



EF-Series AI con NVIDIA

NetApp Solutions

NetApp
November 12, 2024

Tabla de contenidos

- EF-Series AI con NVIDIA 1
- EF-Series AI de NetApp con NVIDIA 1
- NVIDIA DGX SuperPOD con NetApp: Guía de diseño 1
- DISEÑO NVA-1156: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS 5
- NVA-1156-PUESTA en MARCHA: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS ... 5
- TR-4755: Guía de puesta en marcha de BeeGFS con NetApp E-Series 5

EF-Series AI con NVIDIA

EF-Series AI de NetApp con NVIDIA

Descripción general de las soluciones de infraestructura convergente de EF-Series de NetApp y NVIDIA.

EF-Series AI con NVIDIA DGX H100 y H200 Systems y BeeGFS

- ["Guía de diseño"](#)
- ["Guía de puesta en marcha"](#)
- ["Guía de puesta en marcha de BeeGFS en NetApp"](#)

NVIDIA DGX SuperPOD con NetApp: Guía de diseño

Esta arquitectura verificada de NetApp describe el diseño del NVIDIA DGX SuperPOD con elementos básicos NetApp® BeeGFS®. Esta solución es una plataforma de centro de datos de pila completa validada en un clúster de aceptación dedicado de NVIDIA.



Amina Bennani, Christian Whiteside, David Arnette y Sathish Thyagarajan, NetApp

Resumen ejecutivo

La IA está revolucionando las experiencias de los consumidores y impulsando la innovación en todos los sectores, en el panorama tecnológico actual que evoluciona a gran velocidad. No obstante, también presenta retos a los departamentos DE TECNOLOGÍA, que están sometidos a presión de poner en marcha soluciones de computación de alto rendimiento (HPC) capaces de manejar las exigentes demandas de las cargas de trabajo de la IA. A medida que las organizaciones compiten por aprovechar la potencia de la IA, crece la urgencia de encontrar una solución fácil de poner en marcha, escalar y gestionar.

DGX SuperPOD de NVIDIA es una plataforma de infraestructura de centro de datos de IA que se ofrece como solución lista para usar para QUE LOS DEPARTAMENTOS DE tecnología den soporte a las cargas de trabajo de IA más complejas que enfrentan las empresas actuales. En el centro de cualquier modelo de aprendizaje profundo (DL) preciso se encuentran grandes volúmenes de datos, por lo que se requiere una solución de almacenamiento de alto rendimiento que pueda servir y volver a servir estos datos de forma eficiente. La solución BeeGFS de NetApp, que consiste en cabinas de almacenamiento EF600 de NetApp con el sistema de archivos paralelo BeeGFS, permite a DGX SuperPOD de NVIDIA aprovechar toda su capacidad. La solución BeeGFS de NetApp ha sido validada por NVIDIA para integrarse y escalar con la arquitectura SuperPOD. Esto se traduce en una puesta en marcha y gestión simplificadas del centro de datos de IA y en una escalabilidad prácticamente ilimitada del rendimiento y la capacidad.

