheitlicher Multiprotokollunterstützung für SAN, NAS und Objektspeicher skalieren und wachsen und mit einheitlicher und einzelner ONTAP Datenverwaltungssoftware für Daten vor Ort oder in der Cloud maximale Flexibilität erzielen. Darüber hinaus kann die Systemintegrität mit KI-basierten prädiktiven Analysen von Active IQ und Cloud Insights optimiert werden.

#### Kompromisslose Datensicherheit

AFF A90 -Systeme enthalten eine vollständige Suite integrierter und anwendungskonsistenter Datenschutzsoftware von NetApp . Es bietet integrierten Datenschutz und hochmoderne Anti-Ransomware-Lösungen zur Vorbeugung und Wiederherstellung nach einem Angriff. Das Schreiben schädlicher Dateien auf die Festplatte kann blockiert werden und Speicheranomalien können einfach überwacht werden, um Erkenntnisse zu gewinnen.

#### Vereinfachte, unterbrechungsfreie Upgrades

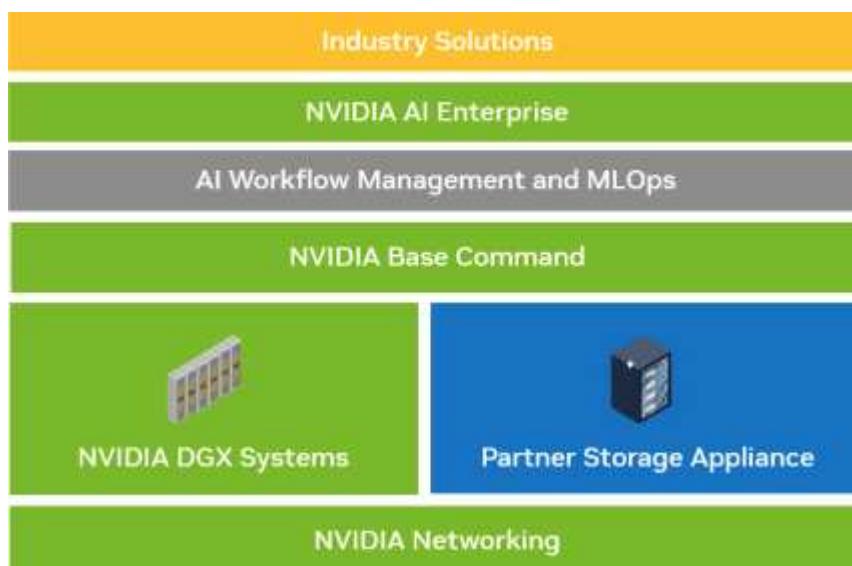
Der AFF A90 ist als unterbrechungsfreies In-Chassis-Upgrade für bestehende A800-Kunden verfügbar. NetApp vereinfacht die Aktualisierung und Beseitigung von Störungen unternehmenskritischer Vorgänge durch unsere erweiterten Funktionen für Zuverlässigkeit, Verfügbarkeit, Wartungsfreundlichkeit und Verwaltbarkeit (RASM). Darüber hinaus steigert NetApp die Betriebseffizienz weiter und vereinfacht die täglichen Aktivitäten der IT-Teams, da die ONTAP -Software automatisch Firmware-Updates für alle Systemkomponenten anwendet.

Für die größten Bereitstellungen bieten AFF A1K -Systeme die höchsten Leistungs- und Kapazitätsoptionen, während andere NetApp Speichersysteme wie AFF A70 und AFF C800 Optionen für kleinere Bereitstellungen zu niedrigeren Kosten bieten.

## NVIDIA DGX BasePOD

NVIDIA DGX BasePOD ist eine integrierte Lösung, die aus NVIDIA -Hardware- und Softwarekomponenten, MLOps-Lösungen und Speicher von Drittanbietern besteht. Durch die Nutzung bewährter Methoden des Scale-Out-Systemdesigns mit NVIDIA Produkten und validierten Partnerlösungen können Kunden eine effiziente und verwaltbare Plattform für die KI-Entwicklung implementieren. Abbildung 1 zeigt die verschiedenen Komponenten von NVIDIA DGX BasePOD.

#### NVIDIA DGX BasePOD-Lösung



## NVIDIA DGX H100-Systeme

Das NVIDIA DGX H100™-System ist das KI-Kraftpaket, das durch die bahnbrechende Leistung der NVIDIA H100 Tensor Core GPU beschleunigt wird.

### NVIDIA DGX H100-System

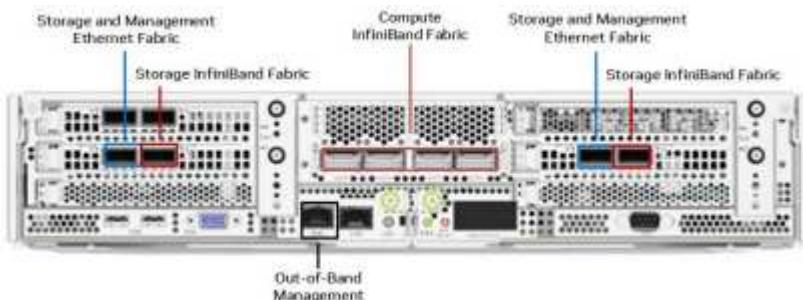


Die wichtigsten Spezifikationen des DGX H100-Systems sind:

- Acht NVIDIA H100-GPUs.
- 80 GB GPU-Speicher pro GPU, insgesamt 640 GB.
- Vier NVIDIA NVSwitch-Chips.
- Duale 56-Core Intel Xeon Platinum 8480-Prozessoren mit PCIe 5.0-Unterstützung.
- 2 TB DDR5-Systemspeicher.
- Vier OSFP-Ports für acht NVIDIA ConnectX-7-Adapter (InfiniBand/Ethernet) mit einem Port und zwei NVIDIA ConnectX-7-Adapter (InfiniBand/Ethernet) mit zwei Ports.
- Zwei 1,92 TB M.2 NVMe-Laufwerke für DGX OS, acht 3,84 TB U.2 NVMe-Laufwerke für Speicher/Cache.
- 10,2 kW maximale Leistung.

Die hinteren Anschlüsse des DGX H100 CPU-Fachs sind unten dargestellt. Vier der OSFP-Ports bedienen acht ConnectX-7-Adapter für das InfiniBand-Compute-Fabric. Jedes Paar ConnectX-7-Adapter mit zwei Anschlüssen bietet parallele Pfade zu den Speicher- und Verwaltungsstrukturen. Der Out-of-Band-Port wird für den BMC Zugriff verwendet.

### Rückseite des NVIDIA DGX H100



## NVIDIA Netzwerk

### NVIDIA Quantum-2 QM9700-Switch

### NVIDIA Quantum-2 QM9700 InfiniBand-Schalter



NVIDIA Quantum-2 QM9700-Switches mit 400 Gb/s InfiniBand-Konnektivität versorgen das Compute Fabric in NVIDIA Quantum-2 InfiniBand BasePOD-Konfigurationen mit Strom. Für das InfiniBand-Compute-Fabric werden ConnectX-7-Einzelportadapter verwendet. Jedes NVIDIA DGX-System verfügt über zwei Verbindungen zu jedem QM9700-Switch und bietet so mehrere Pfade mit hoher Bandbreite und geringer Latenz zwischen den Systemen.

### **NVIDIA Spectrum-3 SN4600-Switch**

#### *NVIDIA Spectrum-3 SN4600-Schalter*



NVIDIA Spectrum™-3 SN4600-Switches bieten insgesamt 128 Ports (64 pro Switch), um redundante Konnektivität für die In-Band-Verwaltung des DGX BasePOD bereitzustellen. Der NVIDIA SN4600-Switch kann Geschwindigkeiten zwischen 1 GbE und 200 GbE bereitstellen. Für über Ethernet angeschlossene Speichergeräte werden ebenfalls die NVIDIA SN4600-Switches verwendet. Die Ports der NVIDIA DGX Dual-Port ConnectX-7-Adapter werden sowohl für die In-Band-Verwaltung als auch für die Speicherkonnektivität verwendet.

### **NVIDIA Spectrum SN2201-Switch**

#### *NVIDIA Spectrum SN2201-Schalter*



NVIDIA Spectrum SN2201-Switches bieten 48 Ports, um Konnektivität für Out-of-Band-Management bereitzustellen. Out-of-Band-Management bietet konsolidierte Verwaltungskonnektivität für alle Komponenten in DGX BasePOD.

### **NVIDIA ConnectX-7-Adapter**

#### *NVIDIA ConnectX-7-Adapter*



Der NVIDIA ConnectX-7-Adapter kann einen Durchsatz von 25/50/100/200/400 G bereitstellen. NVIDIA DGX-Systeme verwenden sowohl die ConnectX-7-Adapter mit einem als auch mit zwei Anschlüssen, um Flexibilität bei DGX BasePOD-Bereitstellungen mit 400 Gb/s InfiniBand und Ethernet zu bieten.

