



AIPod de NetApp con sistemas NVIDIA DGX

NetApp Solutions

NetApp
April 12, 2024

Tabla de contenidos

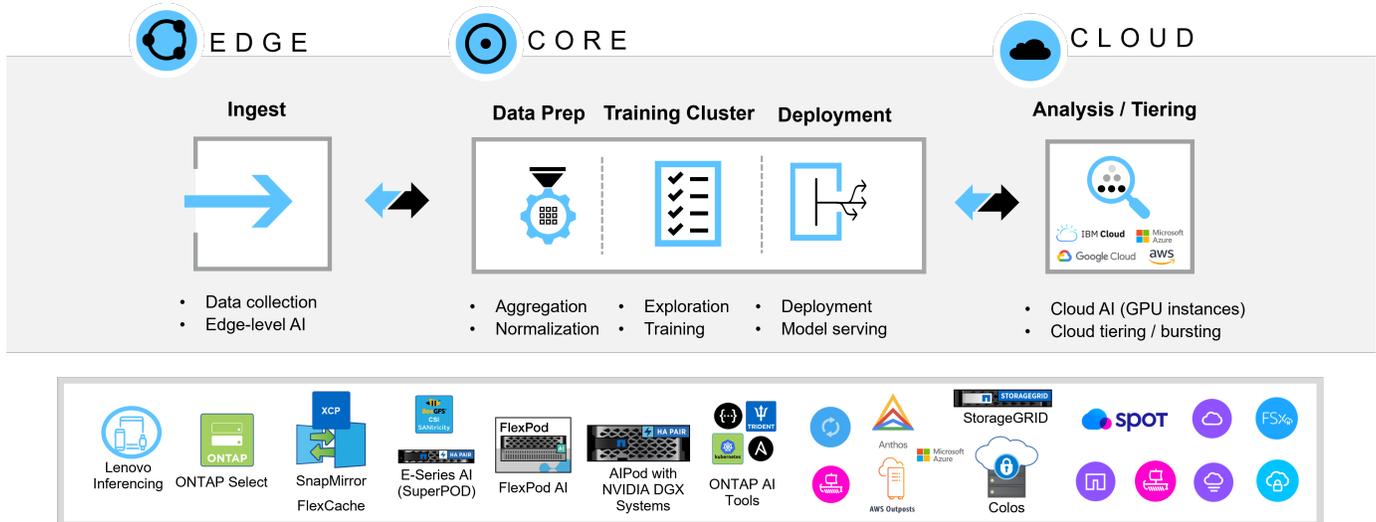
- NetApp AIPod con sistemas NVIDIA DGX: Introducción 1
 - NetApp AIPod con sistemas NVIDIA DGX: Introducción 1
 - AIPod de NetApp con sistemas NVIDIA DGX: Componentes de hardware 2
 - AIPod de NetApp con sistemas NVIDIA DGX: Componentes de software 6
- NetApp AIPod con sistemas NVIDIA DGX: Arquitectura de la solución 10
- AIPod de NetApp con sistemas NVIDIA DGX: Validación de la solución y guía de tamaño 13
- AIPod de NetApp con sistemas NVIDIA DGX: Conclusión e información adicional 15

NetApp AIPOd con sistemas NVIDIA DGX: Introducción

Ingeniería de soluciones de NetApp

NetApp™ AIPOd con NVIDIA DGX™ los sistemas y los sistemas de almacenamiento conectados al cloud de NetApp simplifican las puestas en marcha de infraestructuras para cargas de trabajo de aprendizaje automático (ML) e inteligencia artificial (IA) al eliminar complejidades y conjeturas en la fase de diseño. Basándose en el diseño de NVIDIA DGX BasePOD para proporcionar un rendimiento informático excepcional para cargas de trabajo de nueva generación, AIPOd con sistemas NVIDIA DGX añade sistemas de almacenamiento NetApp AFF que permiten a los clientes empezar poco a poco y crecer sin interrupciones, a la vez que gestionan de forma inteligente los datos desde el perímetro al núcleo y al cloud, y viceversa. NetApp AIPOd forma parte de la cartera más amplia de soluciones de IA de NetApp, muestra la figura a continuación:

Cartera de soluciones de IA de NetApp



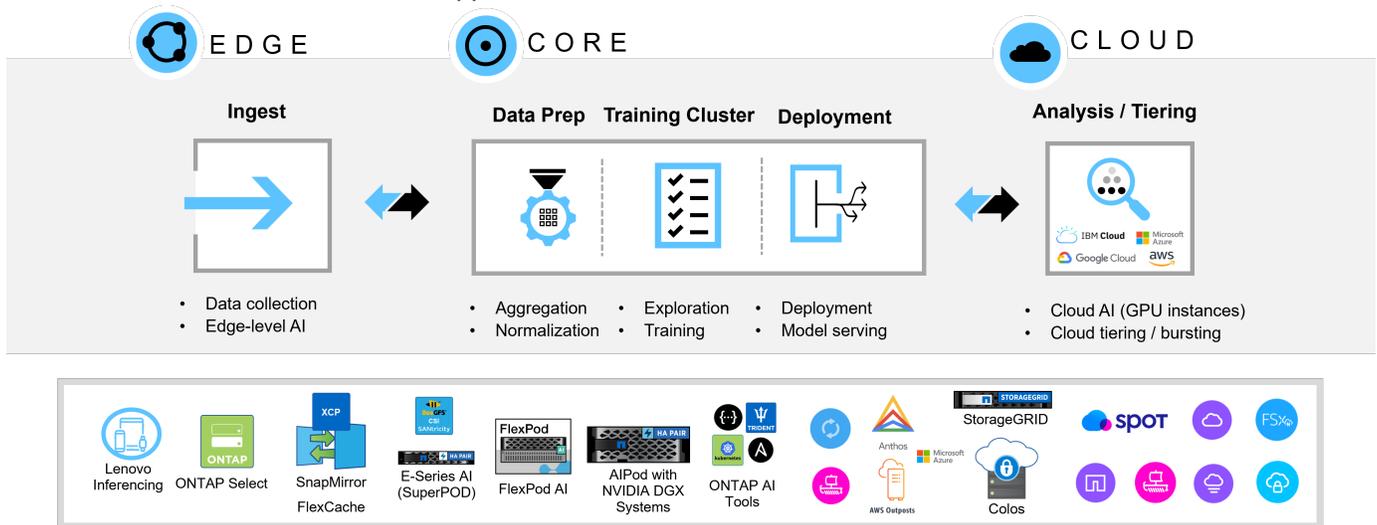
En este documento se describen los componentes clave de la arquitectura de referencia AIPOd, la información sobre conectividad del sistema y la guía para dimensionar la solución. Este documento está destinado a ingenieros de soluciones de NetApp y partners, y para aquellos clientes interesados en poner en marcha una infraestructura de alto rendimiento para cargas de trabajo de análisis y ML/DL.

