NVIDIA AI Enterprise con NetApp y VMware

NetApp artificial intelligence solutions

NetApp
August 18, 2025

This PDF was generated from https://docs.netapp.com/es-es/netapp-solutions-ai/software/ai-nvaie-vmw-intro.html on August 18, 2025. Always check docs.netapp.com for the latest.

Tabla de contenidos

NVIDIA AI Enterprise con NetApp y VMware . . . 1
- NVIDIA AI Enterprise con NetApp y VMware . . . 1
- Descripción general de la tecnología . . . 1
  - NVIDIA AI Enterprise . . . 2
  - Nube de GPU NVIDIA (NGC) . . . 2
  - VMware vSphere . . . 2
  - ONTAP de NetApp . . . 2
  - Kit de herramientas DataOps de NetApp . . . 3
- Arquitectura . . . 4
- Configuración inicial . . . 5
  - Prerrequisitos . . . 5
  - Instalar el software de host empresarial NVIDIA AI . . . 6
- Utilice el software NVIDIA NGC . . . 6
  - Configuración . . . 6
  - Ejemplo de caso de uso: trabajo de capacitación de TensorFlow . . . 8
- Dónde encontrar información adicional . . . 10
  - Expresiones de gratitud . . . 11

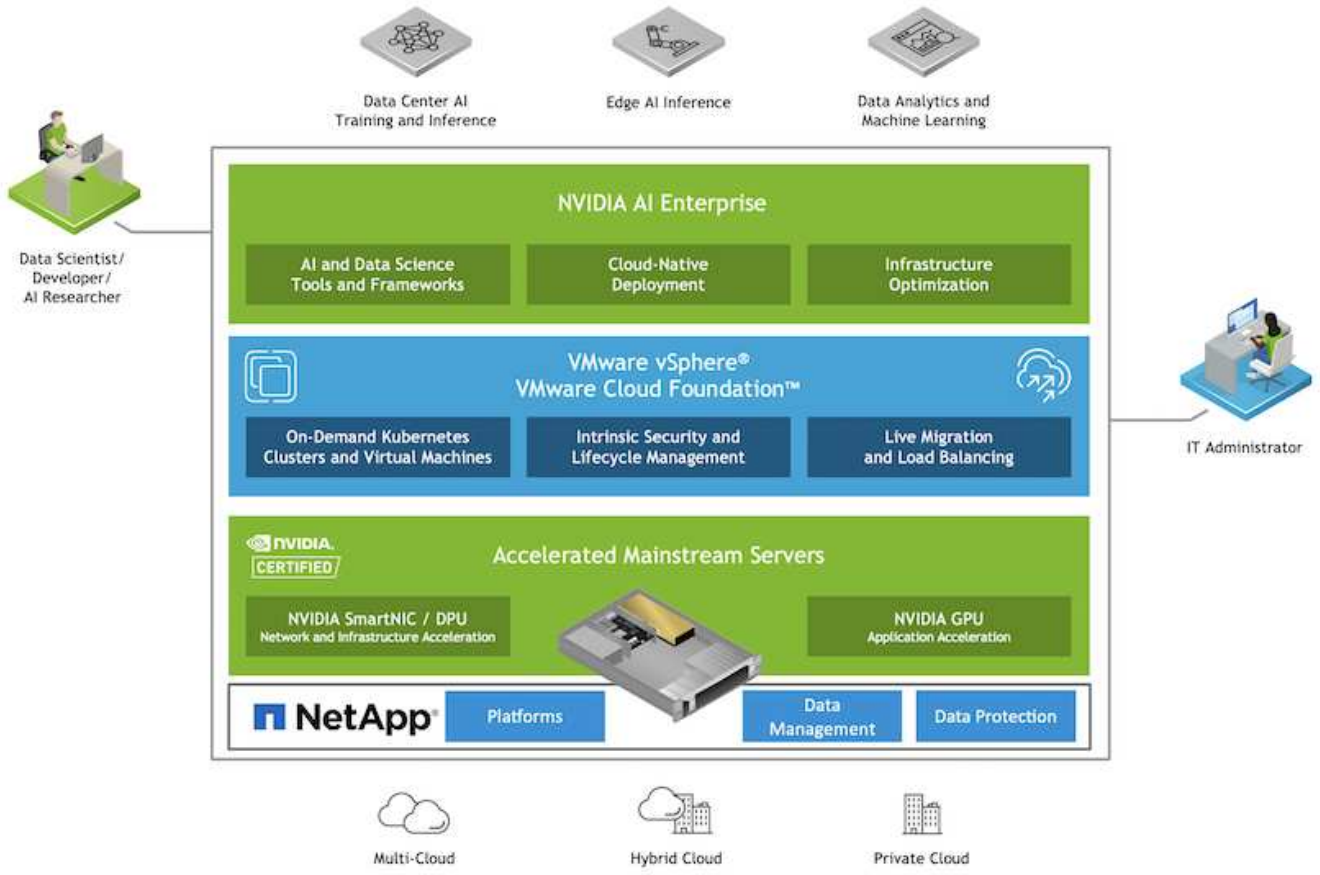
NVIDIA AI Enterprise con NetApp y VMware