## **NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Softwarekomponenten**

Dieser Abschnitt konzentriert sich auf die Softwarekomponenten des NetApp AI Pod mit

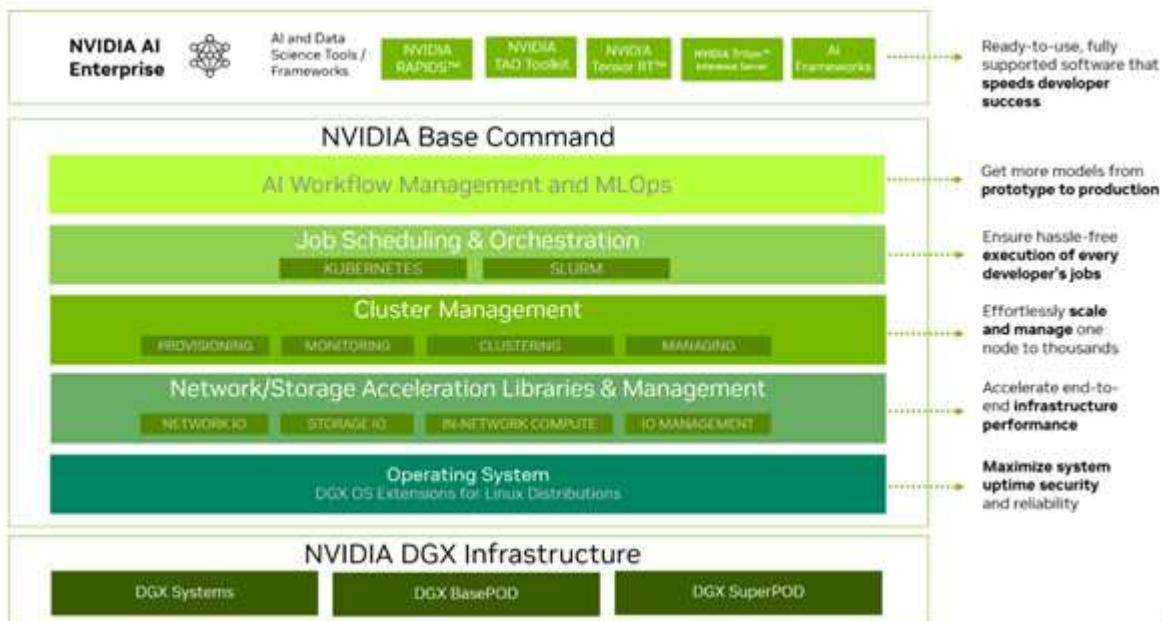
# NVIDIA DGX-Systemen.

## NVIDIA -Software

### NVIDIA Basisbefehl

NVIDIA Base Command™ treibt jeden DGX BasePOD an und ermöglicht es Unternehmen, das Beste der NVIDIA -Softwareinnovation zu nutzen. Unternehmen können das volle Potenzial ihrer Investition mit einer bewährten Plattform ausschöpfen, die Orchestrierung und Clusterverwaltung auf Unternehmensniveau, Bibliotheken zur Beschleunigung der Rechen-, Speicher- und Netzwerkinfrastruktur sowie ein für KI-Workloads optimiertes Betriebssystem (OS) umfasst.

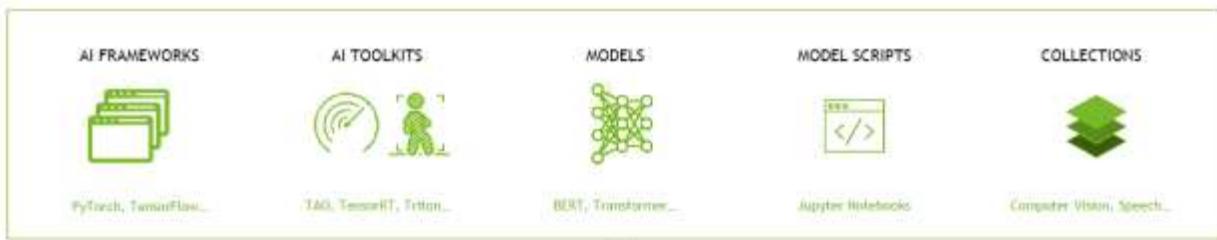
### NVIDIA BaseCommand-Lösung



### NVIDIA GPU Cloud (NGC)

NVIDIA NGC bietet Software, die den Anforderungen von Datenwissenschaftlern, Entwicklern und Forschern mit unterschiedlichem KI-Fachwissen gerecht wird. Auf NGC gehostete Software wird auf einen aggregierten Satz gängiger Schwachstellen und Gefährdungen (CVEs), Kryptografie und privater Schlüssel geprüft. Es ist für die Skalierung auf mehrere GPUs und in vielen Fällen auf mehrere Knoten getestet und konzipiert, um sicherzustellen, dass Benutzer ihre Investition in DGX-Systeme maximieren.

### NVIDIA GPU Cloud



## NVIDIA AI Enterprise

NVIDIA AI Enterprise ist die End-to-End-Softwareplattform, die generative KI für jedes Unternehmen erreichbar macht und die schnellste und effizienteste Laufzeit für generative KI-Basismodelle bietet, die für die Ausführung auf der NVIDIA DGX-Plattform optimiert sind. Mit produktionsreifer Sicherheit, Stabilität und Verwaltbarkeit rationalisiert es die Entwicklung generativer KI-Lösungen. NVIDIA AI Enterprise ist in DGX BasePOD enthalten, damit Unternehmensentwickler auf vortrainierte Modelle, optimierte Frameworks, Microservices, beschleunigte Bibliotheken und Unternehmenssupport zugreifen können.

## NetApp Software

### NetApp ONTAP

ONTAP 9, die neueste Generation der Speicherverwaltungssoftware von NetApp, ermöglicht Unternehmen die Modernisierung ihrer Infrastruktur und den Übergang zu einem Cloud-fähigen Rechenzentrum. Durch die Nutzung branchenführender Datenverwaltungsfunktionen ermöglicht ONTAP die Verwaltung und den Schutz von Daten mit einem einzigen Satz von Tools, unabhängig davon, wo sich diese Daten befinden. Sie können Daten auch frei dorthin verschieben, wo sie benötigt werden: an den Rand, in den Kern oder in die Cloud. ONTAP 9 umfasst zahlreiche Funktionen, die die Datenverwaltung vereinfachen, kritische Daten beschleunigen und schützen und Infrastrukturfunktionen der nächsten Generation in Hybrid-Cloud-Architekturen ermöglichen.

### Beschleunigen und schützen Sie Daten

ONTAP bietet ein Höchstmaß an Leistung und Datenschutz und erweitert diese Funktionen auf folgende Weise:

- Leistung und geringere Latenz. ONTAP bietet den höchstmöglichen Durchsatz bei der geringstmöglichen Latenz, einschließlich Unterstützung für NVIDIA GPUDirect Storage (GDS) mit NFS über RDMA, parallelem NFS (pNFS) und NFS-Sitzungs-Trunking.
- Datenschutz. ONTAP bietet integrierte Datenschutzfunktionen und die branchenweit stärkste Anti-Ransomware-Garantie mit gemeinsamer Verwaltung über alle Plattformen hinweg.
- NetApp Volume Encryption (NVE). ONTAP bietet native Verschlüsselung auf Volume-Ebene mit Unterstützung für integriertes und externes Schlüsselmanagement.
- Mandantenfähigkeit der Speicherung und Multifaktor-Authentifizierung. ONTAP ermöglicht die gemeinsame Nutzung von Infrastrukturressourcen mit höchster Sicherheit.

## Vereinfachen Sie die Datenverwaltung

Das Datenmanagement ist für den IT-Betrieb in Unternehmen und für Datenwissenschaftler von entscheidender Bedeutung, damit für KI-Anwendungen und das Training von KI/ML-Datensätzen die richtigen Ressourcen verwendet werden. Die folgenden zusätzlichen Informationen zu NetApp -Technologien fallen nicht in den Geltungsbereich dieser Validierung, können jedoch je nach Bereitstellung relevant sein.