["Siguiendo: AIPOd de NetApp con sistemas NVIDIA DGX: Componentes de hardware"](#)

NetApp AIPOd con sistemas NVIDIA DGX: Introducción

Ingeniería de soluciones de NetApp

NetApp™ AIPOd con NVIDIA DGX™ los sistemas y los sistemas de almacenamiento conectados al cloud de NetApp simplifican las puestas en marcha de infraestructuras para cargas de trabajo de aprendizaje automático (ML) e inteligencia artificial (IA) al eliminar complejidades y conjeturas en la fase de diseño. Basándose en el diseño de NVIDIA DGX BasePOD para proporcionar un rendimiento informático excepcional para cargas de trabajo de nueva generación, AIPOd con sistemas NVIDIA DGX añade sistemas de almacenamiento NetApp AFF que permiten a los clientes empezar poco a poco y crecer sin interrupciones, a la vez que gestionan de forma inteligente los datos desde el perímetro al núcleo y al cloud, y viceversa. NetApp AIPOd forma parte de la cartera más amplia de soluciones de IA de NetApp, muestra la figura a continuación:



En este documento se describen los componentes clave de la arquitectura de referencia AIPod, la información sobre conectividad del sistema y la guía para dimensionar la solución. Este documento está destinado a ingenieros de soluciones de NetApp y partners, y para aquellos clientes interesados en poner en marcha una infraestructura de alto rendimiento para cargas de trabajo de análisis y ML/DL.

["Siguiente: AIPod de NetApp con sistemas NVIDIA DGX: Componentes de hardware"](#)

AIPod de NetApp con sistemas NVIDIA DGX: Componentes de hardware

["Anterior: AIPod de NetApp con sistemas NVIDIA DGX: Introducción"](#)

Sistemas de almacenamiento de NetApp AFF

Con los vanguardistas sistemas de almacenamiento de NetApp AFF, los departamentos TECNOLÓGICOS pueden satisfacer los requisitos de almacenamiento del negocio con un rendimiento de primera clase, una flexibilidad superior, integración con el cloud y la mejor gestión de datos. Los sistemas AFF han sido diseñados específicamente para flash y ayudan a acelerar, gestionar y proteger los datos esenciales para la empresa.

Sistemas de almacenamiento A900 de AFF

La serie NetApp AFF A900, impulsada por el software para la gestión de datos NetApp ONTAP, ofrece protección de datos integrada, funcionalidades opcionales anti-ransomware y el alto rendimiento y la resiliencia necesarios para dar soporte a las cargas de trabajo empresariales más importantes. Acaba con las interrupciones de las operaciones esenciales para la empresa, reduce al máximo el ajuste del rendimiento y protege sus datos de los ataques de ransomware. Proporciona:

- Rendimiento líder en el sector
- Seguridad de los datos sin concesiones
- Actualizaciones simplificadas y sin interrupciones

Sistema de almacenamiento A900 de NetApp AFF



Rendimiento líder del sector

El AFF A900 gestiona con facilidad las cargas de trabajo de última generación, como el aprendizaje profundo, la IA y los análisis de alta velocidad, así como las bases de datos empresariales tradicionales como Oracle, SAP HANA, Microsoft SQL Server y aplicaciones virtualizadas. Mantiene las aplicaciones vitales para el negocio ejecutándose a máxima velocidad con hasta 2,4m 000 IOPS por pareja de alta disponibilidad y una latencia de hasta 100µs, y aumenta el rendimiento hasta en un 50 % con respecto a los modelos de NetApp anteriores. Con NFS a través de RDMA, pNFS y Session Trunking, los clientes pueden lograr el alto nivel de rendimiento de red necesario para las aplicaciones de nueva generación usando la infraestructura de red existente del centro de datos.

Además, los clientes pueden escalar y crecer con la compatibilidad multiprotocolo unificada para el almacenamiento SAN, NAS y de objetos y ofrecer la máxima flexibilidad con un único software de gestión de datos ONTAP unificado, tanto para los datos locales como en el cloud. Además, el estado del sistema se puede optimizar con los análisis predictivos basados en IA que ofrecen Active IQ y Cloud Insights.

Seguridad de los datos garantizada

Los sistemas A900 de AFF contienen una suite completa del software de NetApp para la protección de datos consistente con las aplicaciones e integrado. Ofrece una protección de datos integrada y soluciones innovadoras anti-ransomware que facilitan la recuperación previa y posterior al ataque. Los archivos maliciosos pueden bloquearse para que no se escriban en el disco, y las anomalías de almacenamiento se supervisan fácilmente para obtener información.

Actualizaciones simplificadas y fluidas

La AFF A900 está disponible como una actualización fluida e integrada para los clientes actuales de A700.

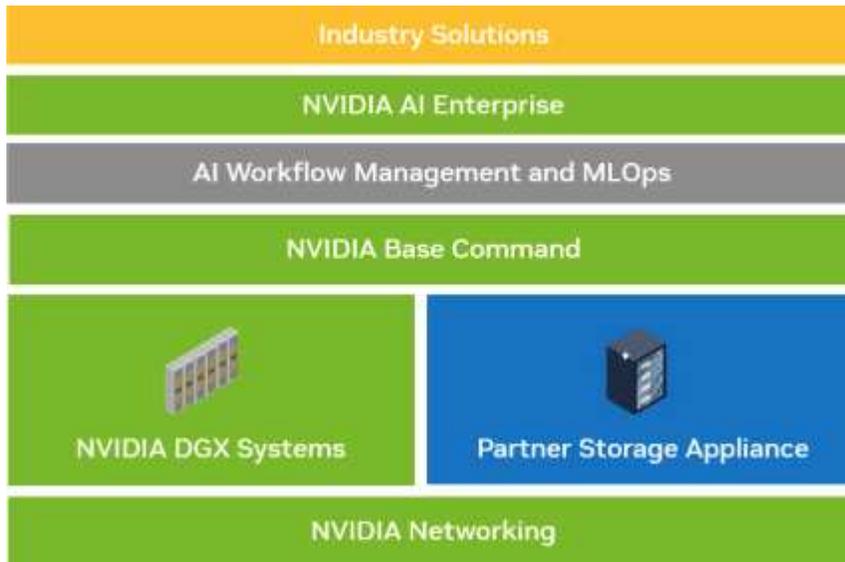
NetApp facilita la actualización y la eliminación de las interrupciones de las operaciones vitales para la empresa gracias a nuestras funciones avanzadas de fiabilidad, disponibilidad, aptitud y gestión (RASM). Además, NetApp aumenta aún más la eficiencia operativa y simplifica las actividades cotidianas de los equipos TECNOLÓGICOS, ya que el software ONTAP aplica automáticamente las actualizaciones de firmware de todos los componentes del sistema.

