NVIDIA DGX SuperPOD com NetApp - Guia de Design

NetApp artificial intelligence solutions

NetApp
August 18, 2025

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

Índice

NVIDIA DGX SuperPOD com NetApp - Guia de Design . . . 1
Sumário executivo . . . 1
Visão geral da solução . . . 1
Tecnologia de soluções . . . 1
Resumo do caso de uso . . . 3
Requisitos de tecnologia . . . 3
Requisitos de hardware . . . 3
Requisitos de software . . . 4
Verificação da solução . . . 4
Conclusão . . . 4
Onde encontrar informações adicionais . . . 4

NVIDIA DGX SuperPOD com NetApp - Guia de Design

Esta Arquitetura Verificada da NetApp descreve o design do NVIDIA DGX SuperPOD com blocos de construção NetApp BeeGFS. Esta solução é uma plataforma de data center full-stack validada em um cluster de aceitação dedicado na NVIDIA.

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

Sumário executivo

No cenário tecnológico de rápida evolução de hoje, a IA está revolucionando as experiências do consumidor e impulsionando a inovação em todos os setores. No entanto, também apresenta desafios significativos para os departamentos de TI, que estão sob pressão para implantar soluções de computação de alto desempenho (HPC) capazes de lidar com as intensas demandas das cargas de trabalho de IA. À medida que as organizações correm para aproveitar o poder da IA, cresce a urgência por uma solução que seja fácil de implantar, dimensionar e gerenciar.

NVIDIA DGX SuperPOD é uma plataforma de infraestrutura de data center de IA fornecida como uma solução pronta para uso para TI, para dar suporte às cargas de trabalho de IA mais complexas enfrentadas pelas empresas de hoje. No centro de qualquer modelo preciso de aprendizado profundo (DL) estão grandes volumes de dados, exigindo uma solução de armazenamento de alto rendimento que possa servir e servir novamente esses dados de forma eficiente. A solução NetApp BeeGFS, composta por matrizes de armazenamento NetApp EF600 com o sistema de arquivos paralelo BeeGFS, permite que o NVIDIA DGX SuperPOD libere toda a sua capacidade. A solução NetApp BeeGFS foi validada pela NVIDIA para integração e escalabilidade com a arquitetura SuperPOD. O resultado é uma implantação e um gerenciamento simplificados de data center de IA, ao mesmo tempo em que oferece escalabilidade praticamente ilimitada para desempenho e capacidade.

Visão geral da solução

A solução NetApp BeeGFS, alimentada pelos sistemas de armazenamento NetApp EF600 NVMe de alto desempenho e pelo sistema de arquivos paralelos BeeGFS escalável, oferece uma base de armazenamento robusta e eficiente para cargas de trabalho de IA exigentes. Sua arquitetura de disco compartilhado garante alta disponibilidade, mantendo desempenho e acessibilidade consistentes, mesmo diante de desafios do sistema. Esta solução fornece uma arquitetura escalável e flexível que pode ser personalizada para atender a diversos requisitos de armazenamento. Os clientes podem expandir facilmente o desempenho e a capacidade de armazenamento integrando blocos de construção de armazenamento adicionais para lidar até mesmo com as cargas de trabalho mais exigentes.

