NVIDIA AI Enterprise搭配NetApp和VMware

NetApp Solutions

NetApp
April 12, 2024

目錄

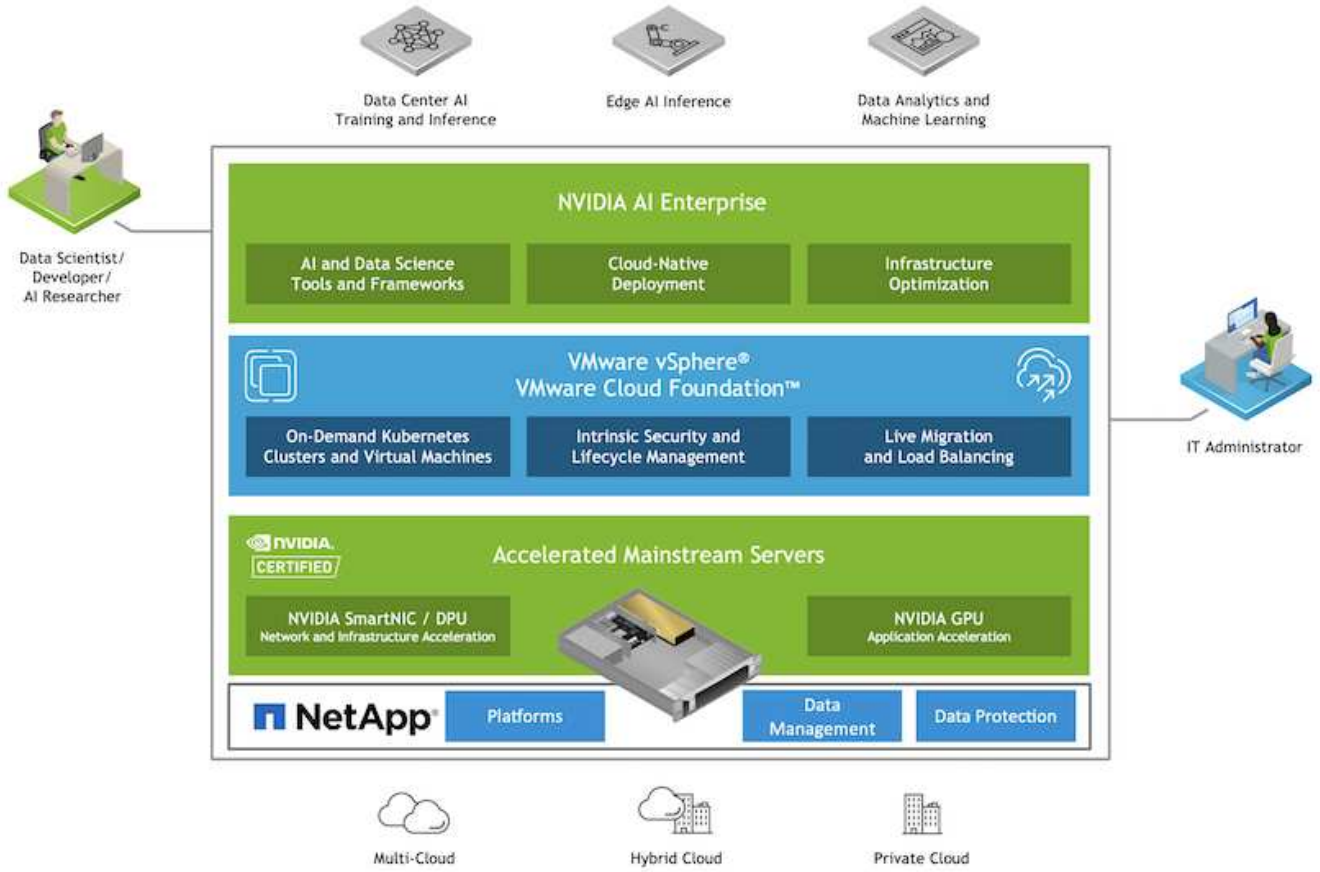
NVIDIA AI Enterprise搭配NetApp和VMware

NVIDIA AI Enterprise搭配NetApp和VMware

Mike Oglesby、NetApp

對於IT架構設計師和管理員而言、AI工具可能複雜且不熟悉。此外、許多AI平台還未做好企業準備。NVIDIA AI Enterprise採用NetApp和VMware技術、提供精簡的企業級AI架構。