Para las mayores puestas en marcha, los sistemas AFF A900 ofrecen las opciones más elevadas de rendimiento y capacidad, mientras que otros sistemas de almacenamiento de NetApp, como AFF A800, AFF C800, AFF A400, AFF C400 y AFF A250 ofrecen opciones para puestas en marcha de menor tamaño a puntos de coste.

DGX BasePOD de NVIDIA

NVIDIA DGX BasePOD es una solución integrada que consta de componentes de hardware y software de NVIDIA, soluciones de MLOps y almacenamiento de terceros. Al aprovechar las mejores prácticas del diseño de sistemas de escalado horizontal con productos NVIDIA y soluciones de partners validadas, los clientes pueden implementar una plataforma eficiente y gestionable para el desarrollo de la IA. La figura 1 destaca los distintos componentes de NVIDIA DGX BasePOD.

Solución NVIDIA DGX BasePOD



SISTEMAS DGX H100 DE NVIDIA

El sistema NVIDIA DGX H100™ es el centro neurálgico de la IA que se acelera gracias al innovador rendimiento de la GPU Tensor Core de NVIDIA H100.

SISTEMA NVIDIA DGX H100

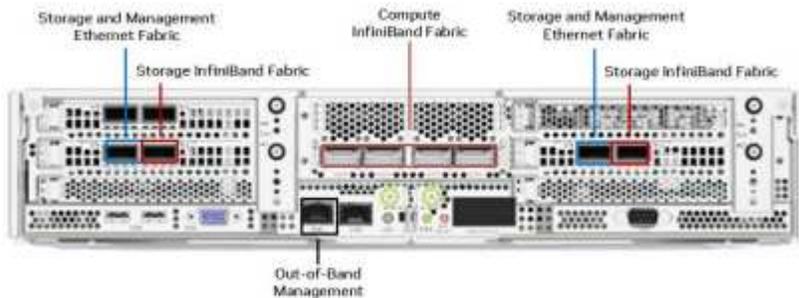


Las especificaciones clave del sistema DGX H100 son:

- Ocho GPU NVIDIA H100.
- 80 GB de memoria GPU por GPU, para un total de 640GB.
- Cuatro chips NVIDIA NVSwitch™.
- Procesadores Intel® Xeon® Platinum 8480 duales de 56 núcleos con soporte PCIe 5,0.
- 2 TB de memoria del sistema DDR5.
- Cuatro puertos OSFP que sirven a ocho adaptadores NVIDIA ConnectX-7 (InfiniBand/Ethernet) de un puerto y dos adaptadores NVIDIA ConnectX-7 (InfiniBand/Ethernet) de doble puerto.
- Dos unidades NVMe M,2 de 1,92 TB para el sistema operativo DGX y ocho unidades NVMe U,2 de 3,84 TB para almacenamiento/caché.
- 10,2 kW de potencia máxima.

A continuación, se muestran los puertos posteriores de la bandeja de CPU DGX H100. Cuatro de los puertos OSFP sirven a ocho adaptadores ConnectX-7 para la estructura de computación InfiniBand. Cada par de adaptadores ConnectX-7 de doble puerto proporciona vías paralelas para los tejidos de almacenamiento y gestión. El puerto fuera de banda se utiliza para el acceso BMC.

Panel posterior NVIDIA DGX H100



Conexión a redes NVIDIA

Switch NVIDIA Quantum-2 QM9700

Conmutador InfiniBand NVIDIA Quantum-2 QM9700



Los switches NVIDIA Quantum-2 QM9700 con conectividad InfiniBand de 400GB Gb/s potencian la estructura de computación en las configuraciones InfiniBand BasePOD de NVIDIA Quantum-2. Se utilizan adaptadores de un solo puerto ConnectX-7 para la estructura de computación InfiniBand. Cada sistema NVIDIA DGX tiene conexiones dobles a cada switch de QM9700 GbE, lo que proporciona múltiples rutas de alto ancho de banda y baja latencia entre los sistemas.

Conmutador NVIDIA Spectrum-3 SN4600

Conmutador NVIDIA Spectrum-3 SN4600



Los switches NVIDIA Spectrum-3 SN4600 ofrecen 128 puertos en total (64 por switch) con el fin de proporcionar conectividad redundante para la gestión en banda del DGX BasePOD. El switch NVIDIA SN4600 puede proporcionar velocidades de entre 1 GbE y 200 GbE. Para los dispositivos de almacenamiento

conectados con Ethernet, también se utilizan los switches NVIDIA SN4600. Los puertos de los adaptadores ConnectX-7 de doble puerto NVIDIA DGX se utilizan para la conectividad de almacenamiento y gestión en banda.

Conmutador NVIDIA Spectrum SN2201

Conmutador NVIDIA Spectrum SN2201



Los switches NVIDIA Spectrum SN2201 ofrecen puertos 48 GbE para proporcionar conectividad para la gestión fuera de banda. La gestión fuera de banda proporciona conectividad de gestión consolidada para todos los componentes de DGX BasePOD.

Adaptador NVIDIA ConnectX-7

Adaptador NVIDIA ConnectX-7



El adaptador NVIDIA ConnectX-7 puede proporcionar 25/50/100/200/400G de rendimiento. Los sistemas NVIDIA DGX utilizan los adaptadores ConnectX-7 de puerto único o doble para proporcionar flexibilidad en puestas en marcha de DGX BasePOD con InfiniBand de 400GB Gb/s y Ethernet 100/200GB Gb.