NVIDIA AI Enterprise con NetApp y VMware

Mike Oglesby, NetApp

Para los arquitectos y administradores de TI, las herramientas de IA pueden ser complicadas y desconocidas. Además, muchas plataformas de IA no están preparadas para el uso empresarial. NVIDIA AI Enterprise, impulsado por NetApp y VMware, se creó para ofrecer una arquitectura de IA optimizada y de clase empresarial.

NVIDIA AI Enterprise es una suite integral de software de inteligencia artificial y análisis de datos nativa de la nube, optimizada, certificada y respaldada por NVIDIA para ejecutarse en VMware vSphere con sistemas certificados NVIDIA. Este software facilita la implementación, la gestión y el escalamiento simples y rápidos de cargas de trabajo de IA en el entorno de nube híbrida moderno. NVIDIA AI Enterprise, con tecnología de NetApp y VMware, ofrece gestión de datos y cargas de trabajo de IA de nivel empresarial en un paquete simplificado y familiar.



Descripción general de la tecnología

Esta sección proporciona una descripción general de la tecnología para NVIDIA AI Enterprise con NetApp y VMware.

NVIDIA AI Enterprise

NVIDIA AI Enterprise es una suite integral de software de inteligencia artificial y análisis de datos nativa de la nube, optimizada, certificada y respaldada por NVIDIA para ejecutarse en VMware vSphere con sistemas certificados NVIDIA. Este software facilita la implementación, la gestión y el escalamiento simples y rápidos de cargas de trabajo de IA en el entorno de nube híbrida moderno.

Nube de GPU NVIDIA (NGC)

NVIDIA NGC alberga un catálogo de software optimizado para GPU para que los profesionales de IA desarrollen sus soluciones de IA. También proporciona acceso a varios servicios de IA, incluidos NVIDIA Base Command para entrenamiento de modelos, NVIDIA Fleet Command para implementar y monitorear modelos, y NGC Private Registry para acceder y administrar de forma segura software de IA propietario. Además, los clientes de NVIDIA AI Enterprise pueden solicitar soporte a través del portal NGC.

VMware vSphere

VMware vSphere es la plataforma de virtualización de VMware, que transforma los centros de datos en infraestructuras informáticas agregadas que incluyen recursos de CPU, almacenamiento y redes. vSphere administra estas infraestructuras como un entorno operativo unificado y proporciona a los administradores las herramientas para administrar los centros de datos que participan en ese entorno.

Los dos componentes principales de vSphere son ESXi y vCenter Server. ESXi es la plataforma de virtualización donde los administradores crean y ejecutan máquinas virtuales y dispositivos virtuales. vCenter Server es el servicio a través del cual los administradores gestionan múltiples hosts conectados en una red y agrupan los recursos del host.