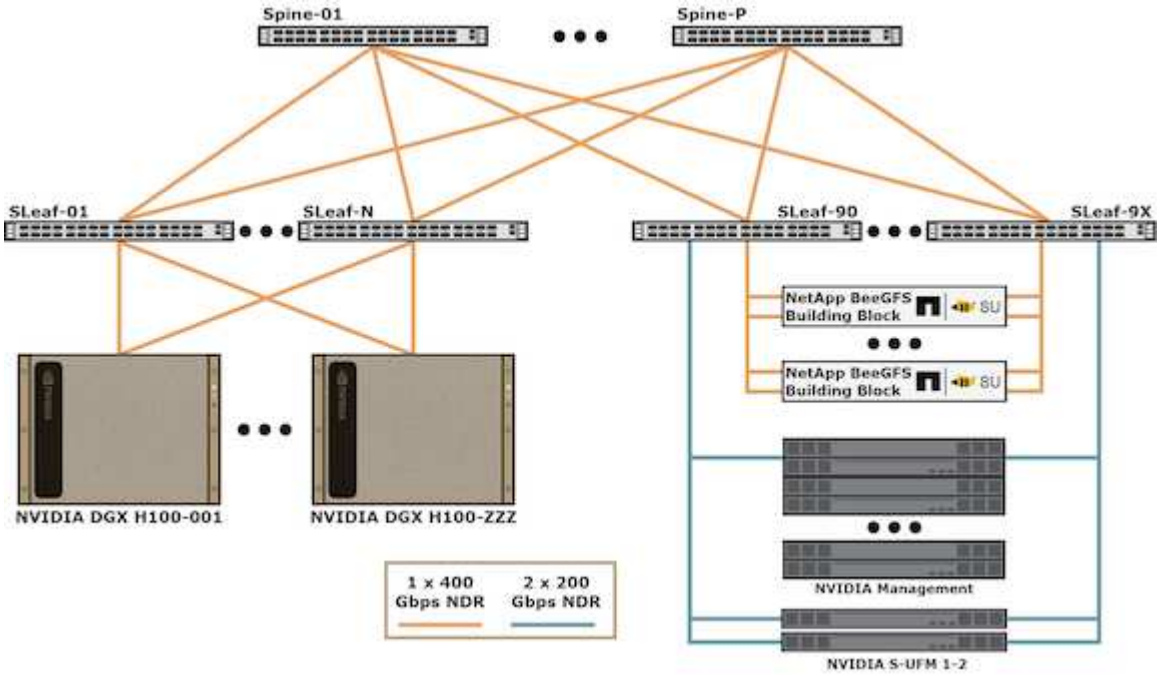
Tecnologia de soluções

- O NVIDIA DGX SuperPOD aproveita os sistemas DGX H100 e H200 com um armazenamento

compartilhado conectado externamente validado:

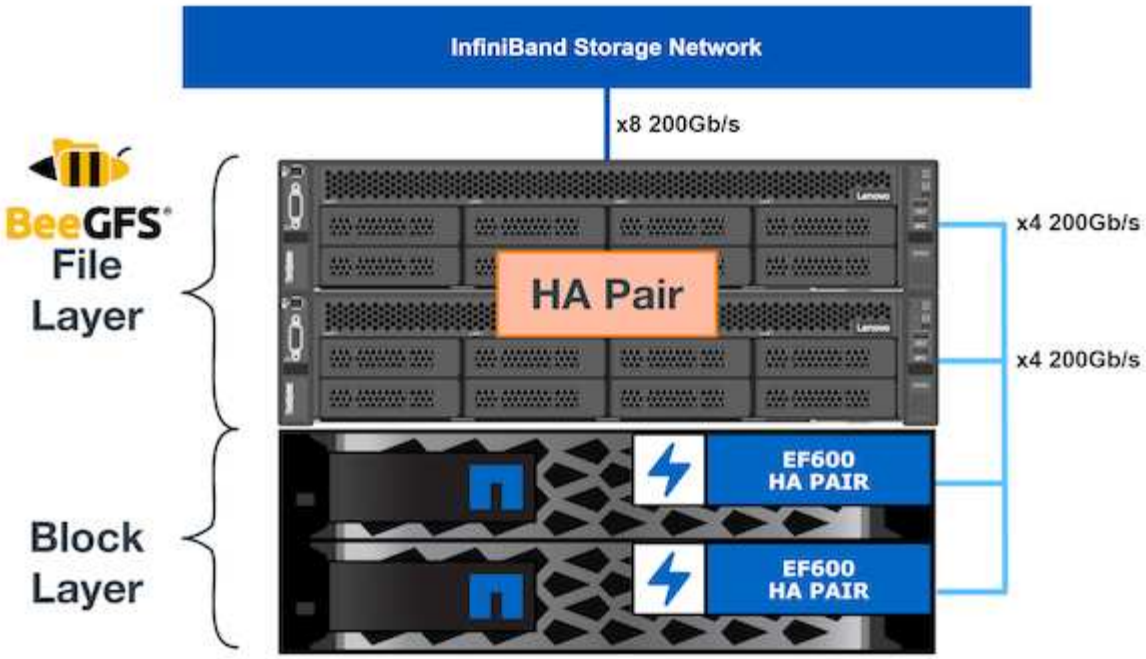
- Cada unidade escalável (SU) DGX SuperPOD consiste em 32 sistemas DGX e é capaz de 640 petaFLOPS de desempenho de IA com precisão FP8. A NetApp recomenda dimensionar a solução de armazenamento NetApp BeeGFS com pelo menos 2 blocos de construção para uma única configuração DGX SuperPOD.