["Siguiente artículo: AIPod de NetApp con sistemas NVIDIA DGX: Componentes de software"](#)

AIPod de NetApp con sistemas NVIDIA DGX: Componentes de software

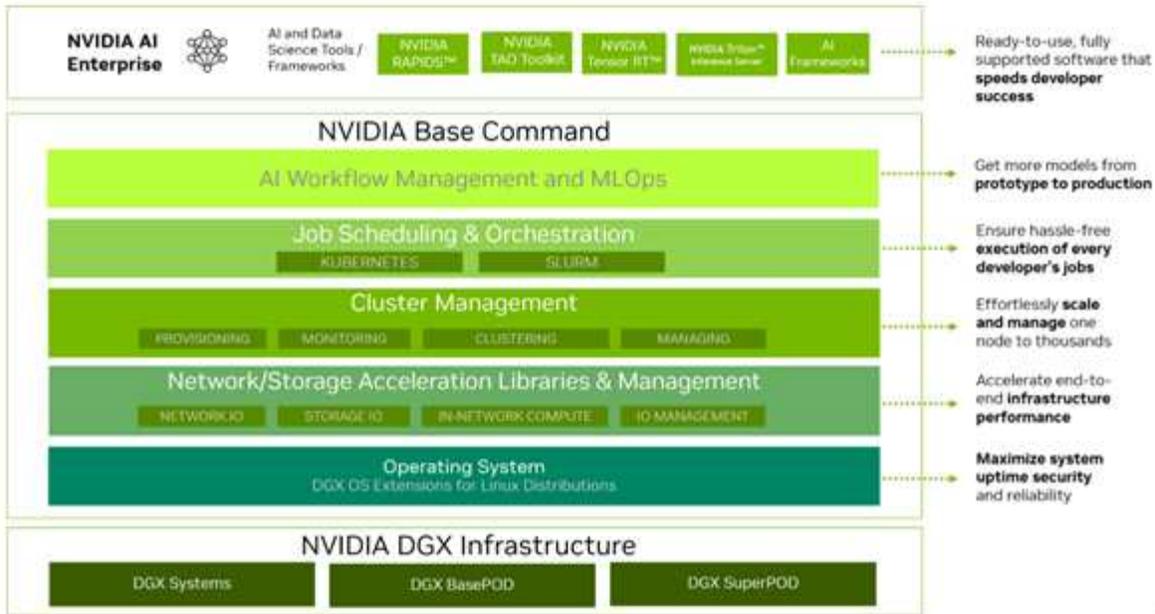
["Anterior: AIPod de NetApp con sistemas NVIDIA DGX: Componentes de hardware"](#)

Software NVIDIA

Comando base NVIDIA

NVIDIA Base Command™ impulsa todos los DGX BasePOD, lo que permite a las organizaciones aprovechar lo mejor de la innovación en software de NVIDIA. Las empresas pueden liberar todo el potencial de su inversión con una plataforma contrastada que incluye orquestación y gestión de clústeres de clase empresarial, bibliotecas que aceleran la computación, la infraestructura de redes y almacenamiento, y un sistema operativo (SO) optimizado para cargas de trabajo de IA.

Solución NVIDIA BaseCommand



GPU CLOUD DE NVIDIA (NGC)

NVIDIA NGC™ proporciona software para satisfacer las necesidades de los científicos de datos, desarrolladores e investigadores con varios niveles de experiencia en IA. El software alojado en NGC se somete a análisis contra un conjunto agregado de vulnerabilidades y exposiciones comunes (CVE), cripto y claves privadas. Está probado y diseñado para escalar a varias GPU y, en muchos casos, a varios nodos, garantizando que los usuarios maximicen la inversión en sistemas DGX.

NVIDIA GPU Cloud



IA Enterprise de NVIDIA

NVIDIA AI Enterprise es la plataforma de software integral que lleva la IA generativa al alcance de todas las empresas, lo que proporciona el tiempo de ejecución más rápido y eficiente para modelos de base de IA generativa optimizados para ejecutarse en la plataforma NVIDIA DGX. Gracias a la seguridad, estabilidad y capacidad de gestión del nivel de producción, se optimiza el desarrollo de soluciones de IA generativas. NVIDIA AI Enterprise se incluye en DGX BasePOD para que los desarrolladores empresariales accedan a modelos previamente entrenados, marcos optimizados, microservicios, bibliotecas aceleradas y soporte empresarial.

Software NetApp

ONTAP de NetApp

ONTAP 9, la última generación del software de gestión del almacenamiento de NetApp, permite a las empresas modernizar su infraestructura y realizar la transición a un centro de datos preparado para el cloud. ONTAP ofrece las mejores capacidades de gestión de datos y permite la gestión y protección de los datos con un solo conjunto de herramientas, sin importar dónde residan. También puede mover los datos libremente a donde sea necesario: El perímetro, el núcleo o el cloud. ONTAP 9 incluye numerosas funciones que simplifican la gestión de datos, aceleran y protegen los datos esenciales y permiten disfrutar de funcionalidades de infraestructura de nueva generación en arquitecturas de cloud híbrido.

Acelere y proteja sus datos

ONTAP no solo ofrece niveles de rendimiento y protección de datos superiores, sino que amplía estas capacidades de las siguientes maneras:

- Rendimiento y menor latencia. ONTAP ofrece la mayor salida posible con la menor latencia posible, incluida la compatibilidad con el almacenamiento GPUDirect (GDS) de NVIDIA mediante NFS a través de RDMA, Parallel NFS (pNFS) y conexión de enlaces de sesiones NFS.
- Protección de datos. ONTAP ofrece funcionalidades de protección de datos integradas y la mayor garantía antiransomware del sector, con una gestión común en todas las plataformas.
- Cifrado de volúmenes de NetApp (NVE). ONTAP ofrece cifrado nativo en el nivel de volumen y permite la gestión de claves incorporada o externa.
- Almacenamiento multi-tenancy y autenticación multifactor. ONTAP permite compartir recursos de infraestructura con los niveles más altos de seguridad.

Simplificar la gestión de los datos

La gestión de los datos es crucial para las operaciones TECNOLÓGICAS empresariales y los científicos de datos, para que se utilicen recursos apropiados para las aplicaciones de IA y para entrenar conjuntos de datos de IA/ML. La siguiente información adicional sobre las tecnologías de NetApp no está disponible para esta validación, pero puede ser relevante en función de su puesta en marcha.

El software para la gestión de datos ONTAP incluye las siguientes funciones para mejorar y simplificar las operaciones, y reducir el coste total de funcionamiento:

- Las copias Snapshot y clones permiten la colaboración, la experimentación paralela y una gestión de datos mejorada para los flujos de trabajo DE ML/DL.
- SnapMirror posibilita un movimiento fluido de datos en entornos de cloud híbrido y multisitio, para permitir que los datos se muevan cuando y donde sea necesario.
- Compactación de datos inline y deduplicación expandida. La compactación de datos reduce el espacio perdido dentro de los bloques de almacenamiento, mientras que la deduplicación aumenta la capacidad efectiva de forma significativa. Esto es aplicable a los datos almacenados localmente y a los datos organizados en niveles en el cloud.
- Calidad de servicio (AQoS) mínima, máxima y adaptativa. Los controles granulares de calidad de servicio (QoS) ayudan a mantener los niveles de rendimiento para aplicaciones críticas en entornos altamente compartidos.
- Los FlexGroup de NetApp permiten distribuir datos a través de todos los nodos del clúster de almacenamiento, lo que proporciona una capacidad masiva y un mayor rendimiento para conjuntos de datos extremadamente grandes.