ONTAP de NetApp

ONTAP 9, la última generación de software de gestión de almacenamiento de NetApp, permite a las empresas modernizar la infraestructura y realizar la transición a un centro de datos preparado para la nube. Al aprovechar las capacidades de gestión de datos líderes en la industria, ONTAP permite la gestión y protección de datos con un único conjunto de herramientas, independientemente de dónde residan esos datos. También puede mover datos libremente a donde sea necesario: el borde, el núcleo o la nube. ONTAP 9 incluye numerosas características que simplifican la gestión de datos, aceleran y protegen datos críticos y habilitan capacidades de infraestructura de próxima generación en arquitecturas de nube híbrida.

Simplificar la gestión de datos

La gestión de datos es crucial para las operaciones de TI de la empresa y los científicos de datos, de modo que se utilicen los recursos adecuados para las aplicaciones de IA y el entrenamiento de conjuntos de datos de IA/ML. La siguiente información adicional sobre las tecnologías de NetApp está fuera del alcance de esta validación, pero podría ser relevante según su implementación.

El software de gestión de datos ONTAP incluye las siguientes características para optimizar y simplificar las operaciones y reducir el costo total de operación:

- Compactación de datos en línea y deduplicación ampliada. La compactación de datos reduce el espacio desperdiciado dentro de los bloques de almacenamiento y la deduplicación aumenta significativamente la capacidad efectiva. Esto se aplica a los datos almacenados localmente y a los datos almacenados en la nube.
- Calidad de servicio mínima, máxima y adaptativa (AQoS). Los controles granulares de calidad de servicio (QoS) ayudan a mantener los niveles de rendimiento de las aplicaciones críticas en entornos altamente

compartidos.

- FabricPool de NetApp . Proporciona niveles automáticos de datos fríos en opciones de almacenamiento en la nube pública y privada, incluidas Amazon Web Services (AWS), Azure y la solución de almacenamiento NetApp StorageGRID . Para obtener más información sobre FabricPool, consulte "TR-4598: Prácticas recomendadas de FabricPool" .

Acelerar y proteger los datos

ONTAP ofrece niveles superiores de rendimiento y protección de datos y amplía estas capacidades de las siguientes maneras:

- Rendimiento y menor latencia. ONTAP ofrece el mayor rendimiento posible con la menor latencia posible.
- Protección de datos. ONTAP proporciona capacidades de protección de datos integradas con gestión común en todas las plataformas.
- Cifrado de volumen de NetApp (NVE). ONTAP ofrece cifrado nativo a nivel de volumen con soporte para administración de claves interna y externa.
- Autenticación multitenencia y multifactor. ONTAP permite compartir recursos de infraestructura con los más altos niveles de seguridad.

Infraestructura a prueba de futuro

ONTAP ayuda a satisfacer necesidades comerciales exigentes y en constante cambio con las siguientes características:

- Escalabilidad fluida y operaciones sin interrupciones. ONTAP admite la incorporación de capacidad sin interrupciones a controladores existentes y a clústeres de escalamiento horizontal. Los clientes pueden actualizar a las últimas tecnologías, como NVMe y FC de 32 Gb, sin migraciones de datos costosas ni interrupciones.
- Conexión a la nube. ONTAP es el software de gestión de almacenamiento más conectado a la nube, con opciones para almacenamiento definido por software (ONTAP Select) e instancias nativas de la nube (Google Cloud NetApp Volumes) en todas las nubes públicas.
- Integración con aplicaciones emergentes. ONTAP ofrece servicios de datos de nivel empresarial para plataformas y aplicaciones de próxima generación, como vehículos autónomos, ciudades inteligentes e Industria 4.0, utilizando la misma infraestructura que respalda las aplicaciones empresariales existentes.