Die ONTAP Datenmanagementsoftware umfasst die folgenden Funktionen zur Optimierung und Vereinfachung von Abläufen und zur Senkung Ihrer Gesamtbetriebskosten:

- Snapshots und Klone ermöglichen Zusammenarbeit, paralleles Experimentieren und verbesserte Datenverwaltung für ML/DL-Workflows.
- SnapMirror ermöglicht eine nahtlose Datenbewegung in Hybrid-Cloud- und Multi-Site-Umgebungen und stellt Daten bereit, wo und wann sie benötigt werden.
- Inline-Datenkomprimierung und erweiterte Deduplizierung. Durch die Datenkomprimierung wird der verschwendete Speicherplatz in Speicherblöcken reduziert und durch die Deduplizierung wird die effektive Kapazität erheblich erhöht. Dies gilt für lokal gespeicherte Daten und für in der Cloud gespeicherte Daten.
- Minimale, maximale und adaptive Dienstqualität (AQoS). Durch granulare Quality of Service (QoS)-Kontrollen wird die Aufrechterhaltung des Leistungsniveaus kritischer Anwendungen in Umgebungen mit hoher gemeinsamer Nutzung unterstützt.
- NetApp FlexGroups ermöglichen die Verteilung von Daten auf alle Knoten im Speichercluster und bieten enorme Kapazität und höhere Leistung für extrem große Datensätze.
- NetApp FabricPool. Bietet automatisches Tiering von Cold Data für öffentliche und private Cloud-Speicheroptionen, einschließlich Amazon Web Services (AWS), Azure und der NetApp StorageGRID-Speicherlösung. Weitere Informationen zu FabricPool finden Sie unter "[TR-4598: Best Practices für FabricPool](#)".
- NetApp FlexCache. Bietet Remote-Volume-Caching-Funktionen, die die Dateiverteilung vereinfachen, die WAN-Latenz reduzieren und die WAN-Bandbreitenkosten senken. FlexCache ermöglicht die verteilte Produktentwicklung über mehrere Standorte hinweg sowie einen beschleunigten Zugriff auf Unternehmensdatensätze von entfernten Standorten aus.

## Zukunftssichere Infrastruktur

ONTAP unterstützt Sie mit den folgenden Funktionen bei der Erfüllung anspruchsvoller und sich ständig ändernder Geschäftsanforderungen:

- Nahtlose Skalierung und unterbrechungsfreier Betrieb. ONTAP unterstützt die Online-Kapazitätserweiterung bestehender Controller und Scale-Out-Cluster. Kunden können ohne kostspielige Datenmigrationen oder Ausfälle auf die neuesten Technologien wie NVMe und 32 GB FC upgraden.
- Cloud-Verbindung. ONTAP ist die Speicherverwaltungssoftware mit der stärksten Cloud-Anbindung und bietet Optionen für softwaredefinierten Speicher (ONTAP Select) und Cloud-native Instanzen (Google Cloud NetApp Volumes) in allen öffentlichen Clouds.
- Integration mit neuen Anwendungen. ONTAP bietet Datendienste der Enterprise-Klasse für Plattformen und Anwendungen der nächsten Generation, wie etwa autonome Fahrzeuge, Smart Cities und Industrie 4.0, und nutzt dabei dieselbe Infrastruktur, die auch bestehende Unternehmens-Apps unterstützt.

## NetApp DataOps Toolkit

Das NetApp DataOps Toolkit ist ein Python-basiertes Tool, das die Verwaltung von Entwicklungs-/Schulungsarbeitsbereichen und Inferenzservern vereinfacht, die durch leistungsstarken, skalierbaren NetApp -Speicher unterstützt werden. Das DataOps Toolkit kann als eigenständiges Dienstprogramm betrieben

werden und ist in Kubernetes-Umgebungen, in denen NetApp Trident zur Automatisierung von Speichervorgängen genutzt wird, noch effektiver. Zu den wichtigsten Funktionen gehören:

- Stellen Sie schnell neue JupyterLab-Arbeitsbereiche mit hoher Kapazität bereit, die durch leistungsstarken, skalierbaren NetApp Speicher unterstützt werden.
- Stellen Sie schnell neue NVIDIA Triton Inference Server-Instanzen bereit, die durch NetApp -Speicher der Enterprise-Klasse unterstützt werden.
- Nahezu sofortiges Klonen von JupyterLab-Arbeitsbereichen mit hoher Kapazität, um Experimente oder schnelle Iterationen zu ermöglichen.
- Nahezu sofortige Snapshots von JupyterLab-Arbeitsbereichen mit hoher Kapazität für Backups und/oder Rückverfolgbarkeit/Baselining.
- Nahezu sofortige Bereitstellung, Klonen und Snapshots von Datenvolumes mit hoher Kapazität und hoher Leistung.

## NetApp Trident

Trident ist ein vollständig unterstützter Open-Source-Speicherorchestrator für Container und Kubernetes-Distributionen, einschließlich Anthos. Trident funktioniert mit dem gesamten NetApp -Speicherportfolio, einschließlich NetApp ONTAP, und unterstützt auch NFS-, NVMe/TCP- und iSCSI-Verbindungen. Trident beschleunigt den DevOps-Workflow, indem es Endbenutzern ermöglicht, Speicher von ihren NetApp -Speichersystemen bereitzustellen und zu verwalten, ohne dass ein Speicheradministrator eingreifen muss.

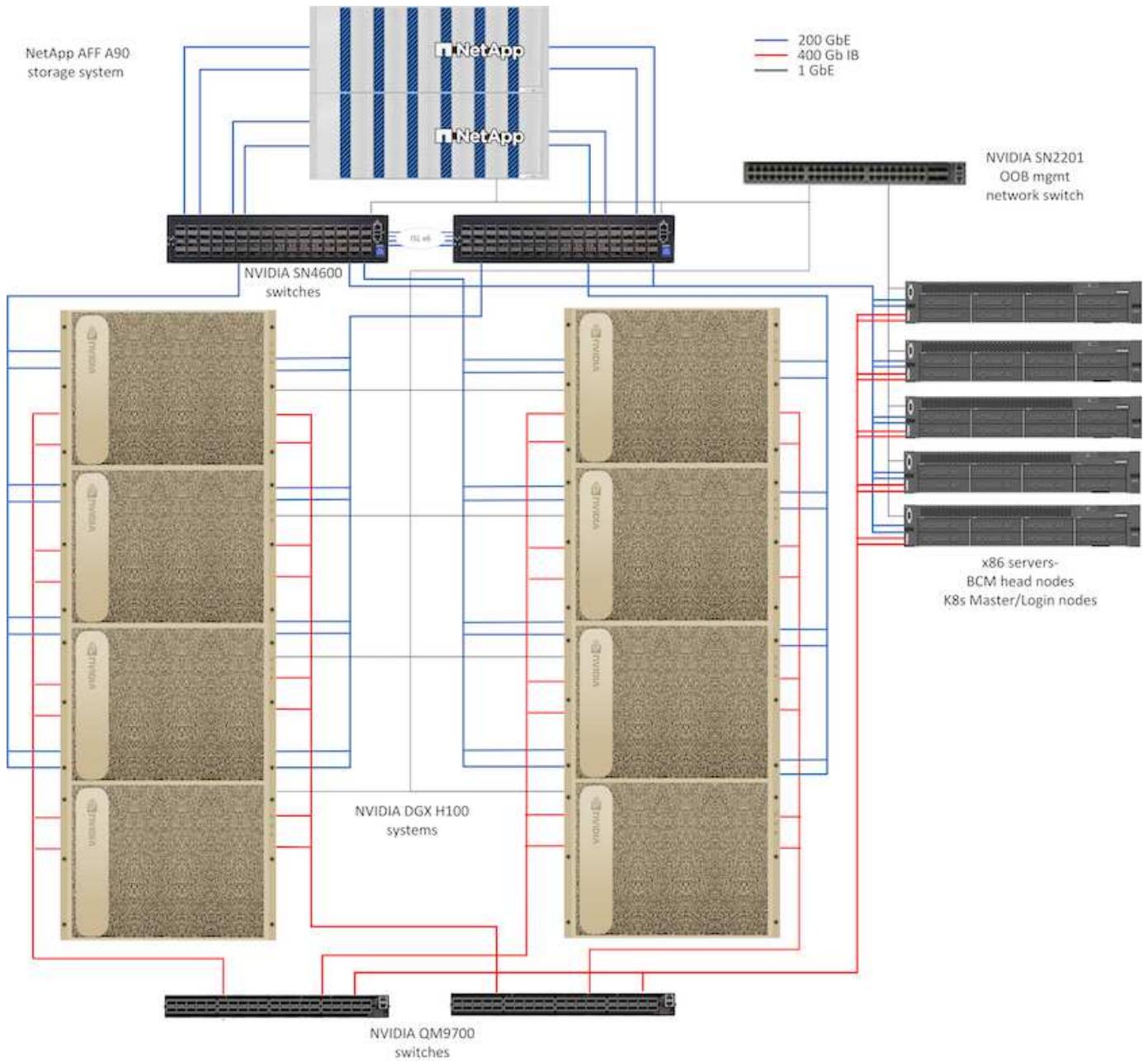
## NVA-1173 NetApp AIpod mit NVIDIA DGX H100-Systemen – Lösungsarchitektur

Dieser Abschnitt konzentriert sich auf die Architektur für den NetApp AIpod mit NVIDIA DGX-Systemen.

### NetApp AIpod mit DGX-Systemen

Diese Referenzarchitektur nutzt separate Fabrics für die Verbindung von Rechenclustern und den Speicherzugriff mit 400 Gb/s InfiniBand (IB)-Konnektivität zwischen Rechenknoten. Die folgende Zeichnung zeigt die Gesamtlösungstopologie von NetApp AIpod mit DGX H100-Systemen.