NVIDIA AI Enterprise是一套端點對端點、雲端原生的AI與資料分析軟體套件、經過NVIDIA最佳化、認證及支援、可在採用NVIDIA認證系統的VMware vSphere上執行。此軟體可在現代化的混合雲環境中、輕鬆快速地部署、管理及擴充AI工作負載。NVIDIA AI Enterprise採用NetApp與VMware技術、以簡化且熟悉的套件提供企業級AI工作負載與資料管理功能。



技術總覽

NVIDIA AI Enterprise

NVIDIA AI Enterprise是一套端點對端點、雲端原生的AI與資料分析軟體套件、經過NVIDIA最佳化、認證及支援、可在採用NVIDIA認證系統的VMware vSphere上執行。此軟體可在現代化的混合雲環境中、輕鬆快速地部署、管理及擴充AI工作負載。

NVIDIA GPU雲端（NGC）

NVIDIA NGC主打GPU最佳化軟體目錄、讓AI從業者得以開發AI解決方案。此外、它還能存取各種AI服務、包括NVIDIA Base Command for Model訓練、NVIDIA車隊Command for Deploy and Monitor model、以及NGC Private登錄、以安全地存取及管理專屬AI軟體。此外、NVIDIA AI Enterprise客戶也可以透過NGC入口網站要求支援。

VMware vSphere

VMware vSphere是VMware的虛擬化平台、可將資料中心轉換成彙總式運算基礎架構、其中包括CPU、儲存設備和網路資源。vSphere將這些基礎架構管理為統一化的作業環境、並提供系統管理員工具來管理參與該環境的資料中心。

vSphere的兩個核心元件為ESXi和vCenter Server。ESXi是系統管理員建立及執行虛擬機器和虛擬應用裝置的虛擬化平台。vCenter Server是一項服務、可讓系統管理員管理網路和集區主機資源中連線的多個主機。

NetApp ONTAP

NetApp最新一代的儲存管理軟體、即支援企業將基礎架構現代化、並移轉至雲端就緒的資料中心。ONTAP利用領先業界的資料管理功能ONTAP 、無論資料位於何處、只要使用一組工具、即可管理及保護資料。您也可以自由地將資料移至任何需要的位置：邊緣、核心或雲端。支援眾多功能、可簡化資料管理、加速及保護關鍵資料、並在混合雲架構中提供新一代基礎架構功能。ONTAP

簡化資料管理

資料管理對於企業IT營運和資料科學家而言至關重要、因此可將適當的資源用於AI應用程式和訓練AI/ML資料集。下列關於NetApp技術的其他資訊超出此驗證範圍、但可能會因您的部署而有所差異。

包含下列功能的資料管理軟體、可簡化及簡化作業、並降低您的總營運成本：ONTAP

- 即時資料精簡與擴充重複資料刪除技術。資料壓縮可減少儲存區塊內的空間浪費、重複資料刪除技術可大幅提升有效容量。這適用於本機儲存的資料、以及分層至雲端的資料。
- 最低、最大及可調適的服務品質（AQO）。精細的服務品質（QoS）控制有助於維持高共享環境中關鍵應用程式的效能等級。
- NetApp FabricPool自動將冷資料分層至公有和私有雲端儲存選項、包括Amazon Web Services（AWS）、Azure和NetApp StorageGRID 等儲存解決方案。如需FabricPool 更多有關資訊、請參閱 "TR-4598：FabricPool 最佳實務做法"。