Kit de herramientas DataOps de NetApp

NetApp DataOps Toolkit es una herramienta basada en Python que simplifica la gestión de espacios de trabajo de desarrollo/entrenamiento y servidores de inferencia respaldados por almacenamiento NetApp de alto rendimiento y escalabilidad horizontal. Las capacidades clave incluyen:

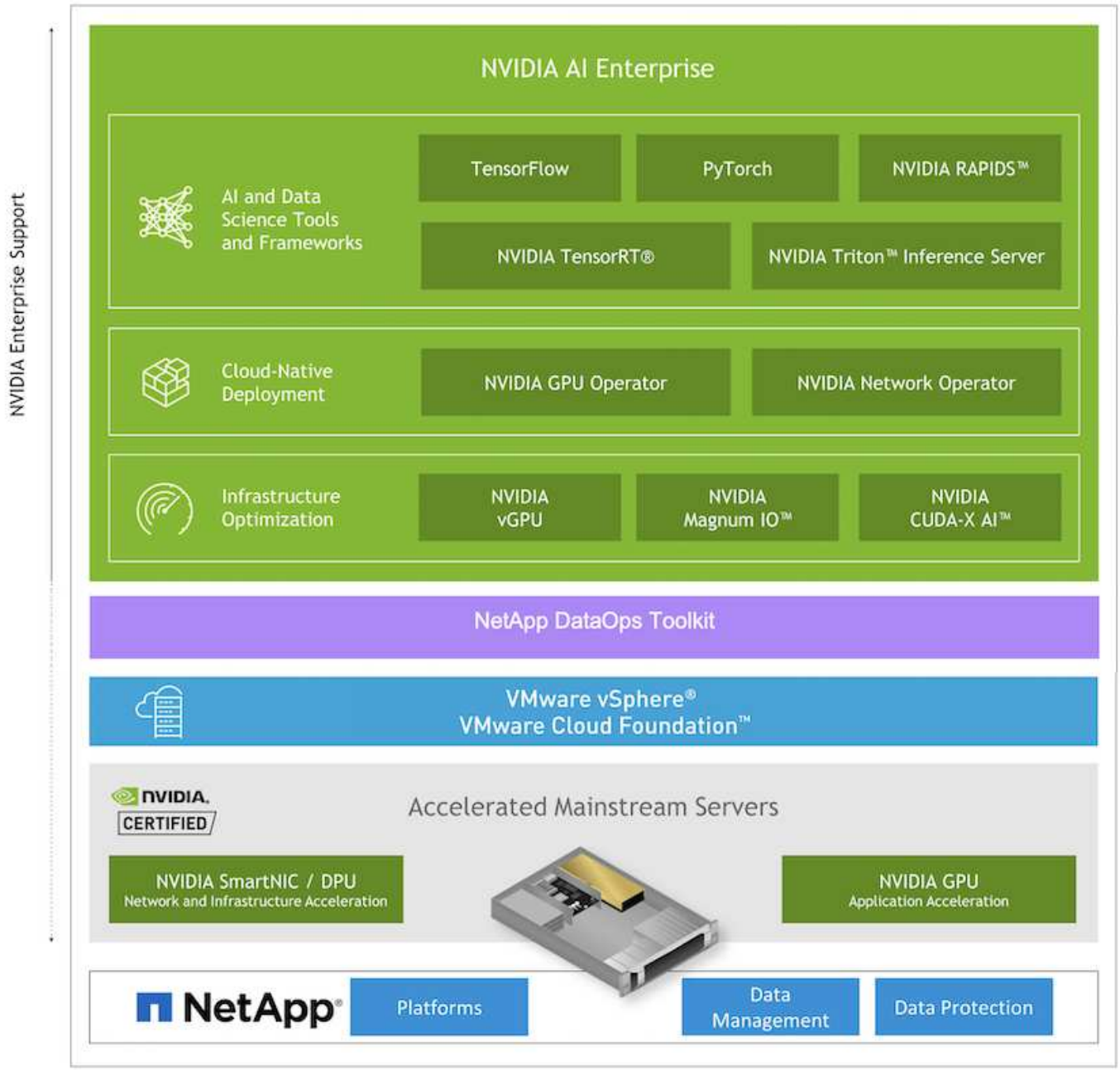
- Aprovisione rápidamente nuevos espacios de trabajo JupyterLab de alta capacidad respaldados por almacenamiento NetApp de alto rendimiento y escalabilidad horizontal.
- Aprovisione rápidamente nuevas instancias de NVIDIA Triton Inference Server respaldadas por almacenamiento NetApp de clase empresarial.
- Clone casi instantáneamente espacios de trabajo de JupyterLab de alta capacidad para permitir la experimentación o la iteración rápida.
- Guarde casi instantáneamente instantáneas de espacios de trabajo de JupyterLab de alta capacidad para realizar copias de seguridad y/o trazabilidad/establecimiento de referencia.
- Aprovisione, clone y cree instantáneas de volúmenes de datos de alto rendimiento y alta capacidad de

manera casi instantánea.

