NVIDIA DGX SuperPOD mit NetApp – Designleitfaden

NetApp artificial intelligence solutions

NetApp
August 18, 2025

This PDF was generated from https://docs.netapp.com/de-de/netapp-solutions-ai/infra/ai-eseries-superpod.html on August 18, 2025. Always check docs.netapp.com for the latest.

Inhalt

NVIDIA DGX SuperPOD mit NetApp – Designleitfaden . . . 1
Zusammenfassung . . . 1
Lösungsübersicht . . . 1
Lösungstechnologie . . . 1
Zusammenfassung des Anwendungsfalls . . . 3
Technologieanforderungen . . . 3
Hardwareanforderungen . . . 3
Softwareanforderungen . . . 4
Lösungsüberprüfung . . . 4
Abschluss . . . 4
Wo Sie weitere Informationen finden . . . 4

NVIDIA DGX SuperPOD mit NetApp – Designleitfaden

Diese NetApp Verified Architecture beschreibt das Design des NVIDIA DGX SuperPOD mit NetApp BeeGFS-Bausteinen. Bei dieser Lösung handelt es sich um eine Full-Stack-Rechenzentrumsplattform, die auf einem dedizierten Akzeptanzcluster bei NVIDIA validiert wurde.

Amine Bennani, Christian Whiteside, David Arnette und Sathish Thyagarajan, NetApp

Zusammenfassung

In der heutigen, sich schnell entwickelnden Technologielandschaft revolutioniert KI das Kundenerlebnis und treibt Innovationen in allen Branchen voran. Allerdings stellt es auch erhebliche Herausforderungen für IT-Abteilungen dar, die unter dem Druck stehen, High-Performance-Computing-Lösungen (HPC) einzusetzen, die den hohen Anforderungen von KI-Workloads gewachsen sind. Während Unternehmen darum wetteifern, die Leistungsfähigkeit der KI zu nutzen, wächst die Dringlichkeit einer Lösung, die einfach bereitzustellen, zu skalieren und zu verwalten ist.

NVIDIA DGX SuperPOD ist eine KI-Rechenzentrumsinfrastrukturplattform, die als schlüsselfertige Lösung für die IT bereitgestellt wird, um die komplexesten KI-Workloads zu unterstützen, mit denen Unternehmen heute konfrontiert sind. Den Kern jedes präzisen Deep-Learning-Modells (DL) bilden große Datenmengen, die eine Hochdurchsatz-Speicherlösung erfordern, die diese Daten effizient bereitstellen und erneut bereitstellen kann. Die NetApp BeeGFS-Lösung, bestehend aus NetApp EF600-Speicherarrays mit dem parallelen BeeGFS-Dateisystem, ermöglicht es dem NVIDIA DGX SuperPOD, sein volles Potenzial zu entfalten. Die NetApp BeeGFS-Lösung wurde von NVIDIA für die Integration und Skalierung mit der SuperPOD-Architektur validiert. Das Ergebnis ist eine vereinfachte Bereitstellung und Verwaltung von KI-Rechenzentren bei gleichzeitig nahezu unbegrenzter Skalierbarkeit hinsichtlich Leistung und Kapazität.