*Topologie der NetApp Alpod-Lösung*



## Netzwerkdesign

In dieser Konfiguration verwendet das Compute-Cluster-Fabric ein Paar QM9700 400Gb/s IB-Switches, die für eine hohe Verfügbarkeit miteinander verbunden sind. Jedes DGX H100-System ist über acht Verbindungen mit den Switches verbunden, wobei die Ports mit geraden Nummern mit einem Switch und die Ports mit ungeraden Nummern mit dem anderen Switch verbunden sind.

Für den Zugriff auf das Speichersystem, die In-Band-Verwaltung und den Clientzugriff wird ein Paar SN4600-Ethernet-Switches verwendet. Die Switches sind über Inter-Switch-Links verbunden und mit mehreren VLANs konfiguriert, um die verschiedenen Verkehrstypen zu isolieren. Grundlegendes L3-Routing wird zwischen bestimmten VLANs aktiviert, um mehrere Pfade zwischen Client- und Speicherschnittstellen auf demselben Switch sowie zwischen Switches für hohe Verfügbarkeit zu ermöglichen. Bei größeren Bereitstellungen kann das Ethernet-Netzwerk durch Hinzufügen zusätzlicher Switch-Paare für Spine-Switches und zusätzlicher Leaves nach Bedarf auf eine Leaf-Spine-Konfiguration erweitert werden.

Zusätzlich zur Computerverbindung und den Hochgeschwindigkeits-Ethernet-Netzwerken sind alle physischen

Geräte für die Out-of-Band-Verwaltung auch mit einem oder mehreren SN2201-Ethernet-Switches verbunden. Bitte beachten Sie die "[Bereitstellungsdetails](#)". Weitere Informationen zur Netzwerkkonfiguration finden Sie auf der Seite.

## Speicherzugriffsübersicht für DGX H100-Systeme

Jedes DGX H100-System ist mit zwei ConnectX-7-Adaptoren mit zwei Ports für Verwaltungs- und Speicherverkehr ausgestattet und für diese Lösung sind beide Ports auf jeder Karte mit demselben Switch verbunden. Ein Port von jeder Karte wird dann in eine LACP MLAG-Verbindung konfiguriert, wobei ein Port mit jedem Switch verbunden ist, und VLANs für In-Band-Management, Clientzugriff und Speicherzugriff auf Benutzerebene werden auf dieser Verbindung gehostet.

Der andere Port auf jeder Karte wird für die Verbindung mit den AFF A90 Speichersystemen verwendet und kann je nach Arbeitslastanforderungen in mehreren Konfigurationen verwendet werden. Bei Konfigurationen, die NFS über RDMA zur Unterstützung von NVIDIA Magnum IO GPUDirect Storage verwenden, werden die Ports einzeln mit IP-Adressen in separaten VLANs verwendet. Für Bereitstellungen, die kein RDMA erfordern, können die Speicherschnittstellen auch mit LACP-Bonding konfiguriert werden, um hohe Verfügbarkeit und zusätzliche Bandbreite bereitzustellen. Mit oder ohne RDMA können Clients das Speichersystem mithilfe von NFS v4.1 pNFS und Session Trunking mounten, um parallelen Zugriff auf alle Speicherknoten im Cluster zu ermöglichen. Bitte beachten Sie die "[Bereitstellungsdetails](#)" Seite für weitere Informationen zur Clientkonfiguration.

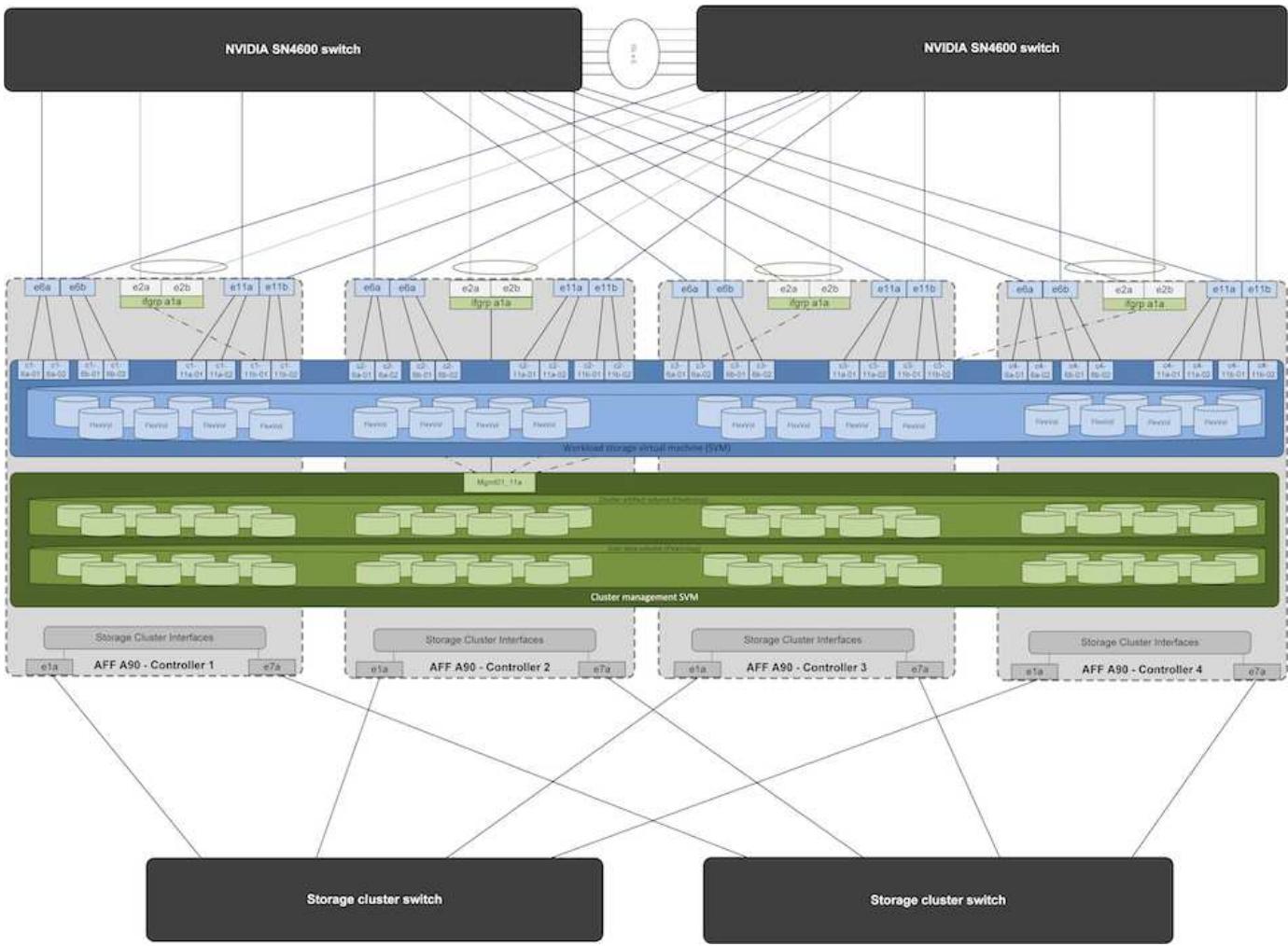
Weitere Einzelheiten zur Konnektivität des DGX H100-Systems finden Sie in der "[NVIDIA BasePOD-Dokumentation](#)".

## Speichersystemdesign

Jedes AFF A90 Speichersystem ist über sechs 200-GbE-Ports von jedem Controller aus verbunden. Vier Ports von jedem Controller werden für den Workload-Datenzugriff von den DGX-Systemen verwendet und zwei Ports von jedem Controller sind als LACP-Schnittstellengruppe konfiguriert, um den Zugriff von den Management-Plane-Servern auf Cluster-Management-Artefakte und Benutzer-Home-Verzeichnisse zu unterstützen. Der gesamte Datenzugriff vom Speichersystem erfolgt über NFS, wobei eine Storage Virtual Machine (SVM) für den Zugriff auf KI-Workloads und eine separate SVM für Clusterverwaltungszwecke vorgesehen ist.

Das Management-SVM benötigt nur ein einziges LIF, das auf den auf jedem Controller konfigurierten 2-Port-Schnittstellengruppen gehostet wird. Andere FlexGroup -Volumes werden auf der Management-SVM bereitgestellt, um Cluster-Management-Artefakte wie Cluster-Knoten-Images, historische Systemüberwachungsdaten und Endbenutzer-Home-Verzeichnisse aufzunehmen. Die folgende Zeichnung zeigt die logische Konfiguration des Speichersystems.

*Logische Konfiguration des NetApp A90-Speicherclusters*



## Management-Plane-Server

Diese Referenzarchitektur umfasst außerdem fünf CPU-basierte Server für den Einsatz auf Verwaltungsebene. Zwei dieser Systeme werden als Hauptknoten für NVIDIA Base Command Manager zur Clusterbereitstellung und -verwaltung verwendet. Die anderen drei Systeme werden verwendet, um zusätzliche Clusterdienste wie Kubernetes-Masterknoten oder Anmeldeknoten für Bereitstellungen zu bereitstellen, die Slurm für die Jobplanung verwenden. Bereitstellungen mit Kubernetes können den NetApp Trident CSI-Treiber nutzen, um automatisierte Bereitstellungs- und Datendienste mit persistentem Speicher für Management- und KI-Workloads auf dem AFF A900 Speichersystem bereitzustellen.