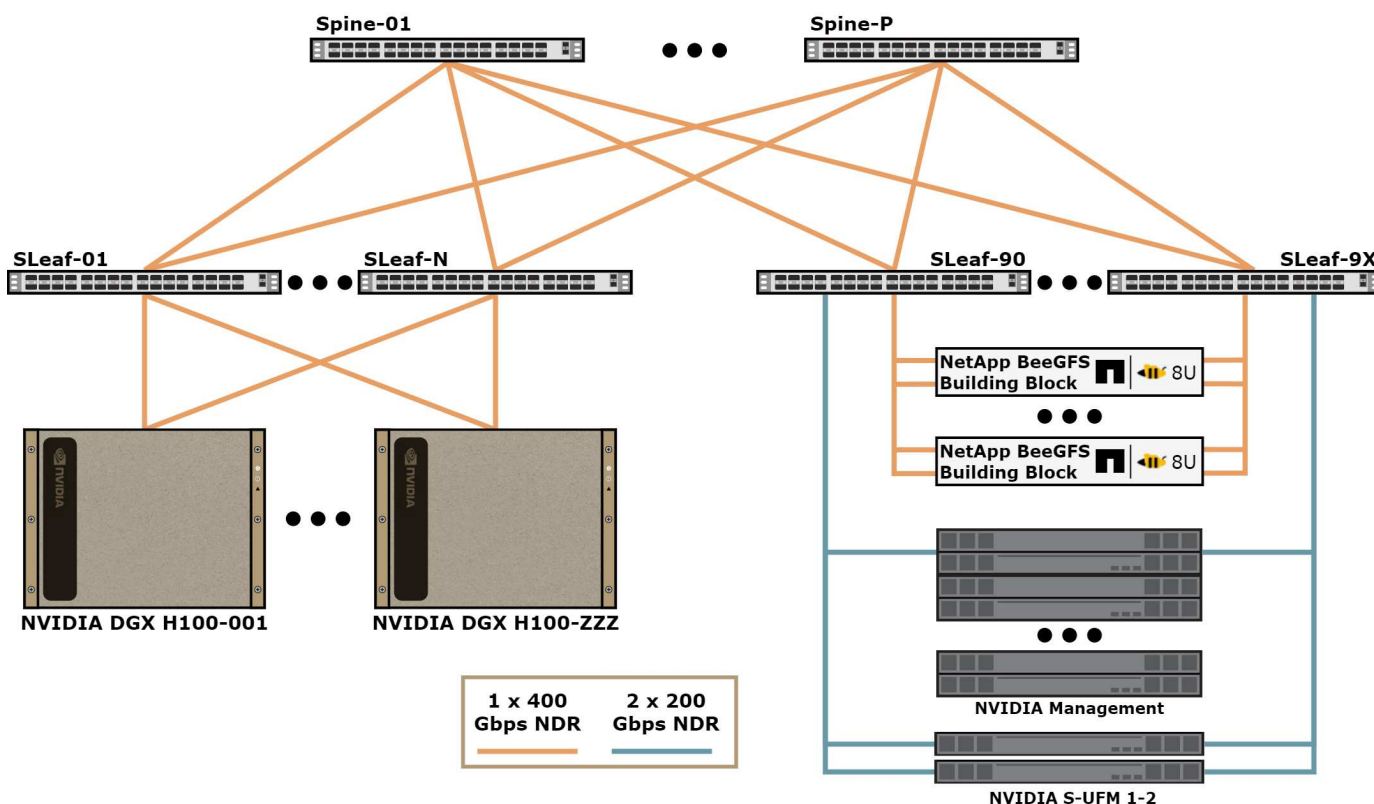
Descripción general de la solución

La solución BeeGFS de NetApp, impulsada por los sistemas de almacenamiento NVMe EF600 de NetApp de alto rendimiento y el sistema de archivos paralelo BeeGFS escalable, ofrece una base de almacenamiento robusta y eficiente para cargas de trabajo de IA exigentes. Su arquitectura de discos compartidos garantiza una alta disponibilidad, manteniendo un rendimiento y una accesibilidad uniformes, incluso cuando se presentan desafíos del sistema. Esta solución proporciona una arquitectura escalable y flexible que puede personalizarse para satisfacer diversos requisitos de almacenamiento. Los clientes pueden expandir fácilmente el rendimiento y la capacidad de su almacenamiento integrando elementos básicos de almacenamiento adicionales para gestionar incluso las cargas de trabajo más exigentes.

Tecnología de soluciones

- DGX SuperPOD de NVIDIA aprovecha los sistemas DGX H100 y H200 con un almacenamiento compartido validado externamente:
 - Cada unidad escalable (SU) de DGX SuperPOD consta de 32 sistemas DGX y es capaz de 640 petaflops de rendimiento de IA con precisión de FP8 PB. NetApp recomienda ajustar el tamaño de la solución de almacenamiento BeeGFS de NetApp con al menos 2 elementos básicos para una única configuración de DGX SuperPOD.