Uma visão geral da solução



- Os blocos de construção do NetApp BeeGFS consistem em dois arrays NetApp EF600 e dois servidores x86:
  - Com os arrays all-flash NetApp EF600 na base do NVIDIA DGX SuperPOD, os clientes obtêm uma base de armazenamento confiável apoiada por seis 9s de tempo de atividade.
  - A camada do sistema de arquivos entre os sistemas NetApp EF600 e NVIDIA DGX é o sistema de arquivos paralelo BeeGFS. O BeeGFS foi criado pelo Centro Fraunhofer de Computação de Alto Desempenho na Alemanha para resolver os problemas dos sistemas de arquivos paralelos legados. O resultado é um sistema de arquivos com uma arquitetura moderna de espaço de usuário que agora é desenvolvido e entregue pela ThinkParQ e usado por muitos ambientes de supercomputação.
  - O suporte da NetApp para BeeGFS alinha a excelente organização de suporte da NetApp com os requisitos do cliente em termos de desempenho e tempo de atividade. Os clientes têm acesso a recursos de suporte superiores, acesso antecipado às versões do BeeGFS e acesso a recursos empresariais selecionados do BeeGFS, como aplicação de cotas e alta disponibilidade (HA).
- A combinação dos NVIDIA SuperPOD SUs e dos blocos de construção NetApp BeeGFS fornece uma solução de IA ágil na qual a computação ou o armazenamento são dimensionados de forma fácil e contínua.

Bloco de construção NetApp BeeGFS



Resumo do caso de uso

Esta solução se aplica aos seguintes casos de uso:

- Inteligência Artificial (IA), incluindo aprendizado de máquina (ML), aprendizado profundo (DL), processamento de linguagem natural (NLP), compreensão de linguagem natural (NLU) e IA generativa (GenAI).
- Treinamento de IA em média e grande escala
- Modelos de visão computacional, fala, áudio e linguagem
- HPC incluindo aplicações aceleradas por interface de passagem de mensagens (MPI) e outras técnicas de computação distribuída
- Cargas de trabalho de aplicativos caracterizadas pelo seguinte:
  - Ler ou escrever em arquivos maiores que 1 GB
  - Leitura ou gravação no mesmo arquivo por vários clientes (dezenas, centenas e milhares)
- Conjuntos de dados multiterabytes ou multipetabytes
- Ambientes que precisam de um único namespace de armazenamento otimizável para uma mistura de arquivos grandes e pequenos

Requisitos de tecnologia

Esta seção aborda os requisitos de tecnologia para a solução NVIDIA DGX SuperPOD com NetApp .

Requisitos de hardware

A Tabela 1 abaixo lista os componentes de hardware necessários para implementar a solução para uma única SU. O dimensionamento da solução começa com 32 sistemas NVIDIA DGX H100 e dois ou três blocos de construção NetApp BeeGFS. Um único bloco de construção NetApp BeeGFS consiste em dois arrays NetApp EF600 e dois servidores x86. Os clientes podem adicionar blocos de construção adicionais conforme o

tamanho da implantação aumenta. Para mais informações, consulte o "Arquitetura de referência NVIDIA DGX H100 SuperPOD" e "NVA-1164-DESIGN: BeeGFS no NetApp NVA Design" .

Hardware	Quantidade
NVIDIA DGX H100 ou H200	32
Switches NVIDIA Quantum QM9700	8 folhas, 4 lombadas
Blocos de construção NetApp BeeGFS	3

Requisitos de software

A Tabela 2 abaixo lista os componentes de software necessários para implementar a solução. Os componentes de software usados em qualquer implementação específica da solução podem variar com base nos requisitos do cliente.

Software
Pilha de software NVIDIA DGX
Gerenciador de comando básico da NVIDIA
Sistema de arquivos paralelos ThinkParQ BeeGFS

Verificação da solução

O NVIDIA DGX SuperPOD com NetApp foi validado em um cluster de aceitação dedicado na NVIDIA usando blocos de construção NetApp BeeGFS. Os critérios de aceitação foram baseados em uma série de testes de aplicação, desempenho e estresse realizados pela NVIDIA. Para mais informações, consulte o "NVIDIA DGX SuperPOD: Arquitetura de referência NetApp EF600 e BeeGFS" .

Conclusão

A NetApp e a NVIDIA têm um longo histórico de colaboração para fornecer um portfólio de soluções de IA ao mercado. O NVIDIA DGX SuperPOD com o array all-flash NetApp EF600 é uma solução comprovada e validada que os clientes podem implementar com confiança. Essa arquitetura totalmente integrada e pronta para uso elimina os riscos da implantação e coloca qualquer pessoa no caminho para vencer a corrida pela liderança em IA.