- FabricPool de NetApp. Proporciona la organización automática en niveles de datos fríos en opciones de almacenamiento en cloud privado como Amazon Web Services (AWS), Azure y la solución de almacenamiento StorageGRID de NetApp. Para obtener más información sobre FabricPool, consulte "[TR-4598: Prácticas recomendadas de FabricPool](#)".
- NetApp FlexCache Proporciona funcionalidades de almacenamiento en caché de volúmenes remotos que simplifican la distribución de archivos, reduce la latencia de redes WAN y reduce los costes de ancho de banda WAN. FlexCache permite el desarrollo de productos distribuidos en varios sitios, así como el acceso acelerado a los conjuntos de datos corporativos desde ubicaciones remotas.

Infraestructura preparada para futuros retos

ONTAP ayuda a satisfacer las exigentes y siempre cambiantes necesidades de su empresa con las siguientes funciones:

- Escalado sencillo y funcionamiento sin interrupciones. ONTAP admite la adición en línea de capacidad a las controladoras existentes y a los clústeres de escalado horizontal. Los clientes pueden empezar a utilizar tecnologías punteras como NVMe y FC 32 GB, sin necesidad de realizar costosas migraciones de datos y sin cortes.
- Conexión de cloud. ONTAP es el software de gestión de almacenamiento con mejor conexión de cloud e incluye opciones de almacenamiento definido por software (ONTAP Select) e instancias nativas del cloud (NetApp Cloud Volumes Service) en todos los clouds públicos.
- Integración con aplicaciones emergentes. ONTAP ofrece servicios de datos de clase empresarial para plataformas y aplicaciones de última generación, como vehículos autónomos, ciudades inteligentes e Industria 4.0, utilizando la misma infraestructura que da soporte a las aplicaciones empresariales existentes.

Kit de herramientas de operaciones de datos de NetApp

El kit de herramientas DataOps de NetApp es una herramienta basada en Python que simplifica la gestión de espacios de trabajo de desarrollo/formación y servidores de inferencia respaldados por un almacenamiento de NetApp de escalado horizontal y de alto rendimiento. DataOps Toolkit puede funcionar como una utilidad independiente y es incluso más eficaz en entornos de Kubernetes si se utiliza Astra Trident de NetApp para automatizar las operaciones de almacenamiento. Estas son algunas funcionalidades clave:

- Aprovechamiento con rapidez nuevos espacios de trabajo de JupyterLab de alta capacidad respaldados por el almacenamiento de NetApp de escalado horizontal y de alto rendimiento.
- Aprovechamiento rápidamente nuevas instancias del servidor de inferencia de NVIDIA Triton, respaldadas por un almacenamiento empresarial de NetApp.
- Clonación casi instantánea de espacios de trabajo JupyterLab de gran capacidad para permitir la experimentación o la iteración rápida.
- Copias Snapshot casi instantáneas de espacios de trabajo de gran capacidad JupyterLab para backup o trazabilidad/creación de bases de datos.
- Aprovechamiento, clonado y copias Snapshot casi instantáneas de volúmenes de datos de gran capacidad y alto rendimiento.

Astra Trident de NetApp

Astra Trident es un orquestador de almacenamiento de código abierto y totalmente compatible para contenedores y distribuciones de Kubernetes, incluido Anthos. Trident funciona con toda la cartera de almacenamiento de NetApp, incluida NetApp ONTAP, y también es compatible con conexiones NFS, NVMe/TCP e iSCSI. Trident acelera el flujo de trabajo de DevOps al permitir que los usuarios finales

aprovisionen y gestionen el almacenamiento desde sus sistemas de almacenamiento de NetApp sin necesidad de intervención del administrador de almacenamiento.

["Siguiendo: AIPod de NetApp con sistemas NVIDIA DGX: Arquitectura de soluciones"](#)

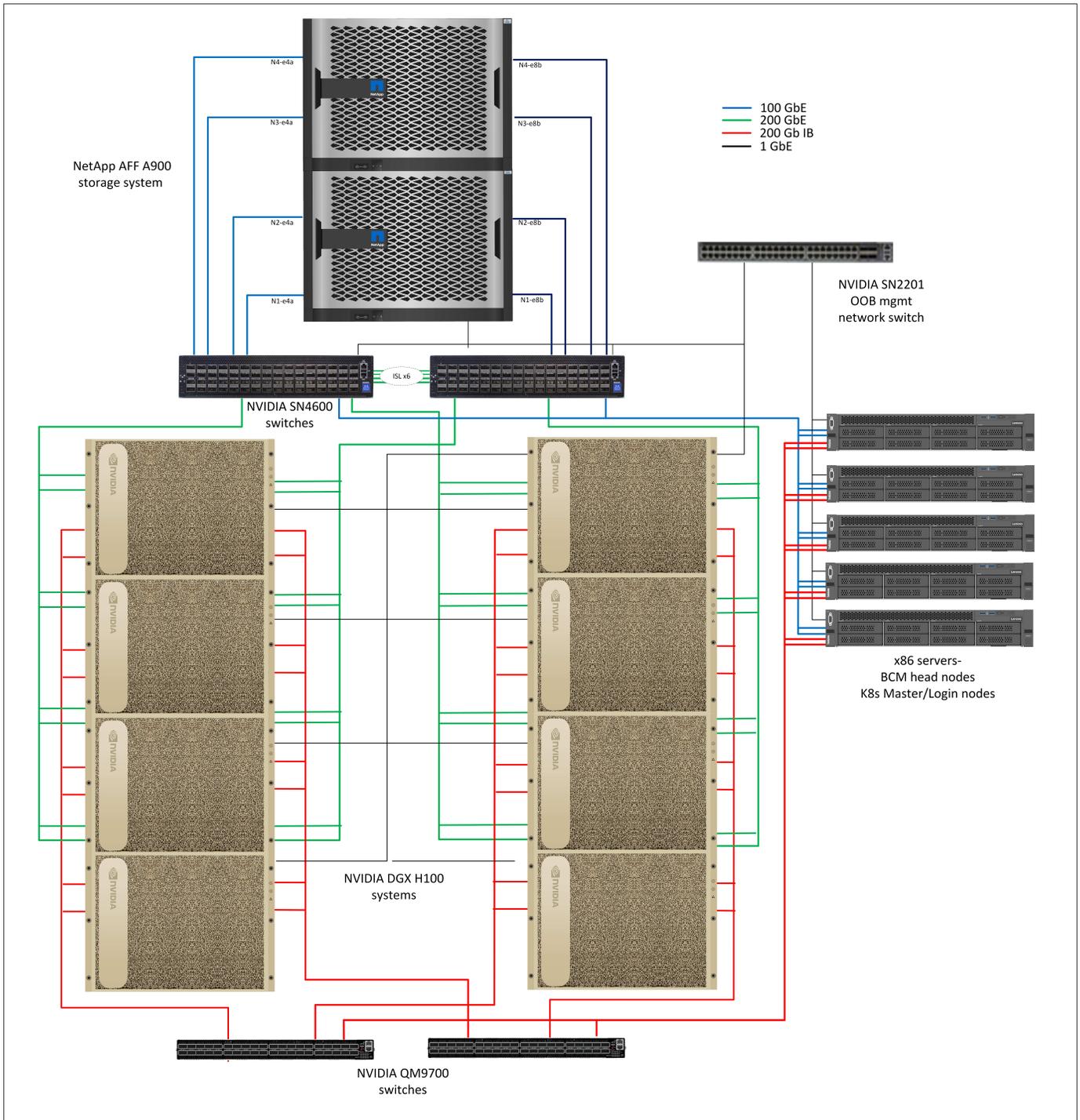
NetApp AIPod con sistemas NVIDIA DGX: Arquitectura de la solución