Una vista de alto nivel de la solución

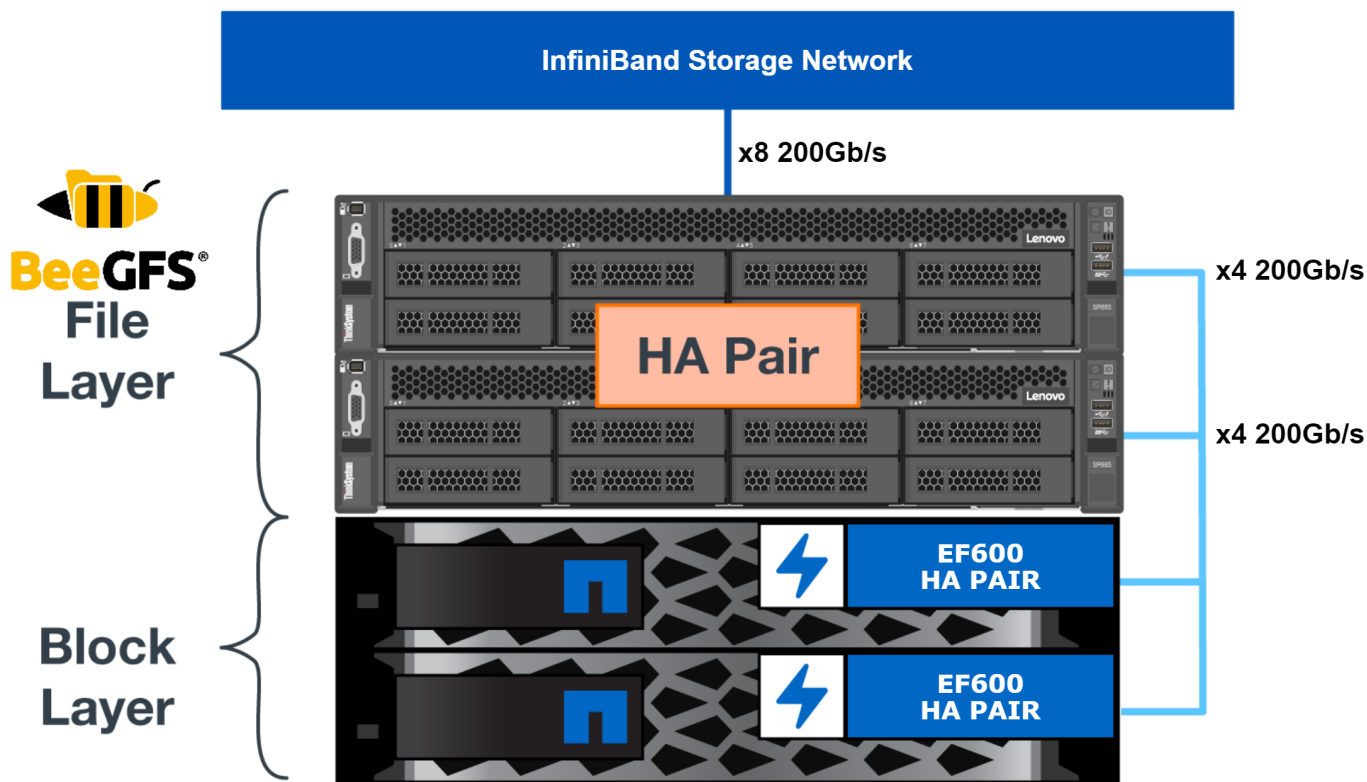


- Los elementos básicos de BeeGFS de NetApp constan de dos cabinas EF600 de NetApp y dos servidores x86:
 - Con las cabinas all-flash EF600 de NetApp como base de NVIDIA DGX SuperPOD, los clientes obtienen una base de almacenamiento fiable respaldada por un tiempo de actividad de seis 9s ms.
 - La capa del sistema de archivos entre NetApp EF600 y los sistemas NVIDIA DGX es el sistema de archivos paralelo BeeGFS. BeeGFS fue creado en el Centro Fraunhofer de Computación de Alto Rendimiento en Alemania para dar solución a los puntos débiles de los sistemas de archivos paralelos

heredados. El resultado es un sistema de archivos con una arquitectura moderna de espacio del usuario que ahora desarrolla y distribuye ThinkParQ, y al que recurren muchos entornos de supercomputación.

- El soporte de NetApp para BeeGFS alinea la excelente organización de soporte de NetApp con los requisitos de rendimiento y tiempo de actividad del cliente. Los clientes obtienen acceso a recursos de soporte superiores, acceso previo a lanzamientos de BeeGFS y acceso a funciones empresariales selectas de BeeGFS, como el cumplimiento de cuotas y alta disponibilidad.
- La combinación de los elementos básicos NVIDIA SuperPOD SUS y BeeGFS de NetApp proporciona una solución de IA ágil en la que la computación o el almacenamiento se pueden escalar de forma fácil y fluida.

NetApp BeeGFS Building Block



Resumen de casos de uso

Esta solución se aplica a los siguientes casos de uso:

- Inteligencia artificial (IA) incluido el aprendizaje automático (ML), el aprendizaje profundo (DL), el procesamiento del lenguaje natural (NLP), el conocimiento del lenguaje natural (NLU) y la IA generativa (GenAI).
- Formación de IA a escala media y grande
- Modelos de visión computarizada, habla, audio y lenguaje
- HPC que incluye aplicaciones aceleradas mediante la interfaz de paso de mensajes (MPI) y otras técnicas informáticas distribuidas
- Cargas de trabajo de aplicaciones que se caracterizan por las siguientes características:
 - Leer o escribir en archivos de más de 1 GB

- Leyendo o escribiendo en el mismo archivo por varios clientes (10s, 100s y 1000s).
- Conjuntos de datos de varios terabytes o varios petabytes
- Entornos que requieren un único espacio de nombres de almacenamiento optimizable para una combinación de archivos grandes y pequeños

Requisitos tecnológicos

En esta sección se tratan los requisitos tecnológicos de la solución NVIDIA DGX SuperPOD con NetApp.