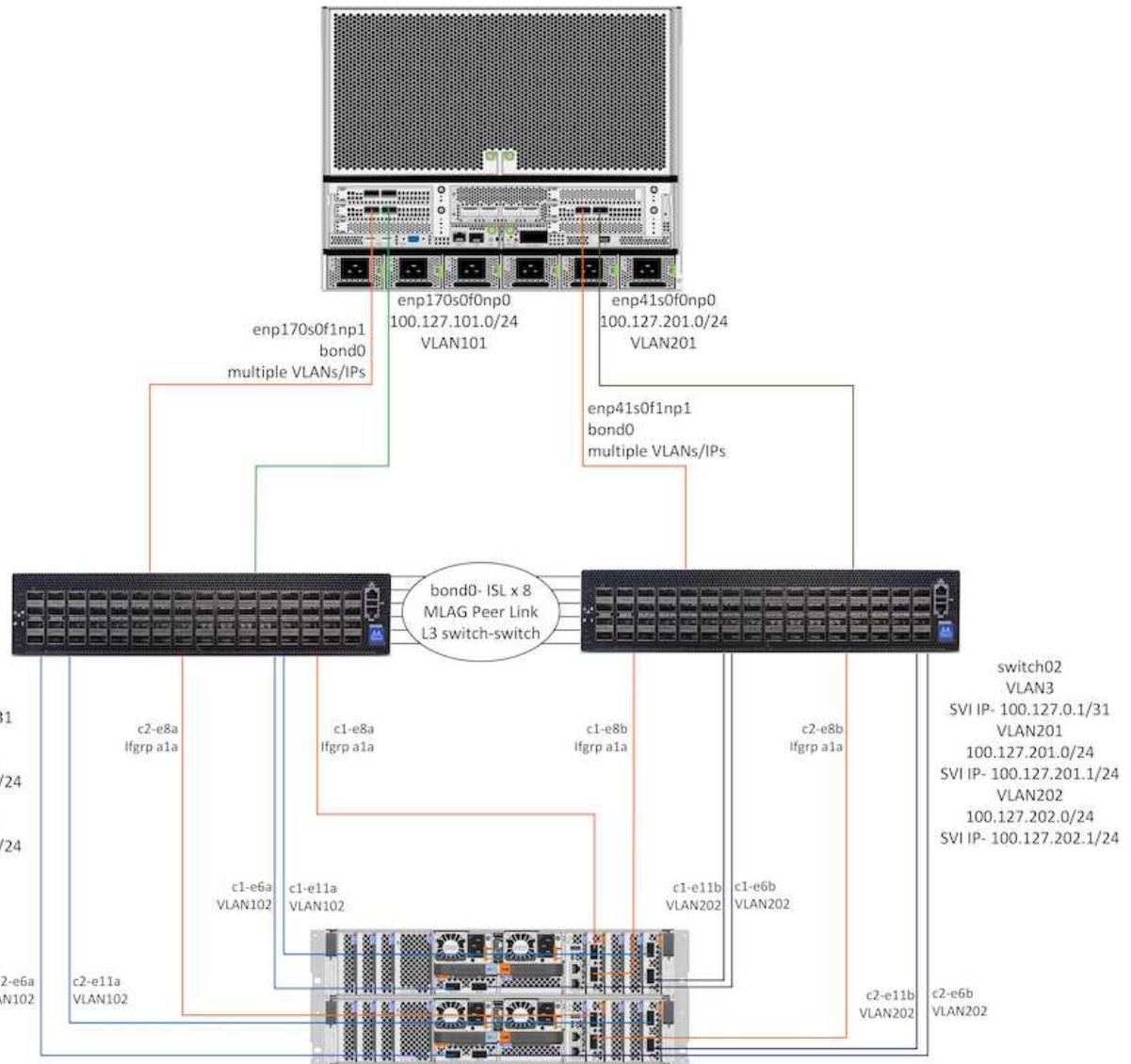
Jeder Server ist physisch mit den IB-Switches und den Ethernet-Switches verbunden, um die Clusterbereitstellung und -verwaltung zu ermöglichen, und mit NFS-Mounts zum Speichersystem über die Verwaltungs-SVM zur Speicherung von Clusterverwaltungssarifakten konfiguriert, wie zuvor beschrieben.

## NVA-1173 NetApp AIPod mit NVIDIA DGX-Systemen – Bereitstellungsdetails

In diesem Abschnitt werden die Bereitstellungsdetails beschrieben, die während der Validierung dieser Lösung verwendet wurden. Die verwendeten IP-Adressen sind Beispiele und sollten je nach Bereitstellungsumgebung geändert werden. Weitere Informationen zu bestimmten Befehlen, die bei der Implementierung dieser Konfiguration verwendet werden, finden Sie in der entsprechenden Produktdokumentation.

Das folgende Diagramm zeigt detaillierte Netzwerk- und Konnektivitätsinformationen für 1 DGX H100-System und 1 HA-Paar AFF A90 Controller. Die Bereitstellungshinweise in den folgenden Abschnitten basieren auf den Details in diesem Diagramm.

### NetApp Alpod-Netzwerkkonfiguration



Die folgende Tabelle zeigt beispielhafte Verkabelungszuweisungen für bis zu 16 DGX-Systeme und 2 AFF A90 HA-Paare.

Schalter und Port	Gerät	Geräteanschluss
Switch1-Ports 1-16	DGX-H100-01 bis -16	enp170s0f0np0, Steckplatz 1, Port 1
Switch1-Ports 17-32	DGX-H100-01 bis -16	enp170s0f1np1, Steckplatz 1, Port 2
Switch1-Ports 33-36	AFF-A90-01 bis -04	Anschluss E6a
Switch1-Ports 37-40	AFF-A90-01 bis -04	Port E11a
Switch1-Ports 41-44	AFF-A90-01 bis -04	Port E2A

Schalter und Port	Gerät	Geräteanschluss
Switch1-Ports 57-64	ISL zu Switch2	Anschlüsse 57-64
Switch2-Ports 1-16	DGX-H100-01 bis -16	enp41s0f0np0, Steckplatz 2, Port 1
Switch2-Ports 17-32	DGX-H100-01 bis -16	enp41s0f1np1, Steckplatz 2, Port 2
Switch2-Ports 33-36	AFF-A90-01 bis -04	Port E6b
Switch2-Ports 37-40	AFF-A90-01 bis -04	Port E11b
Switch2-Ports 41-44	AFF-A90-01 bis -04	Port E2B
Switch2-Ports 57-64	ISL zu Switch1	Anschlüsse 57-64

Die folgende Tabelle zeigt die Softwareversionen für die verschiedenen Komponenten, die bei dieser Validierung verwendet wurden.

Gerät	Softwareversion
NVIDIA SN4600-Switches	Cumulus Linux v5.9.1
NVIDIA DGX-System	DGX OS v6.2.1 (Ubuntu 22.04 LTS)
Mellanox OFED	24.01
NetApp AFF A90	NetApp ONTAP 9.14.1

## Speichernetzwerkkonfiguration

In diesem Abschnitt werden die wichtigsten Details zur Konfiguration des Ethernet-Speichernetzwerks beschrieben. Informationen zur Konfiguration des InfiniBand-Rechnernetzwerks finden Sie im "[NVIDIA BasePOD-Dokumentation](#)". Weitere Einzelheiten zur Switch-Konfiguration finden Sie im "[NVIDIA Cumulus Linux-Dokumentation](#)".

Die grundlegenden Schritte zur Konfiguration der SN4600-Switches werden unten beschrieben. Dieser Vorgang setzt voraus, dass die Verkabelung und die grundlegende Switch-Einrichtung (Verwaltung der IP-Adresse, Lizenzierung usw.) abgeschlossen sind.