Arquitectura

Esta solución se basa en una arquitectura probada y familiar con sistemas certificados por NetApp, VMware y NVIDIA. Consulte la siguiente tabla para obtener más detalles.

Componente	Detalles
Software de inteligencia artificial y análisis de datos	"NVIDIA AI Enterprise para VMware"
Plataforma de virtualización	"VMware vSphere"
Plataforma de cómputo	"Sistemas certificados por NVIDIA"
Plataforma de gestión de datos	"ONTAP de NetApp"



Configuración inicial

Esta sección describe las tareas de configuración inicial que deben realizarse para utilizar NVIDIA AI Enterprise con NetApp y VMware.

Prerrequisitos

Antes de realizar los pasos que se describen en esta sección, asumimos que ya ha implementado VMware vSphere y NetApp ONTAP. Consulte la"Matriz de soporte de productos empresariales de IA de NVIDIA" para obtener detalles sobre las versiones de vSphere compatibles. Consulte la"Documentación de soluciones de NetApp y VMware" para obtener detalles sobre la implementación de VMware vSphere con NetApp ONTAP.

Instalar el software de host empresarial NVIDIA AI

Para instalar el software de host NVIDIA AI Enterprise, siga las instrucciones descritas en las secciones 1 a 4 del"Guía de inicio rápido de NVIDIA AI Enterprise" .

Utilice el software NVIDIA NGC

Configuración

Esta sección describe las tareas de configuración inicial que deben realizarse para utilizar el software empresarial NVIDIA NGC dentro de un entorno NVIDIA AI Enterprise.

Prerrequisitos

Antes de realizar los pasos que se describen en esta sección, asumimos que ya ha implementado el software de host NVIDIA AI Enterprise siguiendo las instrucciones que se describen en la"Configuración inicial" página.

Crear una máquina virtual invitada de Ubuntu con vGPU

Primero, debes crear una máquina virtual invitada Ubuntu 20.04 con vGPU. Para crear una máquina virtual invitada Ubuntu 20.04 con vGPU, siga las instrucciones que se detallan en el"Guía de implementación empresarial de NVIDIA AI" .

Descargar e instalar el software invitado de NVIDIA

A continuación, debe instalar el software invitado NVIDIA necesario dentro de la máquina virtual invitada que creó en el paso anterior. Para descargar e instalar el software invitado NVIDIA necesario dentro de la máquina virtual invitada, siga las instrucciones descritas en las secciones 5.1 a 5.4 del manual."Guía de inicio rápido de NVIDIA AI Enterprise" .

> Al realizar las tareas de verificación descritas en la sección 5.4, es posible que necesite utilizar una etiqueta de versión de imagen de contenedor CUDA diferente, ya que la imagen de contenedor CUDA se ha actualizado desde la redacción de esta guía. En nuestra validación, utilizamos 'nvidia/cuda:11.0.3-base-ubuntu20.04'.

Descargar contenedor(es) del marco de IA/análisis

A continuación, debe descargar las imágenes de contenedor del marco de análisis o IA necesarias de NVIDIA NGC para que estén disponibles en su máquina virtual invitada. Para descargar contenedores del marco dentro de la máquina virtual invitada, siga las instrucciones que se describen en"Guía de implementación empresarial de NVIDIA AI" .

Instalar y configurar el kit de herramientas NetApp DataOps

A continuación, debe instalar NetApp DataOps Toolkit para entornos tradicionales dentro de la máquina virtual invitada. El kit de herramientas NetApp DataOps se puede utilizar para administrar volúmenes de datos escalables en su sistema ONTAP directamente desde la terminal dentro de la máquina virtual invitada. Para instalar NetApp DataOps Toolkit dentro de la máquina virtual invitada, realice las siguientes tareas.