["Anterior: ONTAP AI - Componentes de software"](#)

AI Pod de NetApp con sistemas DGX H100

Esta arquitectura de referencia aprovecha las estructuras separadas para la interconexión en clústeres de computación y el acceso al almacenamiento, con conectividad InfiniBand (IB) de 400GB Gb/s entre los nodos de computación. El siguiente dibujo muestra la topología general de la solución de NetApp AIPod con sistemas DGX H100.

Topología de la solución AIPod de NetApp



Configuración de red

En esta configuración, la estructura de clústeres de computación utiliza un par de switches IB de QM9700 400GB Gb/s, que se conectan entre sí para lograr una alta disponibilidad. Cada sistema DGX H100 está conectado a los switches mediante ocho conexiones, con puertos números pares conectados a un switch y puertos con número impar conectados al otro switch.

Para el acceso al sistema de almacenamiento, la gestión en banda y el acceso de clientes, se utiliza un par de switches Ethernet SN4600 Gb. Los switches están conectados con enlaces entre switches y se configuran con varias VLAN para aislar los distintos tipos de tráfico. Para implementaciones mayores, la red Ethernet se puede expandir a una configuración hoja-espina añadiendo pares de switches adicionales para los switches

espina y hojas adicionales según sea necesario.

Además de la interconexión informática y las redes Ethernet de alta velocidad, todos los dispositivos físicos también están conectados a uno o más switches Ethernet de SN2201 Gb para la gestión fuera de banda. Para obtener más información sobre la conectividad del sistema DGX H100, consulte la "[Documentación de NVIDIA BasePOD](#)".

Configuración de cliente para el acceso al almacenamiento

Cada sistema DGX H100 está aprovisionado con dos adaptadores ConnectX-7 de doble puerto para el tráfico de gestión y almacenamiento. Para esta solución, ambos puertos de cada tarjeta están conectados al mismo switch. Después, un puerto de cada tarjeta se configura en un vínculo LACP MLAG con un puerto conectado a cada switch, y las VLAN para la gestión en banda, el acceso de clientes y el acceso al almacenamiento en el nivel de usuario se alojan en este vínculo.

El otro puerto de cada tarjeta se utiliza para la conectividad con los sistemas de almacenamiento AFF A900 y se puede utilizar en varias configuraciones según los requisitos de la carga de trabajo. Para configuraciones que utilizan NFS a través de RDMA para admitir el almacenamiento Magnum IO GPUDirect de NVIDIA, los puertos se configuran en un vínculo activo-pasivo, ya que RDMA no es compatible con ningún otro tipo de vínculo. Para las puestas en marcha que no requieren RDMA, las interfaces de almacenamiento también se pueden configurar mediante enlaces LACP para ofrecer una alta disponibilidad y ancho de banda adicional. Con o sin RDMA, los clientes pueden montar el sistema de almacenamiento mediante pNFS v4,1 y Trunking de sesiones para permitir el acceso en paralelo a todos los nodos de almacenamiento del clúster.

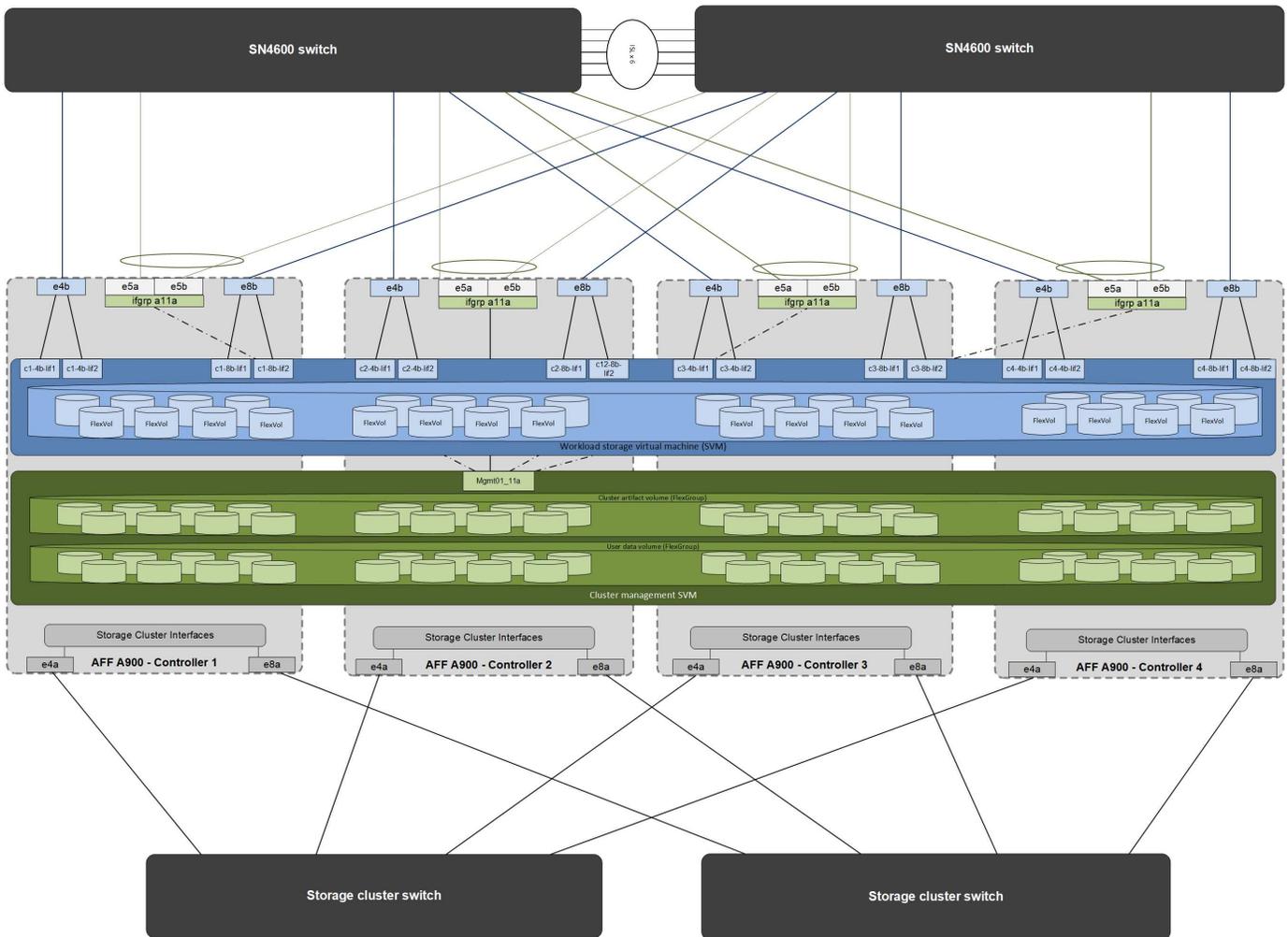
Configuración del sistema de almacenamiento