1. Konfigurieren Sie die ISL-Verbindung zwischen den Switches, um Multi-Link Aggregation (MLAG) und Failover-Verkehr zu ermöglichen
  - Bei dieser Validierung wurden 8 Links verwendet, um mehr als genug Bandbreite für die getestete Speicherkonfiguration bereitzustellen.
  - Spezifische Anweisungen zum Aktivieren von MLAG finden Sie in der Cumulus Linux-Dokumentation.
2. Konfigurieren Sie LACP MLAG für jedes Paar von Client-Ports und Speicher-Ports auf beiden Switches
  - Port swp17 auf jedem Switch für DGX-H100-01 (enp170s0f1np1 und enp41s0f1np1), Port swp18 für DGX-H100-02 usw. (bond1-16)
  - Port swp41 auf jedem Switch für AFF-A90-01 (e2a und e2b), Port swp42 für AFF-A90-02 usw. (bond17-20)
  - nv set interface bondX bond member swpX
  - nv set interface bondx bond mlag id X
3. Fügen Sie alle Ports und MLAG-Verbindungen zur Standard-Bridge-Domäne hinzu

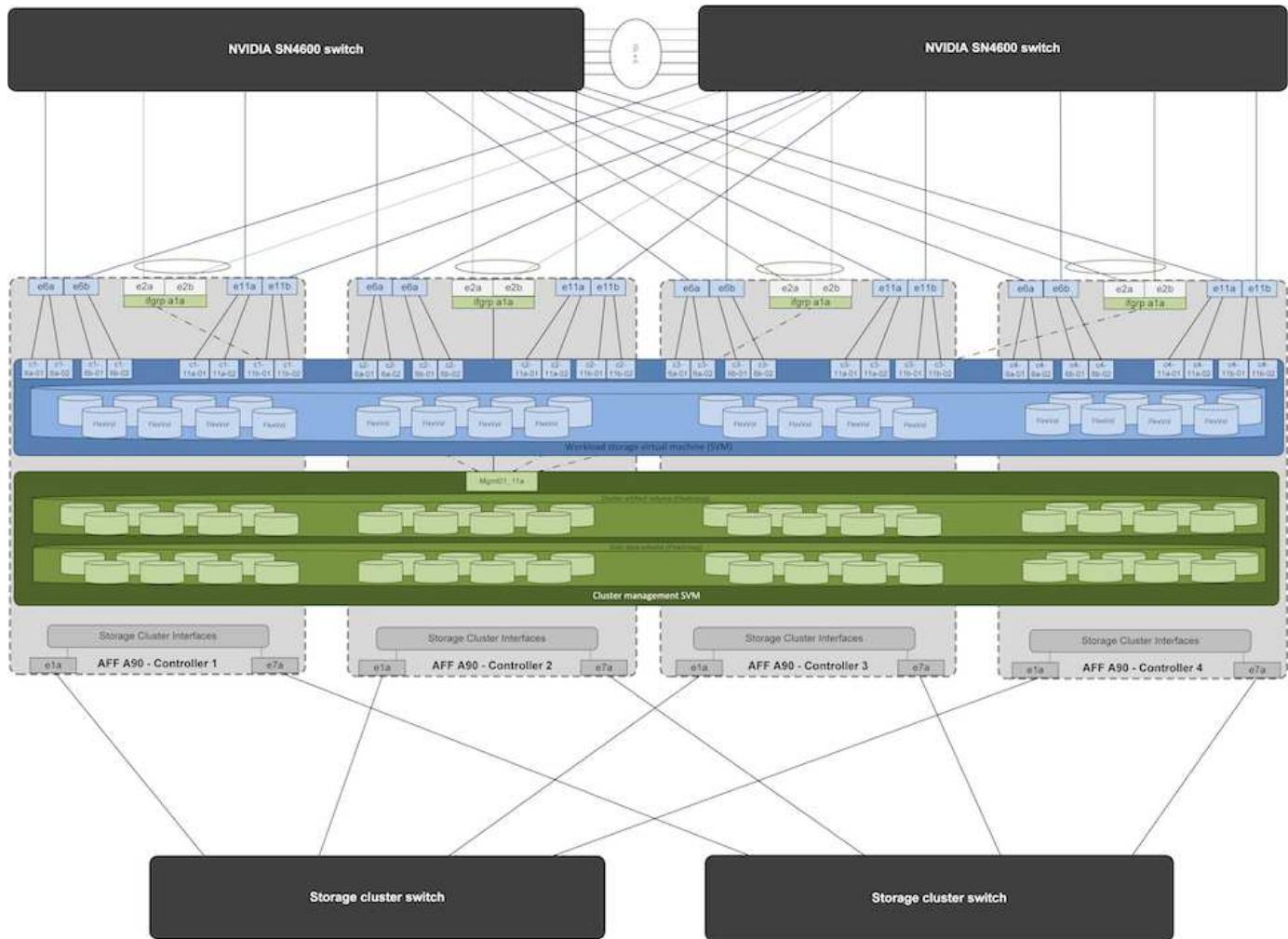
- nv set int swp1-16,33-40 Bridge-Domäne br\_default
  - nv set int bond1-20 Bridge-Domäne br\_default
4. Aktivieren Sie RoCE auf jedem Switch
- nv set roce mode lossless
5. Konfigurieren Sie VLANs – 2 für Client-Ports, 2 für Speicher-Ports, 1 für die Verwaltung, 1 für L3-Switch zu Switch
- Schalter 1-
    - VLAN 3 für L3-Switch-zu-Switch-Routing im Falle eines Client-NIC-Ausfalls
    - VLAN 101 für Speicherport 1 auf jedem DGX-System (enp170s0f0np0, Slot1-Port 1)
    - VLAN 102 für Port e6a und e11a auf jedem AFF A90 Speichercontroller
    - VLAN 301 für die Verwaltung über die MLAG-Schnittstellen zu jedem DGX-System und Speichercontroller
  - Schalter 2-
    - VLAN 3 für L3-Switch-zu-Switch-Routing im Falle eines Client-NIC-Ausfalls
    - VLAN 201 für Speicherport 2 auf jedem DGX-System (enp41s0f0np0, Slot2-Port 1)
    - VLAN 202 für Port e6b und e11b auf jedem AFF A90 Speichercontroller
    - VLAN 301 für die Verwaltung über die MLAG-Schnittstellen zu jedem DGX-System und Speichercontroller
6. Weisen Sie jedem VLAN die entsprechenden physischen Ports zu, z. B. Client-Ports in Client-VLANs und Speicher-Ports in Speicher-VLANs.
- nv set int <swpX> Bridge-Domäne br\_default access <VLAN-ID>
  - MLAG-Ports sollten als Trunk-Ports verbleiben, um bei Bedarf mehrere VLANs über die verbundenen Schnittstellen zu ermöglichen.
7. Konfigurieren Sie Switch Virtual Interfaces (SVI) auf jedem VLAN, um als Gateway zu fungieren und L3-Routing zu aktivieren
- Schalter 1-
    - nv set int vlan3 IP-Adresse 100.127.0.0/31
    - nv set int vlan101 IP-Adresse 100.127.101.1/24
    - nv set int vlan102 IP-Adresse 100.127.102.1/24
  - Schalter 2-
    - nv set int vlan3 IP-Adresse 100.127.0.1/31
    - nv set int vlan201 IP-Adresse 100.127.201.1/24
    - nv set int vlan202 IP-Adresse 100.127.202.1/24
8. Erstellen statischer Routen
- Statische Routen werden automatisch für Subnetze auf demselben Switch erstellt
  - Für das Switch-to-Switch-Routing sind im Falle eines Client-Link-Ausfalls zusätzliche statische Routen erforderlich.
    - Schalter 1-
      - nv set vrf Standardrouter statisch 100.127.128.0/17 über 100.127.0.1

- Schalter 2-
  - nv set vrf default router static 100.127.0.0/17 via 100.127.0.0

## Speichersystemkonfiguration

In diesem Abschnitt werden die wichtigsten Details zur Konfiguration des A90-Speichersystems für diese Lösung beschrieben. Weitere Einzelheiten zur Konfiguration von ONTAP -Systemen finden Sie im "[ONTAP-Dokumentation](#)". Das folgende Diagramm zeigt die logische Konfiguration des Speichersystems.

*Logische Konfiguration des NetApp A90-Speicherclusters*



Die grundlegenden Schritte zur Konfiguration des Speichersystems werden unten beschrieben. Dieser Vorgang setzt voraus, dass die grundlegende Installation des Speicherclusters abgeschlossen ist.

1. Konfigurieren Sie 1 Aggregat auf jedem Controller mit allen verfügbaren Partitionen abzüglich 1 Ersatz
  - `aggr create -node <Knoten> -aggregate <Knoten>_data01 -diskcount <47>`
2. Konfigurieren Sie ifgrps auf jedem Controller
  - Netzport ifgrp erstellen `-node <Knoten> -ifgrp a1a -mode multimode_lacp -distr-function port`
  - net port ifgrp add-port `-node <Knoten> -ifgrp <ifgrp> -ports <Knoten>:e2a,<Knoten>:e2b`
3. Konfigurieren Sie den Mgmt-VLAN-Port auf IFgrp auf jedem Controller

- net port vlan erstellen -node aff-a90-01 -port a1a -vlan-id 31
- net port vlan erstellen -node aff-a90-02 -port a1a -vlan-id 31
- net port vlan erstellen -node aff-a90-03 -port a1a -vlan-id 31
- net port vlan erstellen -node aff-a90-04 -port a1a -vlan-id 31

#### 4. Erstellen von Broadcastdomänen

- Broadcast-Domäne erstellen -Broadcast-Domäne vlan21 -MTU 9000 -Ports aff-a90-01:e6a,aff-a90-01:e11a,aff-a90-02:e6a,aff-a90-02:e11a,aff-a90-03:e6a,aff-a90-03:e11a,aff-a90-04:e6a,aff-a90-04:e11a
- Broadcast-Domäne erstellen -Broadcast-Domäne vlan22 -MTU 9000 -Ports aaff-a90-01:e6b,aff-a90-01:e11b,aff-a90-02:e6b,aff-a90-02:e11b,aff-a90-03:e6b,aff-a90-03:e11b,aff-a90-04:e6b,aff-a90-04:e11b
- Broadcast-Domäne erstellen -Broadcast-Domäne vlan31 -MTU 9000 -Ports aff-a90-01:a1a-31,aff-a90-02:a1a-31,aff-a90-03:a1a-31,aff-a90-04:a1a-31