加速並保護資料

提供優異的效能與資料保護、並以下列方式擴充這些功能：ONTAP

- 效能與較低的延遲。以最低的延遲提供最高的處理量。ONTAP
- 資料保護：支援所有平台的通用管理功能、可提供內建的資料保護功能。ONTAP
- NetApp Volume Encryption（NVE）。支援內建和外部金鑰管理、提供原生Volume層級的加密功能。ONTAP
- 多租戶和多因素驗證。支援以最高安全等級共享基礎架構資源。ONTAP

符合未來需求的基礎架構

下列功能可協助滿足嚴苛且不斷變化的業務需求：ONTAP

- 無縫擴充與不中斷營運。支援在不中斷營運的情況下、將容量新增至現有控制器和橫向擴充叢集。ONTAP客戶可以升級至最新技術、例如NVMe和32GB FC、而不需進行昂貴的資料移轉或中斷運作。
- 雲端連線：NetApp是最具雲端連線能力的儲存管理軟體、可在所有公有雲中選擇軟體定義儲存（AI）和雲端原生執行個體（NetApp）ONTAP ONTAP Select Cloud Volumes Service 。
- 與新興應用程式整合。利用支援現有企業應用程式的相同基礎架構、為新一代平台和應用程式提供企業級資料服務、例如自動駕駛車輛、智慧城市和產業4.0。ONTAP

NetApp DataOps工具套件

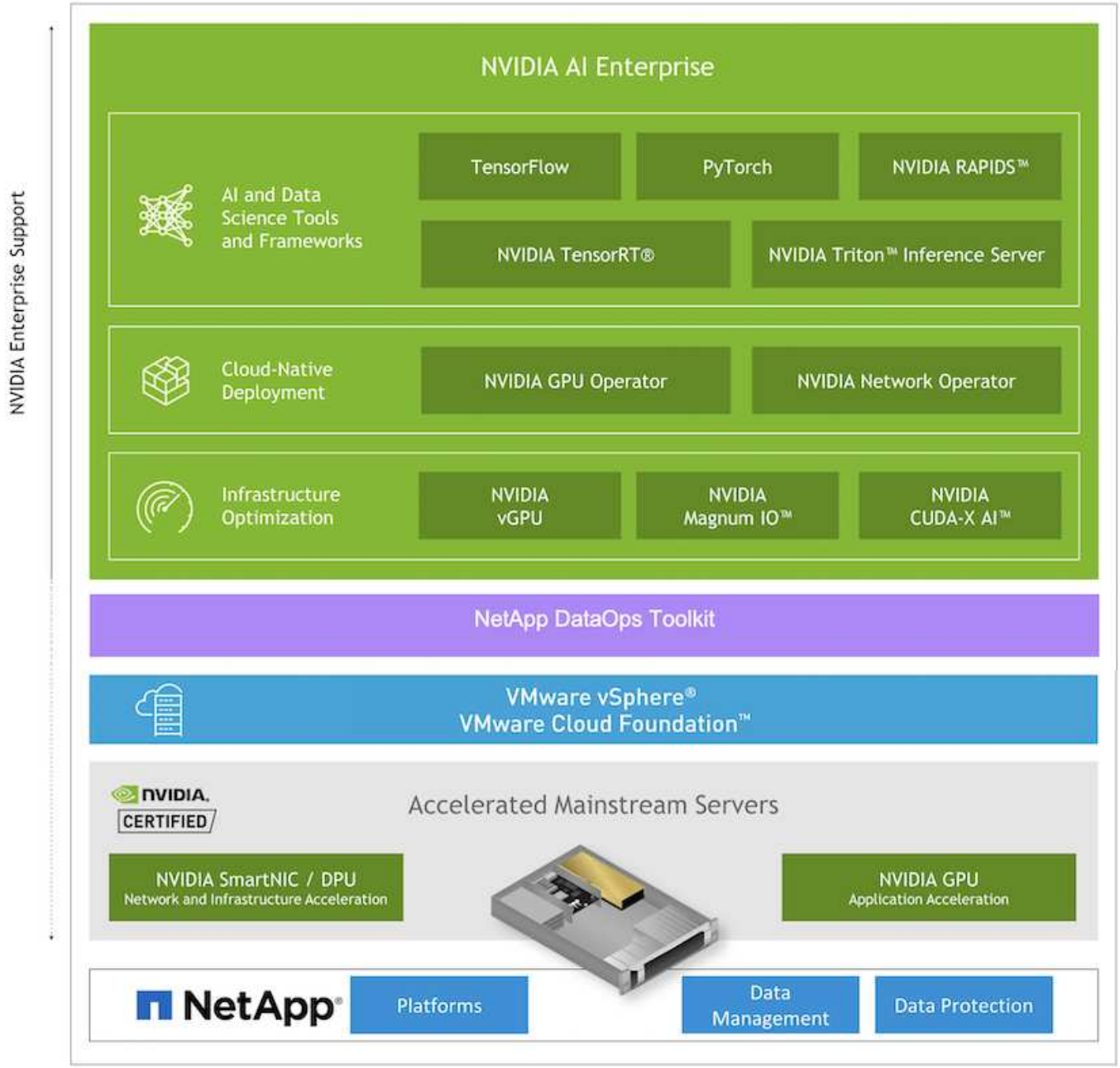
NetApp DataOps Toolkit是一款以Python為基礎的工具、可簡化開發/訓練工作區和推斷伺服器的管理、這些工作區都以高效能橫向擴充的NetApp儲存設備為後盾。主要功能包括：