Onde encontrar informações adicionais

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

- "Arquitetura de referência NVIDIA DGX SuperPOD"
- "Guia de referência de design de data center NVIDIA DGX SuperPOD"
- "NVIDIA DGX SuperPOD: NetApp EF600 e BeeGFS"

Informações sobre direitos autorais

Copyright © 2025 NetApp, Inc. Todos os direitos reservados. Impresso nos EUA. Nenhuma parte deste documento protegida por direitos autorais pode ser reproduzida de qualquer forma ou por qualquer meio — gráfico, eletrônico ou mecânico, incluindo fotocópia, gravação, gravação em fita ou storage em um sistema de recuperação eletrônica — sem permissão prévia, por escrito, do proprietário dos direitos autorais.

O software derivado do material da NetApp protegido por direitos autorais está sujeito à seguinte licença e isenção de responsabilidade:

ESTE SOFTWARE É FORNECIDO PELA NETAPP "NO PRESENTE ESTADO" E SEM QUAISQUER GARANTIAS EXPRESSAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO E ADEQUAÇÃO A UM DETERMINADO PROPÓSITO, CONFORME A ISENÇÃO DE RESPONSABILIDADE DESTE DOCUMENTO. EM HIPÓTESE ALGUMA A NETAPP SERÁ RESPONSÁVEL POR QUALQUER DANO DIRETO, INDIRETO, INCIDENTAL, ESPECIAL, EXEMPLAR OU CONSEQUENCIAL (INCLUINDO, SEM LIMITAÇÕES, AQUISIÇÃO DE PRODUTOS OU SERVIÇOS SOBRESSALENTES; PERDA DE USO, DADOS OU LUCROS; OU INTERRUPÇÃO DOS NEGÓCIOS), INDEPENDENTEMENTE DA CAUSA E DO PRINCÍPIO DE RESPONSABILIDADE, SEJA EM CONTRATO, POR RESPONSABILIDADE OBJETIVA OU PREJUÍZO (INCLUINDO NEGLIGÊNCIA OU DE OUTRO MODO), RESULTANTE DO USO DESTE SOFTWARE, MESMO SE ADVERTIDA DA RESPONSABILIDADE DE TAL DANO.

A NetApp reserva-se o direito de alterar quaisquer produtos descritos neste documento, a qualquer momento e sem aviso. A NetApp não assume nenhuma responsabilidade nem obrigação decorrentes do uso dos produtos descritos neste documento, exceto conforme expressamente acordado por escrito pela NetApp. O uso ou a compra deste produto não representam uma licença sob quaisquer direitos de patente, direitos de marca comercial ou quaisquer outros direitos de propriedade intelectual da NetApp.

O produto descrito neste manual pode estar protegido por uma ou mais patentes dos EUA, patentes estrangeiras ou pedidos pendentes.

LEGENDA DE DIREITOS LIMITADOS: o uso, a duplicação ou a divulgação pelo governo estão sujeitos a restrições conforme estabelecido no subparágrafo (b)(3) dos Direitos em Dados Técnicos - Itens Não Comerciais no DFARS 252.227-7013 (fevereiro de 2014) e no FAR 52.227- 19 (dezembro de 2007).

Os dados aqui contidos pertencem a um produto comercial e/ou serviço comercial (conforme definido no FAR 2.101) e são de propriedade da NetApp, Inc. Todos os dados técnicos e software de computador da NetApp fornecidos sob este Contrato são de natureza comercial e desenvolvidos exclusivamente com despesas privadas. O Governo dos EUA tem uma licença mundial limitada, irrevogável, não exclusiva, intransferível e não sublicenciável para usar os Dados que estão relacionados apenas com o suporte e para cumprir os contratos governamentais desse país que determinam o fornecimento de tais Dados. Salvo disposição em contrário no presente documento, não é permitido usar, divulgar, reproduzir, modificar, executar ou exibir os dados sem a aprovação prévia por escrito da NetApp, Inc. Os direitos de licença pertencentes ao governo dos Estados Unidos para o Departamento de Defesa estão limitados aos direitos identificados na cláusula 252.227-7015(b) (fevereiro de 2014) do DFARS.

Informações sobre marcas comerciais

NETAPP, o logotipo NETAPP e as marcas listadas em http://www.netapp.com/TM são marcas comerciais da NetApp, Inc. Outros nomes de produtos e empresas podem ser marcas comerciais de seus respectivos proprietários.