#### 5. Management-SVM erstellen \*

#### 6. Konfigurieren der Verwaltungs-SVM

- LIF erstellen
  - net int create -vserver basepod-mgmt -lif vlan31-01 -home-node aff-a90-01 -home-port a1a-31 -address 192.168.31.X -netmask 255.255.255.0
- FlexGroup -Volumes erstellen-
  - vol erstellen -vserver basepod-mgmt -volume home -size 10T -auto-provision-as flexgroup -junction -path /home
  - vol erstellen -vserver basepod-mgmt -volume cm -size 10T -auto-provision-as flexgroup -junction -path /cm
- Exportrichtlinie erstellen
  - Exportrichtlinienregel erstellen -vserver basepod-mgmt -policy default -client-match 192.168.31.0/24 -rorule sys -rwrule sys -superuser sys

#### 7. Daten-SVM erstellen \*

#### 8. Daten-SVM konfigurieren

- Konfigurieren Sie SVM für RDMA-Unterstützung
  - vserver nfs modify -vserver basepod-data -rdma enabled
- LIFs erstellen
  - net int create -vserver basepod-data -lif c1-6a-lif1 -home-node aff-a90-01 -home-port e6a -address 100.127.102.101 -netmask 255.255.255.0
  - net int create -vserver basepod-data -lif c1-6a-lif2 -home-node aff-a90-01 -home-port e6a -address 100.127.102.102 -netmask 255.255.255.0
  - net int create -vserver basepod-data -lif c1-6b-lif1 -home-node aff-a90-01 -home-port e6b -address 100.127.202.101 -netmask 255.255.255.0
  - net int create -vserver basepod-data -lif c1-6b-lif2 -home-node aff-a90-01 -home-port e6b -address 100.127.202.102 -netmask 255.255.255.0
  - net int create -vserver basepod-data -lif c1-11a-lif1 -home-node aff-a90-01 -home-port e11a -address 100.127.102.103 -netmask 255.255.255.0
  - net int create -vserver basepod-data -lif c1-11a-lif2 -home-node aff-a90-01 -home-port e11a -address 100.127.102.104 -netmask 255.255.255.0

- net int create -vserver basepod-data -lif c1-11b-lif1 -home-node aff-a90-01 -home-port e11b -address 100.127.202.103 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c1-11b-lif2 -home-node aff-a90-01 -home-port e11b -address 100.127.202.104 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-6a-lif1 -home-node aff-a90-02 -home-port e6a -address 100.127.102.105 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-6a-lif2 -home-node aff-a90-02 -home-port e6a -address 100.127.102.106 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-6b-lif1 -home-node aff-a90-02 -home-port e6b -address 100.127.202.105 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-6b-lif2 -home-node aff-a90-02 -home-port e6b -address 100.127.202.106 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-11a-lif1 -home-node aff-a90-02 -home-port e11a -address 100.127.102.107 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-11a-lif2 -home-node aff-a90-02 -home-port e11a -address 100.127.102.108 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-11b-lif1 -home-node aff-a90-02 -home-port e11b -address 100.127.202.107 -netmask 255.255.255.0
- net int create -vserver basepod-data -lif c2-11b-lif2 -home-node aff-a90-02 -home-port e11b -address 100.127.202.108 -netmask 255.255.255.0

## 9. Konfigurieren von LIFs für RDMA-Zugriff

- Bei Bereitstellungen mit ONTAP 9.15.1 erfordert die RoCE-QoS-Konfiguration für physische Informationen Befehle auf Betriebssystemebene, die in der ONTAP CLI nicht verfügbar sind. Wenden Sie sich an den NetApp -Support, um Hilfe bei der Konfiguration der Ports für die RoCE-Unterstützung zu erhalten. NFS über RDMA funktioniert ohne Probleme
- Ab ONTAP 9.16.1 werden physische Schnittstellen automatisch mit den entsprechenden Einstellungen für die End-to-End-RoCE-Unterstützung konfiguriert.
- net int modifizieren -vserver basepod-data -lif \* -rdma-protocols roce

## 10. Konfigurieren Sie NFS-Parameter auf der Daten-SVM

- nfs ändern -vserver basepod-data -v4.1 aktiviert -v4.1-pnfs aktiviert -v4.1-trunking aktiviert -tcp-max -transfer-size 262144

## 11. FlexGroup -Volumes erstellen-

- vol erstellen -vserver basepod-data -volume data -size 100T -auto-provision-as flexgroup -junction-path /data

## 12. Exportrichtlinie erstellen

- Exportrichtlinienregel erstellen -vserver basepod-data -policy default -client-match 100.127.101.0/24 -rorule sys -rwrule sys -superuser sys
- Exportrichtlinienregel erstellen -vserver basepod-data -policy default -client-match 100.127.201.0/24 -rorule sys -rwrule sys -superuser sys

## 13. Routen erstellen

- Route hinzufügen -vserver basepod\_data -destination 100.127.0.0/17 -gateway 100.127.102.1 Metrik 20
- Route hinzufügen -vserver basepod\_data -destination 100.127.0.0/17 -gateway 100.127.202.1 Metrik 30

- Route hinzufügen -vserver basepod\_data -destination 100.127.128.0/17 -gateway 100.127.202.1 Metrik 20
- Route hinzufügen -vserver basepod\_data -destination 100.127.128.0/17 -gateway 100.127.102.1 Metrik 30

## DGX H100-Konfiguration für RoCE-Speicherzugriff

In diesem Abschnitt werden die wichtigsten Details zur Konfiguration der DGX H100-Systeme beschrieben. Viele dieser Konfigurationselemente können in das auf den DGX-Systemen bereitgestellte Betriebssystem-Image aufgenommen oder beim Booten vom Base Command Manager implementiert werden. Sie sind hier als Referenz aufgeführt. Weitere Informationen zum Konfigurieren von Knoten und Software-Images in BCM finden Sie im "[BCM-Dokumentation](#)" .

### 1. Installieren Sie zusätzliche Pakete

- ipmitool
- python3-pip

### 2. Installieren Sie Python-Pakete

- Paramiko
- matplotlib

### 3. Konfigurieren Sie dpkg nach der Paketinstallation neu

- dpkg --configure -a

### 4. Installieren von MOFED

### 5. Legen Sie die MST-Werte zur Leistungsoptimierung fest

- mstconfig -y -d <aa:00.0,29:00.0> set ADVANCED\_PCI\_SETTINGS=1 NUM\_OF\_VFS=0 MAX\_ACC\_OUT\_READ=44

### 6. Setzen Sie die Adapter nach dem Ändern der Einstellungen zurück

- mlxfwreset -d <aa:00.0,29:00.0> -y zurücksetzen

### 7. MaxReadReq auf PCI-Geräten festlegen

- setpci -s <aa:00.0,29:00.0> 68.W=5957

### 8. Legen Sie die Größe des RX- und TX-Ringpuffers fest

- ethtool -G <enp170s0f0np0,enp41s0f0np0> rx 8192 tx 8192

### 9. Legen Sie PFC und DSCP mit mlnx\_qos fest

- mlnx\_qos -i <enp170s0f0np0,enp41s0f0np0> --pfc 0,0,0,1,0,0,0,0 --trust=dscp --cable\_len=3

### 10. Legen Sie ToS für RoCE-Verkehr auf Netzwerkports fest

- echo 106 > /sys/class/infiniband/<mlx5\_7,mlx5\_1>/tc/1/traffic\_class

### 11. Konfigurieren Sie jede Speicher-NIC mit einer IP-Adresse im entsprechenden Subnetz

- 100.127.101.0/24 für Speicher-NIC 1
- 100.127.201.0/24 für Speicher-NIC 2

### 12. Konfigurieren Sie In-Band-Netzwerkports für LACP-Bonding (enp170s0f1np1,enp41s0f1np1).

### 13. Konfigurieren Sie statische Routen für primäre und sekundäre Pfade zu jedem Speichersubnetz

- Route hinzufügen –net 100.127.0.0/17 gw 100.127.101.1 Metrik 20

- Route hinzufügen –net 100.127.0.0/17 gw 100.127.201.1 Metrik 30
- Route hinzufügen –net 100.127.128.0/17 gw 100.127.201.1 Metrik 20
- Route hinzufügen –net 100.127.128.0/17 gw 100.127.101.1 Metrik 30

#### 14. Mounten Sie das /home-Volume

- mount -o vers=3,nconnect=16,rsize=262144,wsize=262144 192.168.31.X:/home /home

#### 15. Mounten/Datenvolumen

- Beim Mounten des Datenvolumes wurden die folgenden Mount-Optionen verwendet:
  - vers=4.1 # aktiviert pNFS für den parallelen Zugriff auf mehrere Speicherknoten
  - proto=rdma # setzt das Übertragungsprotokoll auf RDMA statt auf das Standard-TCP
  - max\_connect=16 # aktiviert NFS-Sitzungs-Trunking, um die Bandbreite des Speicherports zu aggregieren
  - write=eager # verbessert die Schreibleistung von gepufferten Schreibvorgängen
  - rsize=262144,wsize=262144 # setzt die I/O-Übertragungsgröße auf 256k

## NVA-1173 NetApp AI Pod mit NVIDIA DGX-Systemen – Leitfaden zur Lösungsvalidierung und Größenbestimmung

Dieser Abschnitt konzentriert sich auf die Lösungsvalidierung und Größenrichtlinien für den NetApp AI Pod mit NVIDIA DGX-Systemen.