Requisitos de hardware

La tabla 1 que aparece a continuación enumera los componentes de hardware necesarios para implementar la solución para una única SU. El ajuste de tamaño de la solución comienza con 32 sistemas NVIDIA DGX H100 y dos o tres elementos básicos BeeGFS de NetApp.

Un único elemento básico BeeGFS de NetApp consta de dos cabinas EF600 de NetApp y dos servidores x86. Los clientes pueden agregar elementos básicos adicionales a medida que aumenta el tamaño de la puesta en marcha. Para obtener más información, consulte "[Arquitectura de referencia NVIDIA DGX H100 SuperPOD](#)" y.. "[NVA-1164-DESIGN: BeeGFS en diseño NVA de NetApp](#)".

Hardware subyacente	Cantidad
NVIDIA DGX H100 o H200	32
Switches NVIDIA Quantum QM9700	8 hoja, 4 espina
Elementos básicos BeeGFS de NetApp	3

Requisitos de software

En la tabla 2 que aparece a continuación se enumeran los componentes de software necesarios para implementar la solución. Los componentes que se usan en cualquier implementación particular de la solución pueden variar en función de las necesidades del cliente.

De NetApp
Pila de software NVIDIA DGX
Administrador de comandos base de NVIDIA
Sistema de archivos paralelo BeeGFS de ThinkParQ

Verificación de la solución

NVIDIA DGX SuperPOD con NetApp ha sido validado en un clúster de aceptación dedicado de NVIDIA empleando los elementos básicos BeeGFS de NetApp. Los criterios de aceptación se basaron en una serie de pruebas de aplicación, rendimiento y estrés realizadas por NVIDIA. Para obtener más información, consulte "[NVIDIA DGX SuperPOD: Arquitectura de referencia de NetApp EF600 y BeeGFS](#)".

Conclusión

NetApp y NVIDIA llevan mucho tiempo colaborando para ofrecer una cartera de soluciones de inteligencia artificial al mercado. NVIDIA DGX SuperPOD con la cabina all-flash EF600 de NetApp es una solución demostrada y validada que los clientes pueden poner en marcha con total confianza. Su arquitectura, totalmente integrada y lista para usar, acaba con los riesgos de la puesta en marcha y permite que cualquiera

pueda ganar terreno en el liderazgo de la IA.

Dónde encontrar información adicional

Si quiere más información sobre el contenido de este documento, consulte los siguientes documentos o sitios web:

- ["Arquitectura de referencia de DGX SuperPOD de NVIDIA"](#)
- ["Guía de referencia de diseño del centro de datos de NVIDIA DGX SuperPOD"](#)
- ["DGX SuperPOD de NVIDIA: NetApp EF600 y BeeGFS"](#)

DISEÑO NVA-1156: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS

Abdel Sadek, Tim Chau, Joe McCormick y David Arnette, NetApp

El DISEÑO de NVA-1156 describe una arquitectura verificada de NetApp para cargas de trabajo de aprendizaje automático (ML) y de inteligencia artificial (IA) mediante los sistemas de almacenamiento EF600 de NetApp, el sistema de archivos en paralelo BeeGFS, los sistemas NVIDIA DGX A100 y los switches IB de NVIDIA Quantum QM8700 de 200 Gbps. Este diseño incluye InfiniBand (IB) de 200 Gbps para la estructura de interconexión de clústeres de almacenamiento e informáticos con el fin de proporcionar a los clientes una arquitectura completamente basada en IB para las cargas de trabajo de alto rendimiento. Este documento también incluye los resultados de las pruebas de rendimiento para la arquitectura tal y como se ha implementado.

["DISEÑO NVA-1156: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS"](#)

NVA-1156-PUESTA en MARCHA: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS

Abdel Sadek, Tim Chau, Joe McCormick y David Arnette, NetApp

En este documento se describe una arquitectura verificada de NetApp para cargas de trabajo de aprendizaje automático (ML) e inteligencia artificial (IA) mediante los sistemas de almacenamiento EF600 NVMe de NetApp, el sistema de archivos en paralelo ThinkParQ BeeGFS, los sistemas NVIDIA DGX A100 y los switches Mellanox Quantum QM8700 InfiniBand (IB) de 200 Gbps. También se incluyen instrucciones para ejecutar pruebas de referencia de validación una vez completada la implementación.