Lösungsübersicht

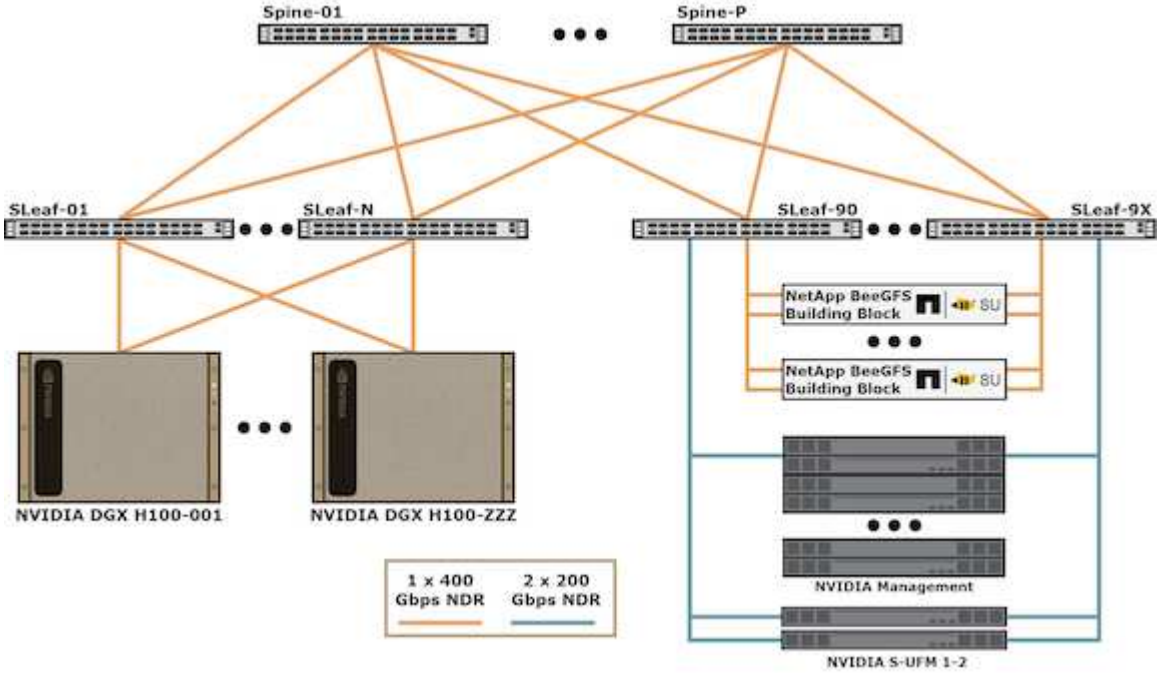
Die NetApp BeeGFS-Lösung, die auf den leistungsstarken NetApp EF600 NVMe-Speichersystemen und dem skalierbaren parallelen BeeGFS-Dateisystem basiert, bietet eine robuste und effiziente Speichergrundlage für anspruchsvolle KI-Workloads. Seine Shared-Disk-Architektur gewährleistet hohe Verfügbarkeit und sorgt für gleichbleibende Leistung und Zugänglichkeit, selbst bei Systemproblemen. Diese Lösung bietet eine skalierbare und flexible Architektur, die an unterschiedliche Speicheranforderungen angepasst werden kann. Kunden können ihre Speicherleistung und -kapazität problemlos erweitern, indem sie zusätzliche Speicherbausteine integrieren, um selbst die anspruchsvollsten Arbeitslasten zu bewältigen.

Lösungstechnologie

- NVIDIA DGX SuperPOD nutzt DGX H100- und H200-Systeme mit einem validierten, extern angeschlossenen gemeinsam genutzten Speicher:

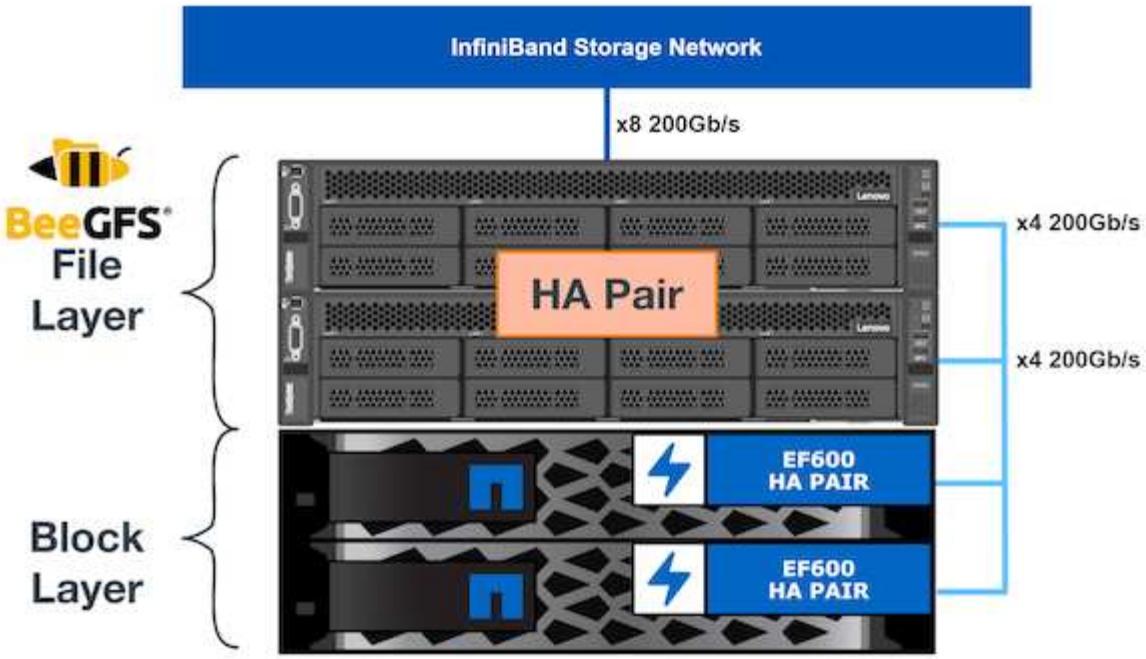
- Jede skalierbare DGX SuperPOD-Einheit (SU) besteht aus 32 DGX-Systemen und ist zu einer KI-Leistung von 640 PetaFLOPS bei FP8-Präzision fähig. NetApp empfiehlt, die NetApp BeeGFS-Speicherlösung mit mindestens 2 Bausteinen für eine einzelne DGX SuperPOD-Konfiguration zu dimensionieren.

Eine Übersicht über die Lösung



- Die NetApp BeeGFS-Bausteine bestehen aus zwei NetApp EF600-Arrays und zwei x86-Servern:
  - Mit den All-Flash-Arrays NetApp EF600 als Grundlage von NVIDIA DGX SuperPOD erhalten Kunden eine zuverlässige Speichergrundlage mit einer Verfügbarkeit von 9 …
  - Die Dateisystemschicht zwischen den NetApp EF600- und den NVIDIA DGX-Systemen ist das parallele Dateisystem BeeGFS. BeeGFS wurde vom Fraunhofer-Zentrum für Hochleistungsrechnen in Deutschland entwickelt, um die Schwachstellen älterer paralleler Dateisysteme zu lösen. Das Ergebnis ist ein Dateisystem mit einer modernen User-Space-Architektur, das jetzt von ThinkParQ entwickelt und bereitgestellt und von vielen Supercomputing-Umgebungen verwendet wird.
  - Der NetApp -Support für BeeGFS richtet die hervorragende Support-Organisation von NetApp an den Kundenanforderungen hinsichtlich Leistung und Verfügbarkeit aus. Kunden erhalten Zugriff auf erstklassige Supportressourcen, frühzeitigen Zugriff auf BeeGFS-Versionen und Zugriff auf ausgewählte BeeGFS-Unternehmensfunktionen wie Quotendurchsetzung und Hochverfügbarkeit (HA).
- Die Kombination aus NVIDIA SuperPOD SUs und NetApp BeeGFS-Bausteinen bietet eine agile KI-Lösung, bei der Rechenleistung oder Speicher einfach und nahtlos skaliert werden können.