1. Instalar pip.

```
$ sudo apt update
$ sudo apt install python3-pip
$ python3 -m pip install netapp-dataops-traditional
```

2. Cierre la sesión de la terminal de la máquina virtual invitada y luego vuelva a iniciarla.
3. Configurar el kit de herramientas NetApp DataOps. Para completar este paso, necesitará detalles de acceso a la API para su sistema ONTAP . Es posible que necesite obtenerlos de su administrador de almacenamiento.

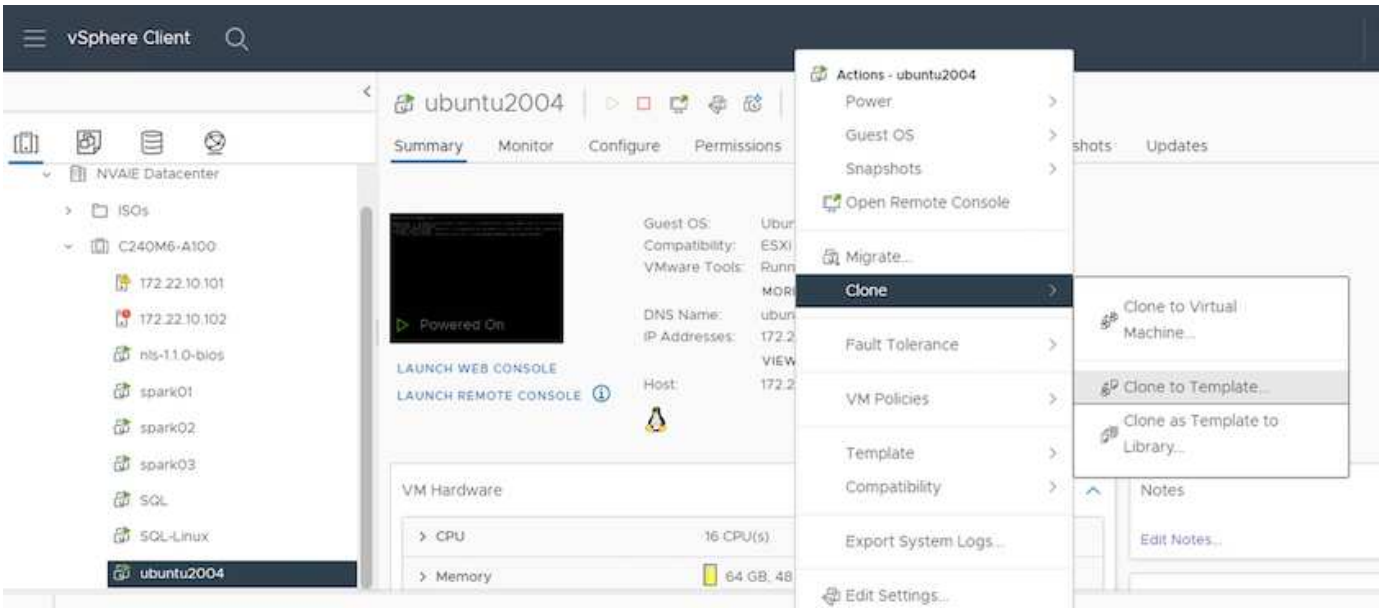
```
$ netapp_dataops_cli.py config

Enter ONTAP management LIF hostname or IP address (Recommendation: Use
SVM management interface): 172.22.10.10
Enter SVM (Storage VM) name: NVAIE-client
Enter SVM NFS data LIF hostname or IP address: 172.22.13.151
Enter default volume type to use when creating new volumes
(flexgroup/flexvol) [flexgroup]:
Enter export policy to use by default when creating new volumes
[default]:
Enter snapshot policy to use by default when creating new volumes
[none]:
Enter unix filesystem user id (uid) to apply by default when creating
new volumes (ex. '0' for root user) [0]:
Enter unix filesystem group id (gid) to apply by default when creating
new volumes (ex. '0' for root group) [0]:
Enter unix filesystem permissions to apply by default when creating new
volumes (ex. '0777' for full read/write permissions for all users and
groups) [0777]:
Enter aggregate to use by default when creating new FlexVol volumes:
aff_a400_01_NVME_SSD_1
Enter ONTAP API username (Recommendation: Use SVM account): admin
Enter ONTAP API password (Recommendation: Use SVM account):
Verify SSL certificate when calling ONTAP API (true/false): false
Do you intend to use this toolkit to trigger BlueXP Copy and Sync
operations? (yes/no): no
Do you intend to use this toolkit to push/pull from S3? (yes/no): no
Created config file: '/home/user/.netapp_dataops/config.json'.
```