["NVA-1156-PUESTA en MARCHA: EF-Series AI de NetApp con sistemas NVIDIA DGX A100 y BeeGFS"](#)

TR-4755: Guía de puesta en marcha de BeeGFS con NetApp E-Series

BeeGFS basado en NetApp con almacenamiento E-Series es una solución demostrada e

integrada con una infraestructura de informática de alto rendimiento sencilla, fiable, escalable y rentable que se adapta al ritmo de sus cargas de trabajo más extremas.

["Guía de puesta en marcha de BeeGFS con E-Series de NetApp"](#)

Información de copyright

Copyright © 2024 NetApp, Inc. Todos los derechos reservados. Imprimido en EE. UU. No se puede reproducir este documento protegido por copyright ni parte del mismo de ninguna forma ni por ningún medio (gráfico, electrónico o mecánico, incluidas fotocopias, grabaciones o almacenamiento en un sistema de recuperación electrónico) sin la autorización previa y por escrito del propietario del copyright.

El software derivado del material de NetApp con copyright está sujeto a la siguiente licencia y exención de responsabilidad:

ESTE SOFTWARE LO PROPORCIONA NETAPP «TAL CUAL» Y SIN NINGUNA GARANTÍA EXPRESA O IMPLÍCITA, INCLUYENDO, SIN LIMITAR, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO, CUYA RESPONSABILIDAD QUEDA EXIMIDA POR EL PRESENTE DOCUMENTO. EN NINGÚN CASO NETAPP SERÁ RESPONSABLE DE NINGÚN DAÑO DIRECTO, INDIRECTO, ESPECIAL, EJEMPLAR O RESULTANTE (INCLUYENDO, ENTRE OTROS, LA OBTENCIÓN DE BIENES O SERVICIOS SUSTITUTIVOS, PÉRDIDA DE USO, DE DATOS O DE BENEFICIOS, O INTERRUPCIÓN DE LA ACTIVIDAD EMPRESARIAL) CUALQUIERA SEA EL MODO EN EL QUE SE PRODUJERON Y LA TEORÍA DE RESPONSABILIDAD QUE SE APLIQUE, YA SEA EN CONTRATO, RESPONSABILIDAD OBJETIVA O AGRAVIO (INCLUIDA LA NEGLIGENCIA U OTRO TIPO), QUE SURJAN DE ALGÚN MODO DEL USO DE ESTE SOFTWARE, INCLUSO SI HUBIEREN SIDO ADVERTIDOS DE LA POSIBILIDAD DE TALES DAÑOS.

NetApp se reserva el derecho de modificar cualquiera de los productos aquí descritos en cualquier momento y sin aviso previo. NetApp no asume ningún tipo de responsabilidad que surja del uso de los productos aquí descritos, excepto aquello expresamente acordado por escrito por parte de NetApp. El uso o adquisición de este producto no lleva implícita ninguna licencia con derechos de patente, de marcas comerciales o cualquier otro derecho de propiedad intelectual de NetApp.

Es posible que el producto que se describe en este manual esté protegido por una o más patentes de EE. UU., patentes extranjeras o solicitudes pendientes.

LEYENDA DE DERECHOS LIMITADOS: el uso, la copia o la divulgación por parte del gobierno están sujetos a las restricciones establecidas en el subpárrafo (b)(3) de los derechos de datos técnicos y productos no comerciales de DFARS 252.227-7013 (FEB de 2014) y FAR 52.227-19 (DIC de 2007).

Los datos aquí contenidos pertenecen a un producto comercial o servicio comercial (como se define en FAR 2.101) y son propiedad de NetApp, Inc. Todos los datos técnicos y el software informático de NetApp que se proporcionan en este Acuerdo tienen una naturaleza comercial y se han desarrollado exclusivamente con fondos privados. El Gobierno de EE. UU. tiene una licencia limitada, irrevocable, no exclusiva, no transferible, no sublicenciable y de alcance mundial para utilizar los Datos en relación con el contrato del Gobierno de los Estados Unidos bajo el cual se proporcionaron los Datos. Excepto que aquí se disponga lo contrario, los Datos no se pueden utilizar, desvelar, reproducir, modificar, interpretar o mostrar sin la previa aprobación por escrito de NetApp, Inc. Los derechos de licencia del Gobierno de los Estados Unidos de América y su Departamento de Defensa se limitan a los derechos identificados en la cláusula 252.227-7015(b) de la sección DFARS (FEB de 2014).

Información de la marca comercial

NETAPP, el logotipo de NETAPP y las marcas que constan en <http://www.netapp.com/TM> son marcas comerciales de NetApp, Inc. El resto de nombres de empresa y de producto pueden ser marcas comerciales de sus respectivos propietarios.