Cada sistema de almacenamiento A900 de AFF está conectado mediante cuatro puertos de 100 GbE desde cada controladora. Dos puertos de cada controladora se utilizan para el acceso a los datos de carga de trabajo desde los sistemas DGX y dos puertos de cada controladora están configurados como un grupo de interfaces LACP para admitir el acceso desde los servidores del plano de gestión para artefactos de gestión de clústeres y directorios iniciales de usuario. Todos los accesos a datos desde el sistema de almacenamiento se realizan mediante NFS, con una máquina virtual de almacenamiento (SVM) dedicada al acceso a las cargas de trabajo de IA y una SVM independiente dedicada a los usos de gestión del clúster.

La SVM de carga de trabajo está configurada con un total de ocho interfaces lógicas (LIF), con dos para cada puerto físico. Esta configuración proporciona el ancho de banda máximo, así como los medios para que cada LIF pueda conmutar por error a otro puerto de la misma controladora, de modo que ambas controladoras permanezcan activas en caso de un fallo de red. Esta configuración también admite NFS sobre RDMA para habilitar el acceso a almacenamiento GPUDirect. La capacidad de almacenamiento se aprovisiona en forma de un único volumen FlexGroup grande que abarca todas las controladoras de almacenamiento del clúster, con 16 volúmenes constituyentes en cada controladora. Puede accederse a esta FlexGroup desde cualquiera de las LIF en la SVM y al utilizar NFSv4,1 con pNFS y conexión de enlaces de sesiones, los clientes establecen conexiones a todas las LIF de la SVM, lo cual permite acceder a los datos locales de cada nodo de almacenamiento en paralelo con el fin de mejorar significativamente el rendimiento. La SVM de carga de trabajo y cada LIF de datos también están configuradas para el acceso a protocolo RDMA. Para obtener más información sobre la configuración de RDMA para ONTAP, consulte la "[Documentación de ONTAP](#)".

La SVM de gestión solo requiere un solo LIF, que está alojado en los grupos de interfaz de 2 puertos configurados en cada controladora. Otros volúmenes FlexGroup se aprovisionan en la SVM de gestión con el fin de albergar artefactos de gestión del clúster, como imágenes de nodos de clúster, datos históricos de supervisión del sistema y directorios iniciales de usuarios finales. El siguiente dibujo muestra la configuración lógica del sistema de almacenamiento.

Configuración lógica del clúster de almacenamiento de NetApp A900



Servidores del plano de gestión

Esta arquitectura de referencia también incluye cinco servidores basados en CPU para los usos del plano de gestión. Dos de estos sistemas se usan como nodos principales de NVIDIA Base Command Manager para la puesta en marcha y la gestión del clúster. Los otros tres sistemas se utilizan para proporcionar servicios de clúster adicionales, como los nodos maestros de Kubernetes o los nodos de inicio de sesión para las implementaciones que utilizan Slurm para la programación de tareas. Las puestas en marcha que utilizan Kubernetes pueden aprovechar el controlador CSI Astra Trident de NetApp para proporcionar aprovisionamiento automatizado y servicios de datos con almacenamiento persistente para cargas de trabajo de gestión y de IA en el sistema de almacenamiento A900 de AFF.

Cada servidor está conectado físicamente a los switches IB y Ethernet para permitir la puesta en marcha y gestión de clústeres. Además, está configurado con montajes NFS en el sistema de almacenamiento a través de la SVM de gestión para almacenamiento de artefactos de gestión de clústeres, tal como se ha descrito anteriormente.

["Próximo: AIPod de NetApp con sistemas NVIDIA DGX: Validación de la solución y guía de tamaño"](#)

AIPod de NetApp con sistemas NVIDIA DGX: Validación de la solución y guía de tamaño

["Anterior: AIPod de NetApp con sistemas NVIDIA DGX: Arquitectura de soluciones"](#)

Validación de la solución

La configuración del almacenamiento de esta solución se validó mediante una serie de cargas de trabajo sintéticas utilizando la herramienta de código abierto FIO. Estas pruebas incluyen patrones de I/O de lectura y escritura destinados a simular la carga de trabajo de almacenamiento generada por los sistemas DGX que realizan trabajos de entrenamiento de aprendizaje profundo. La configuración del almacenamiento se validó utilizando un clúster de servidores CPU de 2 sockets que ejecutaban las cargas de trabajo FIO de forma simultánea para simular un clúster de sistemas DGX. Cada cliente se configuró con la misma configuración de red descrita anteriormente, con la adición de los siguientes detalles.