Crear una plantilla de máquina virtual invitada

Por último, debes crear una plantilla de VM basada en tu VM invitada. Podrás usar esta plantilla para crear rápidamente máquinas virtuales invitadas para utilizar el software NVIDIA NGC.

Para crear una plantilla de VM basada en su VM invitada, inicie sesión en VMware vSphere, haga clic derecho en el nombre de la VM invitada, elija "Clonar", elija "Clonar a plantilla…" y luego siga el asistente.



Ejemplo de caso de uso: trabajo de capacitación de TensorFlow

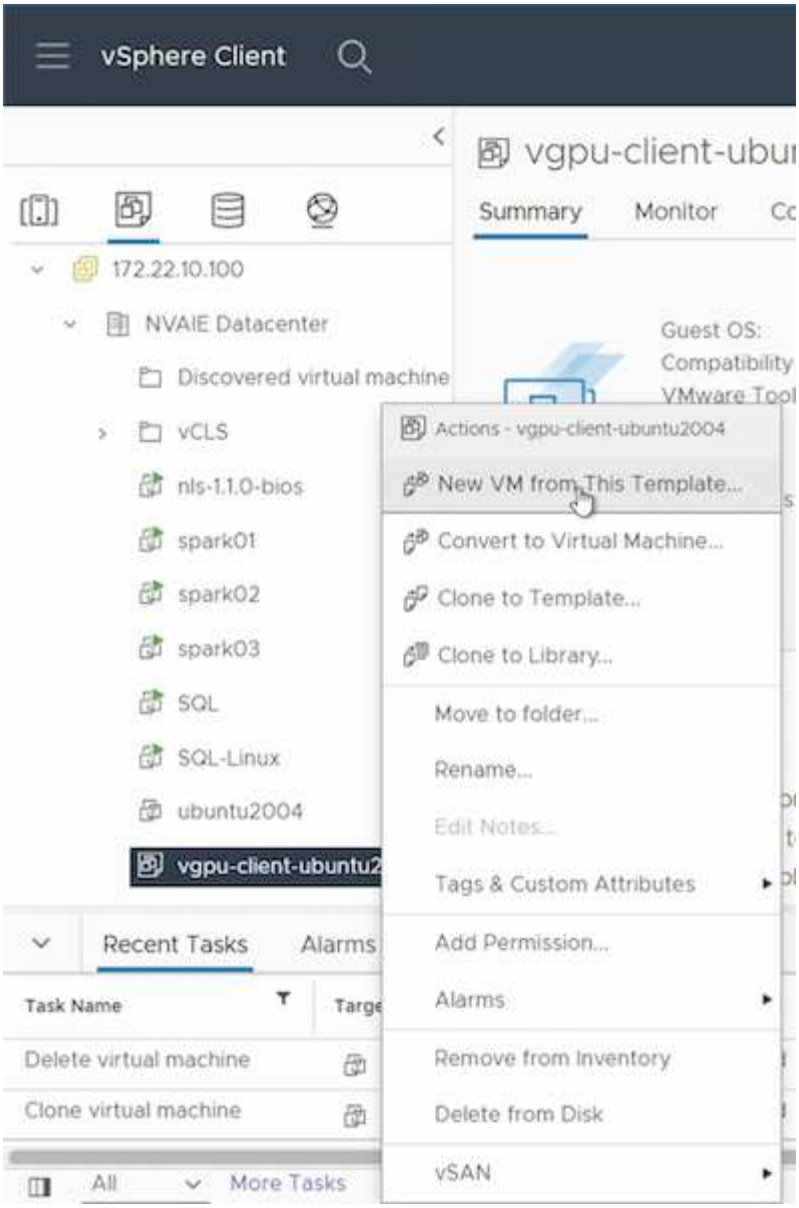
Esta sección describe las tareas que deben realizarse para ejecutar un trabajo de entrenamiento de TensorFlow dentro de un entorno de NVIDIA AI Enterprise.