- 快速配置以高效能橫向擴充NetApp儲存設備為後盾的新高容量JupyterLab工作區。
- 快速配置以企業級NetApp儲存設備為後盾的全新NVIDIA Triton Inference Server執行個體。
- 近乎即時地複製高容量JupyterLab工作區、以進行實驗或快速迭代。
- 近乎即時地儲存高容量JupyterLab工作區的快照、以供備份和/或追蹤/基準化。
- 近乎即時地配置、複製及快照高容量、高效能的資料磁碟區。

架構

本解決方案以NetApp、VMware及NVIDIA認證系統為基礎、打造出備受肯定且熟悉的架構。如需詳細資料、請參閱下表。

元件	詳細資料
AI與資料分析軟體	"NVIDIA AI Enterprise for VMware"
虛擬化平台	"VMware vSphere"
運算平台	"NVIDIA認證系統"
資料管理平台	"NetApp ONTAP"



初始設定

本節說明在NetApp和VMware上使用NVIDIA AI Enterprise時、必須執行的初始設定工作。

先決條件

在您執行本節所述步驟之前、我們假設您已部署VMware vSphere和NetApp ONTAP VMware。請參閱 "NVIDIA AI企業產品支援對照表" 如需支援vSphere版本的詳細資訊、請參閱 "NetApp與VMware解決方案文件" 如需部署VMware vSphere搭配NetApp ONTAP 功能的詳細資訊、

安裝NVIDIA AI Enterprise Host軟體

若要安裝NVIDIA AI Entrprise主機軟體、請依照第1-4節所述的指示進行 "NVIDIA AI企業快速入門指南"。

使用NVIDIA NGC軟體

本節說明在NVIDIA AI Enterprise環境中使用NVIDIA NGC企業軟體所需執行的工作。

設定

本節說明在NVIDIA AI Enterprise環境中使用NVIDIA NGC企業軟體所需執行的初始設定工作。

先決條件

在您執行本節所述步驟之前、我們假設您已依照中所述的指示部署NVIDIA AI Entrprise主機軟體 "初始設定" 頁面。

使用vGPU建立Ubuntu Guest虛擬機器

首先、您必須使用vGPU建立Ubuntu 20.04客體VM。若要使用vGPU建立Ubuntu 20.04客體VM、請遵循中的指示大綱 "NVIDIA AI企業部署指南"。