NetApp BeeGFS-Baustein



Zusammenfassung des Anwendungsfalls

Diese Lösung gilt für die folgenden Anwendungsfälle:

- Künstliche Intelligenz (KI), einschließlich maschinelles Lernen (ML), Deep Learning (DL), Verarbeitung natürlicher Sprache (NLP), Verständnis natürlicher Sprache (NLU) und generative KI (GenAI).
- KI-Training im mittleren bis großen Maßstab
- Computer Vision, Sprache, Audio und Sprachmodelle
- HPC, einschließlich Anwendungen, die durch Message Passing Interface (MPI) und andere verteilte Computertechniken beschleunigt werden
- Anwendungs-Workloads, die durch Folgendes gekennzeichnet sind:
  - Lesen oder Schreiben in Dateien, die größer als 1 GB sind
  - Lesen oder Schreiben in dieselbe Datei durch mehrere Clients (10er, 100er und 1000er)
- Multiterabyte- oder Multipetabyte-Datensätze
- Umgebungen, die einen einzigen Speichernamespace benötigen, der für eine Mischung aus großen und kleinen Dateien optimiert werden kann

Technologieanforderungen

Dieser Abschnitt behandelt die Technologieanforderungen für die NVIDIA DGX SuperPOD Lösung mit NetApp .

Hardwareanforderungen

In der folgenden Tabelle 1 sind die Hardwarekomponenten aufgeführt, die zur Implementierung der Lösung für eine einzelne SU erforderlich sind. Die Dimensionierung der Lösung beginnt mit 32 NVIDIA DGX H100-Systemen und zwei oder drei NetApp BeeGFS-Bausteinen. Ein einzelner NetApp BeeGFS-Baustein besteht aus zwei NetApp EF600-Arrays und zwei x86-Servern. Kunden können zusätzliche Bausteine hinzufügen,

wenn die Bereitstellungsgröße zunimmt. Weitere Informationen finden Sie im "NVIDIA DGX H100 SuperPOD-Referenzarchitektur" Und "NVA-1164-DESIGN: BeeGFS auf NetApp NVA-Design" .

Hardware	Menge
NVIDIA DGX H100 oder H200	32
NVIDIA Quantum QM9700-Switches	8 Blätter, 4 Rücken
NetApp BeeGFS-Bausteine	3

Softwareanforderungen

In der folgenden Tabelle 2 sind die zur Implementierung der Lösung erforderlichen Softwarekomponenten aufgeführt. Die in einer bestimmten Implementierung der Lösung verwendeten Softwarekomponenten können je nach Kundenanforderungen variieren.

Software
NVIDIA DGX-Software-Stack
NVIDIA Base Command Manager
Paralleles Dateisystem ThinkParQ BeeGFS

Lösungsüberprüfung

NVIDIA DGX SuperPOD mit NetApp wurde auf einem dedizierten Akzeptanzcluster bei NVIDIA unter Verwendung von NetApp BeeGFS-Bausteinen validiert. Die Annahmekriterien basierten auf einer Reihe von Anwendungs-, Leistungs- und Belastungstests, die von NVIDIA durchgeführt wurden. Weitere Informationen finden Sie im "NVIDIA DGX SuperPOD: NetApp EF600 und BeeGFS-Referenzarchitektur" .

Abschluss

NetApp und NVIDIA arbeiten seit langem zusammen, um ein Portfolio von KI-Lösungen auf den Markt zu bringen. NVIDIA DGX SuperPOD mit dem NetApp EF600 All-Flash-Array ist eine bewährte, validierte Lösung, die Kunden bedenkenlos einsetzen können. Diese vollständig integrierte, schlüsselfertige Architektur eliminiert das Risiko der Bereitstellung und ermöglicht jedem, das Rennen um die KI-Führung zu gewinnen.

Wo Sie weitere Informationen finden

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

- "NVIDIA DGX SuperPOD -Referenzarchitektur"
- "NVIDIA DGX SuperPOD Referenzhandbuch zum Design von Rechenzentren"
- "NVIDIA DGX SuperPOD: NetApp EF600 und BeeGFS"

Copyright-Informationen

Copyright © 2025 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 BESCHRÄNKT 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 BESCHRÄNKT 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 IRGENDEINER 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.