Prerrequisitos

Antes de realizar los pasos que se describen en esta sección, asumimos que ya ha creado una plantilla de máquina virtual invitada siguiendo las instrucciones que se describen en la"Configuración" página.

Crear una máquina virtual invitada a partir de una plantilla

Primero, debes crear una nueva VM invitada a partir de la plantilla que creaste en la sección anterior. Para crear una nueva VM invitada a partir de su plantilla, inicie sesión en VMware vSphere, haga clic derecho en el nombre de la plantilla, seleccione "Nueva VM a partir de esta plantilla…" y luego siga el asistente.



Crear y montar un volumen de datos

A continuación, debe crear un nuevo volumen de datos en el que almacenar su conjunto de datos de entrenamiento. Puede crear rápidamente un nuevo volumen de datos utilizando el kit de herramientas NetApp DataOps. El comando de ejemplo que sigue muestra la creación de un volumen llamado 'imagenet' con una capacidad de 2 TB.

```
$ netapp_dataops_cli.py create vol -n imagenet -s 2TB
```

Antes de poder rellenar su volumen de datos con datos, debe montarlo dentro de la máquina virtual invitada. Puede montar rápidamente un volumen de datos utilizando el kit de herramientas NetApp DataOps. El comando de ejemplo que sigue muestra el montaje del volumen que se creó en el paso anterior.

```
$ sudo -E netapp_dataops_cli.py mount vol -n imagenet -m ~/imagenet
```

Rellenar volumen de datos

Una vez aprovisionado y montado el nuevo volumen, se puede recuperar el conjunto de datos de entrenamiento de la ubicación de origen y colocarlo en el nuevo volumen. Por lo general, esto implicará extraer los datos de un lago de datos S3 o Hadoop y, a veces, requerirá la ayuda de un ingeniero de datos.

Ejecutar trabajo de entrenamiento de TensorFlow

Ahora, está listo para ejecutar su trabajo de entrenamiento de TensorFlow. Para ejecutar su trabajo de entrenamiento de TensorFlow, realice las siguientes tareas.

1. Extraiga la imagen del contenedor TensorFlow empresarial NVIDIA NGC.

```
$ sudo docker pull nvcr.io/nvaie/tensorflow-2-1:22.05-tf1-nvaie-2.1-py3
```

2. Inicie una instancia del contenedor TensorFlow empresarial NVIDIA NGC. Utilice la opción '-v' para adjuntar su volumen de datos al contenedor.

```
$ sudo docker run --gpus all -v ~/imagenet:/imagenet -it --rm
nvcr.io/nvaie/tensorflow-2-1:22.05-tf1-nvaie-2.1-py3
```

3. Ejecute su programa de entrenamiento de TensorFlow dentro del contenedor. El siguiente comando de ejemplo muestra la ejecución de un programa de entrenamiento ResNet-50 de ejemplo que está incluido en la imagen del contenedor.

```
$ python ./nvidia-examples/cnn/resnet.py --layers 50 -b 64 -i 200 -u
batch --precision fp16 --data_dir /imagenet/data
```

Dónde encontrar 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 NetApp ONTAP : biblioteca de información de ONTAP

  http://mysupport.netapp.com/documentation/productlibrary/index.html?productID=62286

- Kit de herramientas DataOps de NetApp

  "https://github.com/NetApp/netapp-dataops-toolkit"

- NVIDIA AI Enterprise con VMware

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

Expresiones de gratitud

- Bobby Oommen, director sénior de NetApp
- Ramesh Isaac, administrador de sistemas, NetApp
- Roney Daniel, ingeniero de marketing técnico, NetApp

Información de copyright

Copyright © 2025 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.