下載並安裝NVIDIA Guest軟體

接下來、您必須在先前步驟所建立的客體VM中安裝必要的NVIDIA客體軟體。若要在客體VM內下載及安裝所需的NVIDIA客體軟體、請遵循中5.1-5.4節所述的指示 "NVIDIA AI企業快速入門指南"。

> 執行第5.4節所述的驗證工作時、您可能需要使用不同的CUDA Container映像版本標記、因為CUDA Container映像自撰寫指南以來就已更新。在我們的驗證中、我們使用了「nvidia/CUDA：11.0.3-base-ubuntu20.04」。

下載AI /分析架構容器

接下來、您必須從NVIDIA NGC下載所需的AI或分析架構容器映像、以便在您的客體VM中使用。若要在客體VM內下載架構容器、請遵循中所述的指示 "NVIDIA AI企業部署指南"。

安裝及設定NetApp DataOps Toolkit

接下來、您必須在客體VM內安裝適用於傳統環境的NetApp DataOps Toolkit。NetApp DataOps Toolkit可用於直接從ONTAP 客體VM內的終端機、管理您的一套系統上的橫向擴充資料磁碟區。若要在客體VM內安裝NetApp DataOps Toolkit、請執行下列工作。

1. 安裝Pip。

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

2. 登出客體VM終端機、然後重新登入。
3. 設定NetApp DataOps Toolkit。若要完成此步驟、ONTAP 您需要針對您的整套系統提供API存取詳細資料。您可能需要向儲存管理員取得這些資訊。

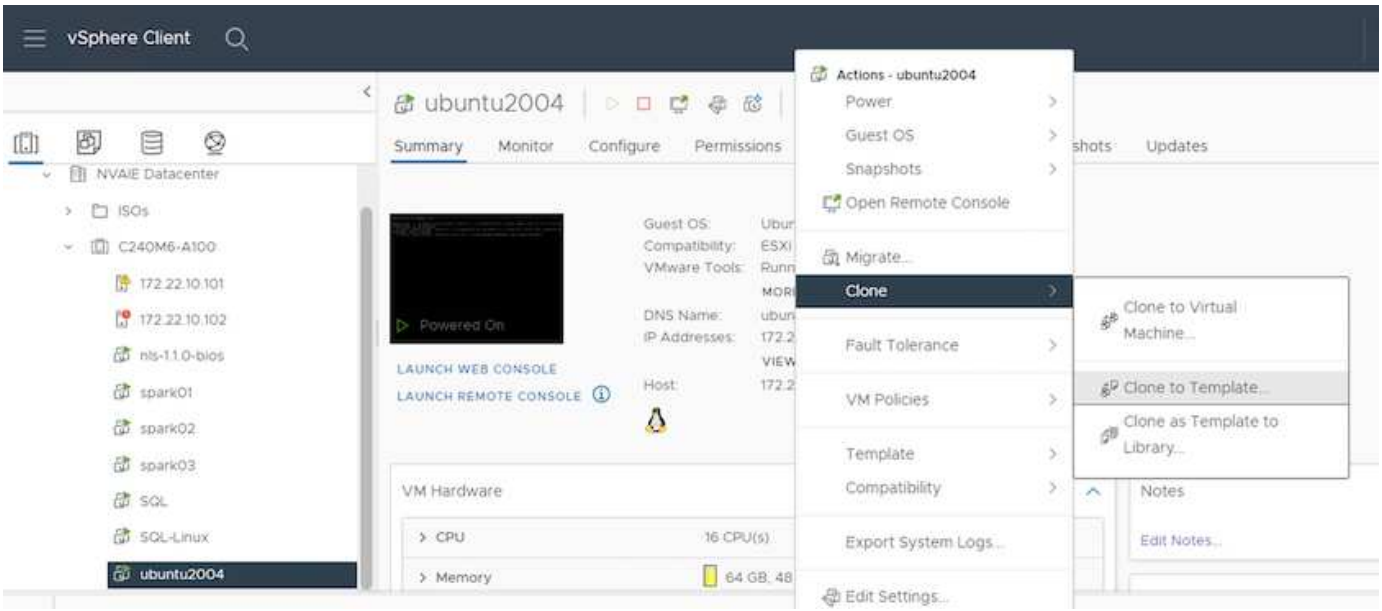
```
$ netapp_dataops_cli.py config

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'.
```