Para esta validación se han utilizado las siguientes opciones de montaje:

- `vers=4,1` # Habilita pNFS para el acceso paralelo a varios nodos de almacenamiento
- `proto=rdma` # establece el protocolo de transferencia en RDMA en lugar del TCP predeterminado
- `puerto=20049` # Especifica el puerto correcto para el servicio NFS RDMA
- `max_connect=16` # Permite el trunking de sesión NFS para agregar ancho de banda de puerto de almacenamiento
- `write=eager` # mejora el rendimiento de escritura de escrituras en búfer
- `Rsize=262144,wsiz=262144` # establece el tamaño de transferencia de I/O en 256K kb

Además, los clientes se configuraron con un valor de NFS `max_session_slots` de 1024. Dado que la solución se probó con NFS a través de RDMA, los puertos de redes de almacenamiento se configuraron con un vínculo activo-pasivo. Para esta validación se han utilizado los siguientes parámetros de enlace:

- `mode=backup activo` # establece el enlace al modo activo / pasivo
- `primario=<interface name>` # las interfaces primarias para todos los clientes se distribuyeron a través de los switches
- `mii-monitor-interval=100` # especifica el intervalo de monitoreo de 100ms
- `fail-over-mac-policy=active` # especifica que la dirección MAC del enlace activo es el MAC del enlace. Esto es necesario para el correcto funcionamiento de RDMA a través de la interfaz vinculada.

El sistema de almacenamiento se configuró como se describe con dos pares de alta disponibilidad A900 (4 controladoras) con dos bandejas de discos NS224 de 24 1,9TB unidades de disco NVMe añadidas a cada par de alta disponibilidad. Tal y como se indica en la sección de arquitectura, la capacidad de almacenamiento de todas las controladoras se combinó mediante el uso de un volumen FlexGroup y la información de todos los clientes se distribuyó entre todas las controladoras del clúster.

Directrices de tamaño del sistema de almacenamiento

NetApp ha completado correctamente la certificación DGX BasePOD y los dos pares de alta disponibilidad A900 como pruebas pueden admitir fácilmente un clúster de ocho sistemas DGX H100. Para puestas en marcha de mayor tamaño con requisitos de rendimiento del almacenamiento superiores, es posible añadir sistemas AFF adicionales al clúster de NetApp ONTAP hasta 12 pares de alta disponibilidad (24 nodos) en un único clúster. Con la tecnología FlexGroup descrita en esta solución, un clúster de 24 nodos puede proporcionar más de 40 PB y hasta 300 Gbps de rendimiento en un solo espacio de nombres. Otros sistemas de almacenamiento de NetApp como AFF A400, A250 y C800 ofrecen un rendimiento menor y opciones de capacidad superior para puestas en marcha de menor tamaño a puntos de coste menores. Como ONTAP 9 admite clústeres de modelo mixto, los clientes pueden comenzar con una huella inicial pequeña e ir aumentando el sistema de almacenamiento a medida que crezcan los requisitos de capacidad y rendimiento. La siguiente tabla muestra una estimación aproximada del número de GPU A100 y H100 admitidas en cada modelo de AFF.

Guía para ajustar el tamaño del sistema de almacenamiento de NetApp

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

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

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

"Siguiente: AIPOd de NetApp con sistemas NVIDIA DGX: Conclusión e información adicional"

AIPOd de NetApp con sistemas NVIDIA DGX: Conclusión e información adicional

"Anterior: AIPOd de NetApp con sistemas NVIDIA DGX: Validación de la solución y guía de tamaño"

Conclusión

La arquitectura de DGX BasePOD es una plataforma de aprendizaje profundo de próxima generación que requiere funcionalidades de gestión de datos y almacenamiento igualmente avanzadas. Combinando DGX BasePOD con sistemas NetApp AFF, el AIPOd de NetApp con la arquitectura de sistemas DGX se puede implementar a casi cualquier escala hasta 48 sistemas DGX H100 en un clúster AFF A900 de 24 nodos. Gracias a la superior integración con el cloud y las capacidades definidas por software de NetApp ONTAP, AFF ofrece una completa gama de canalizaciones de datos que abarcan el perímetro, el núcleo y el cloud y que le permitirán llevar a buen puerto sus proyectos de DL.

Información adicional

Para obtener más información sobre la información descrita en este documento, consulte los siguientes documentos y/o sitios web:

- Software de gestión de datos ONTAP de NetApp: Biblioteca de información de ONTAP

"<https://docs.netapp.com/us-en/ontap-family/>"

- Sistemas DE ALMACENAMIENTO A900 DE NetApp AFF-

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

- Información de RDMA de NetApp ONTAP-

["https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html"](https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html)

- Kit de herramientas de operaciones de datos de NetApp

["https://github.com/NetApp/netapp-dataops-toolkit"](https://github.com/NetApp/netapp-dataops-toolkit)

- Astra Trident de NetApp

["https://docs.netapp.com/us-en/netapp-solutions/containers/rh-os-n_overview_trident.html"](https://docs.netapp.com/us-en/netapp-solutions/containers/rh-os-n_overview_trident.html)

- Blog sobre almacenamiento GPUDirect de NetApp-

["https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/"](https://www.netapp.com/blog/ontap-reaches-171-gpudirect-storage/)

- DGX BasePOD de NVIDIA

["https://www.nvidia.com/en-us/data-center/dgx-basepod/"](https://www.nvidia.com/en-us/data-center/dgx-basepod/)

- SISTEMAS DGX H100 DE NVIDIA

["https://www.nvidia.com/en-us/data-center/dgx-h100/"](https://www.nvidia.com/en-us/data-center/dgx-h100/)

- Conexión a redes NVIDIA

["https://www.nvidia.com/en-us/networking/"](https://www.nvidia.com/en-us/networking/)

- Almacenamiento NVIDIA Magnum IO GPUDirect

["https://docs.nvidia.com/gpudirect-storage"](https://docs.nvidia.com/gpudirect-storage)

- Comando base NVIDIA

["https://www.nvidia.com/en-us/data-center/base-command/"](https://www.nvidia.com/en-us/data-center/base-command/)

- Administrador de comandos base de NVIDIA

["https://www.nvidia.com/en-us/data-center/base-command/manager"](https://www.nvidia.com/en-us/data-center/base-command/manager)

- IA Enterprise de NVIDIA

["https://www.nvidia.com/en-us/data-center/products/ai-enterprise/"](https://www.nvidia.com/en-us/data-center/products/ai-enterprise/)

Reconocimientos

Este documento es el trabajo de los equipos de Ingeniería de Soluciones NetApp y ONTAP: David Arnette, Olga Kornievskaia, Dustin Fischer, Srikanth Kaligotla, Mohit Kumar y Rajeev Badrinath. Los autores también quieren agradecer a NVIDIA y al equipo de ingeniería de NVIDIA DGX BasePOD su apoyo continuo.

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 INTERRUPTIÓN 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.