### Lösungsvalidierung

Die Speicherkonfiguration in dieser Lösung wurde mithilfe einer Reihe synthetischer Workloads unter Verwendung des Open-Source-Tools FIO validiert. Diese Tests umfassen Lese- und Schreib-E/A-Muster, die die Speicherarbeitslast simulieren sollen, die von DGX-Systemen generiert wird, die Deep-Learning-Trainingsjobs ausführen. Die Speicherkonfiguration wurde mithilfe eines Clusters aus 2-Sockel-CPU-Servern validiert, auf denen die FIO-Workloads gleichzeitig ausgeführt wurden, um einen Cluster aus DGX-Systemen zu simulieren. Jeder Client wurde mit der gleichen Netzwerkkonfiguration wie zuvor beschrieben konfiguriert, wobei die folgenden Details hinzugefügt wurden.

Für diese Validierung wurden die folgenden Mount-Optionen verwendet:

vers=4.1	ermöglicht pNFS für den parallelen Zugriff auf mehrere Speicherknoten
proto=rdma	setzt das Übertragungsprotokoll auf RDMA statt auf das Standard-TCP
Port=20049	Geben Sie den richtigen Port für den RDMA-NFS-Dienst an
max_connect=16	ermöglicht NFS-Sitzungsbündelung zur Aggregation der Speicherportbandbreite
schreiben=eifrig	verbessert die Schreibleistung gepufferter Schreibvorgänge
rsize=262144,wsize=262144	setzt die I/O-Transfergröße auf 256k

Darüber hinaus wurden die Clients mit einem NFS-max\_session\_slots-Wert von 1024 konfiguriert. Da die Lösung mit NFS über RDMA getestet wurde, wurden die Ports der Speichernetzwerke mit einer aktiven/passiven Verbindung konfiguriert. Für diese Validierung wurden folgende Bindungsparameter verwendet:

Modus=aktive Sicherung	setzt die Bindung in den Aktiv/Passiv-Modus
primary=<Schnittstellenname>	Primäre Schnittstellen für alle Clients wurden auf die Switches verteilt
mii-monitor-interval=100	gibt Überwachungsintervall von 100ms an
Failover-Mac-Richtlinie = aktiv	gibt an, dass die MAC-Adresse des aktiven Links die MAC der Verbindung ist. Dies ist für den ordnungsgemäßen Betrieb von RDMA über die verbundene Schnittstelle erforderlich.

Das Speichersystem wurde wie beschrieben mit zwei A900 HA-Paaren (4 Controller) mit zwei NS224-Festplattenregalen mit 24 1,9 TB NVMe-Festplatten konfiguriert, die an jedes HA-Paar angeschlossen waren. Wie im Abschnitt zur Architektur erwähnt, wurde die Speicherkapazität aller Controller mithilfe eines FlexGroup Volumes kombiniert und die Daten aller Clients auf alle Controller im Cluster verteilt.

## Leitfaden zur Größenbestimmung von Speichersystemen

NetApp hat die DGX BasePOD-Zertifizierung erfolgreich abgeschlossen und die beiden getesteten A90 HA-Paare können problemlos einen Cluster aus sechzehn DGX H100-Systemen unterstützen. Für größere Bereitstellungen mit höheren Anforderungen an die Speicherleistung können dem NetApp ONTAP -Cluster zusätzliche AFF -Systeme hinzugefügt werden, bis zu 12 HA-Paare (24 Knoten) in einem einzigen Cluster. Mithilfe der in dieser Lösung beschriebenen FlexGroup -Technologie kann ein Cluster mit 24 Knoten über 79 PB und bis zu 552 GBps Durchsatz in einem einzigen Namespace bereitstellen. Andere NetApp Speichersysteme wie AFF A400, A250 und C800 bieten Optionen mit geringerer Leistung und/oder höherer Kapazität für kleinere Bereitstellungen zu niedrigeren Kosten. Da ONTAP 9 Cluster mit gemischten Modellen unterstützt, können Kunden mit einem kleineren anfänglichen Footprint beginnen und dem Cluster weitere oder größere Speichersysteme hinzufügen, wenn die Kapazitäts- und Leistungsanforderungen steigen. Die folgende Tabelle zeigt eine grobe Schätzung der Anzahl der von jedem AFF Modell unterstützten A100- und H100-GPUs.

### Leitfaden zur Größenbestimmung von NetApp-Speichersystemen

		Throughput <sup>2</sup>	Raw capacity (typical <sup>3</sup> / max)	Connectivity	# NVIDIA A100 GPUs supported <sup>4</sup>	# NVIDIA H100 GPUs supported <sup>5</sup>
NetApp® AFF A1K	1 HA pair <sup>1</sup>	56 GB/s	368TB / 14.7PB	200 GbE	1-160	1-80
	12 HA pairs	672 GB/s	4.4PB / 176.4PB		1920	960
AFF A90	1 HA pair	46 GB/s	368TB / 6.6PB	200 GbE	1 – 128	1-64
	12 HA pairs	552 GB/s	4.4PB / 79.2PB		1536	768
AFF A70	1 HA pair	21 GB/s	368TB / 6.6PB	200 GbE	1-48	1-24
	12 HA pairs	252 GB/s	4.4PB / 79.2PB		576	288

## NVA-1173 NetApp AIpod mit NVIDIA DGX-Systemen – Fazit und weitere Informationen

Dieser Abschnitt enthält Verweise auf zusätzliche Informationen zum NetApp AIpod mit NVIDIA DGX-Systemen.

## Abschluss

Die DGX BasePOD-Architektur ist eine Deep-Learning-Plattform der nächsten Generation, die ebenso fortschrittliche Speicher- und Datenverwaltungsfunktionen erfordert. Durch die Kombination von DGX BasePOD mit NetApp AFF -Systemen kann die NetApp AIPod -Architektur mit DGX-Systemen in nahezu jedem Maßstab implementiert werden. In Kombination mit der überlegenen Cloud-Integration und den softwaredefinierten Funktionen von NetApp ONTAP ermöglicht AFF eine vollständige Palette von Datenpipelines, die sich über den Rand, den Kern und die Cloud erstrecken, für erfolgreiche DL-Projekte.

## Weitere Informationen

Weitere Informationen zu den in diesem Dokument beschriebenen Informationen finden Sie in den folgenden Dokumenten und/oder auf den folgenden Websites:

- NetApp ONTAP Datenverwaltungssoftware – ONTAP Informationsbibliothek  
["https://docs.netapp.com/us-en/ontap-family/"](https://docs.netapp.com/us-en/ontap-family/)
- NetApp AFF A90 Speichersysteme-  
<https://www.netapp.com/pdf.html?item=/media/7828-ds-3582-aff-a-series-ai-era.pdf>
- NetApp ONTAP RDMA-Informationen –  
["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 Trident  
["Überblick"](#)
- NetApp GPUDirect Storage Blog –  
["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/)
- NVIDIA DGX H100-Systeme  
["https://www.nvidia.com/en-us/data-center/dgx-h100/"](https://www.nvidia.com/en-us/data-center/dgx-h100/)
- NVIDIA Netzwerk  
["https://www.nvidia.com/en-us/networking/"](https://www.nvidia.com/en-us/networking/)
- NVIDIA Magnum IO – GPUDirect – Speicher  
["https://docs.nvidia.com/gpudirect-storage/"](https://docs.nvidia.com/gpudirect-storage/)
- NVIDIA Basisbefehl

["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/)

## Danksagung

Dieses Dokument ist das Werk der NetApp Solutions- und ONTAP Engineering-Teams – David Arnette, Olga Kornievskaia, Dustin Fischer, Srikanth Kaligotla, Mohit Kumar und Raghuram Sudhaakar. Die Autoren möchten sich außerdem bei NVIDIA und dem NVIDIA DGX BasePOD Entwicklungsteam für ihre anhaltende Unterstützung bedanken.

## **Copyright-Informationen**

Copyright © 2026 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGENDERWEINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

**ERLÄUTERUNG ZU „RESTRICTED RIGHTS“:** Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

## **Markeninformationen**

NETAPP, das NETAPP Logo und die unter <http://www.netapp.com/TM> aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.