建立來賓**VM**範本

最後、您必須根據客體VM建立VM範本。您可以使用此範本快速建立來賓VM、以使用NVIDIA NGC軟體。

若要根據客體VM建立VM範本、請登入VMware vSphere、按一下客體VM名稱、選擇「Clone（複製）」、選擇「Clone to Template（複製到範本）...」、然後依照精靈進行。



範例使用案例- TensorFlow訓練工作

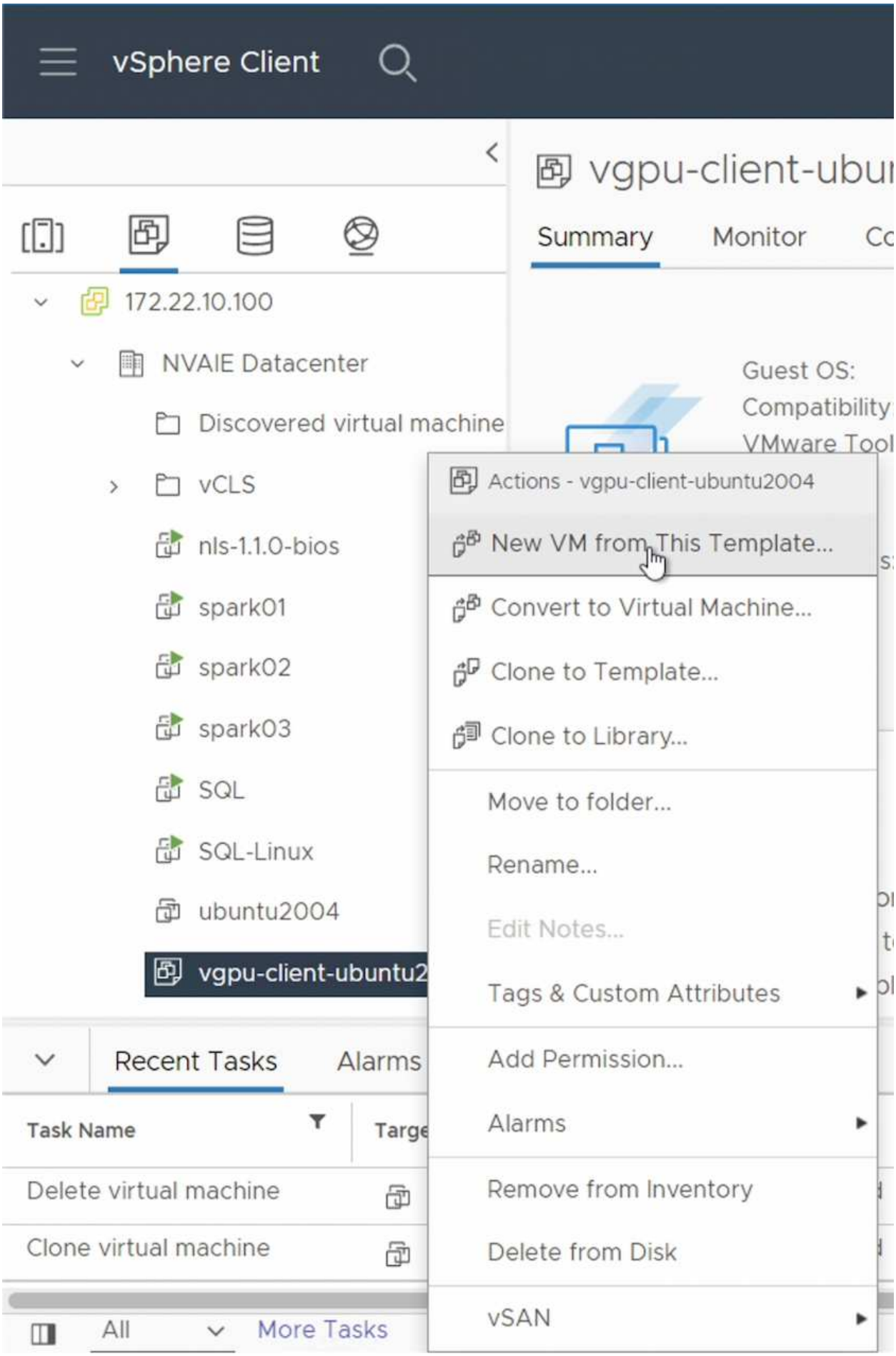
本節說明在NVIDIA AI Enterprise環境中執行TensorFlow訓練工作所需執行的工作。

先決條件

在您執行本節所述步驟之前、我們假設您已依照中所述的指示建立客體VM範本 "設定" 頁面。

從範本建立來賓VM

首先、您必須從上一節建立的範本建立新的來賓VM。若要從範本建立新的來賓VM、請登入VMware vSphere、按一下範本名稱、選擇「New VM from this Template …（從此範本新增VM …）」、然後依照精靈進行。


vSphere Client
vgpu-client-ubu
Summary
Monitor
172.22.10.100
NVAIE Datacenter
Discovered virtual machine
vCLS
nls-1.1.0-bios
spark01
spark02
spark03
SQL
SQL-Linux
ubuntu2004
vgpu-client-ubuntu2
Guest OS:
Compatibility
VMware Tool
Actions - vgpu-client-ubuntu2004
New VM from This Template...
Convert to Virtual Machine...
Clone to Template...
Clone to Library...
Move to folder...
Rename...
Edit Notes...
Tags & Custom Attributes
Add Permission...
Alarms
Remove from Inventory
Delete from Disk
vSAN
Recent Tasks
Alarms
Task Name
Delete virtual machine
Clone virtual machine
All
More Tasks

建立及掛載資料Volume

接下來、您必須建立新的資料量、以便儲存訓練資料集。您可以使用NetApp DataOps Toolkit快速建立新的資料Volume。以下命令範例顯示建立容量為2 TB的名為「imagenNet」的磁碟區。

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

您必須先在客體VM內掛載資料、才能在資料磁碟區中填入資料。您可以使用NetApp DataOps Toolkit快速掛載資料磁碟區。以下命令範例顯示在上一個步驟中建立的磁碟區遠移。

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

填入資料Volume

新磁碟區完成資源配置和掛載之後、即可從來源位置擷取訓練資料集、並放在新磁碟區上。這通常需要從S3或Hadoop資料湖提取資料、有時需要資料工程師提供協助。

執行TensorFlow訓練工作

現在、您已準備好執行TensorFlow訓練工作。若要執行TensorFlow訓練工作、請執行下列工作。

1. 拉出NVIDIA NGC企業級TensorFlow容器映像。

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

2. 啟動NVIDIA NGC企業級TensorFlow容器的執行個體。使用「-v」選項將資料磁碟區附加至容器。

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

3. 在容器內執行TensorFlow訓練方案。以下命令範例顯示執行容器映像所包含的ResNet-50訓練程式範例。

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

何處可找到其他資訊

若要深入瞭解本文件所述資訊、請參閱下列文件和/或網站：

- NetApp ONTAP 數據管理軟體ONTAP —資訊庫

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

- NetApp DataOps工具套件

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

- NVIDIA AI Enterprise搭配VMware

  https://www.nvidia.com/en-us/data-center/products/ai-enterprise/vmware/^]

感謝

- Bobby Oommen、資深NetApp經理
- NetApp系統管理員Ramesh Issac
- NetApp技術行銷工程師Raney Daniel