



NVIDIA AI Enterprise 與 NetApp 和 VMware 合作

NetApp artificial intelligence solutions

NetApp
February 12, 2026

目錄

NVIDIA AI Enterprise 與NetApp和 VMware 合作	1
NVIDIA AI Enterprise 與NetApp和 VMware 合作	1
技術概述	1
NVIDIA AI 企業版	2
NVIDIA GPU 雲端 (NGC)	2
VMware vSphere	2
NetApp ONTAP	2
NetApp DataOps 工具包	3
架構	3
初始設定	4
先決條件	4
安裝NVIDIA AI Enterprise Host 軟體	5
使用NVIDIA NGC 軟體	5
設定	5
範例用例 - TensorFlow 訓練作業	7
如何查找更多信息	9
致謝	9

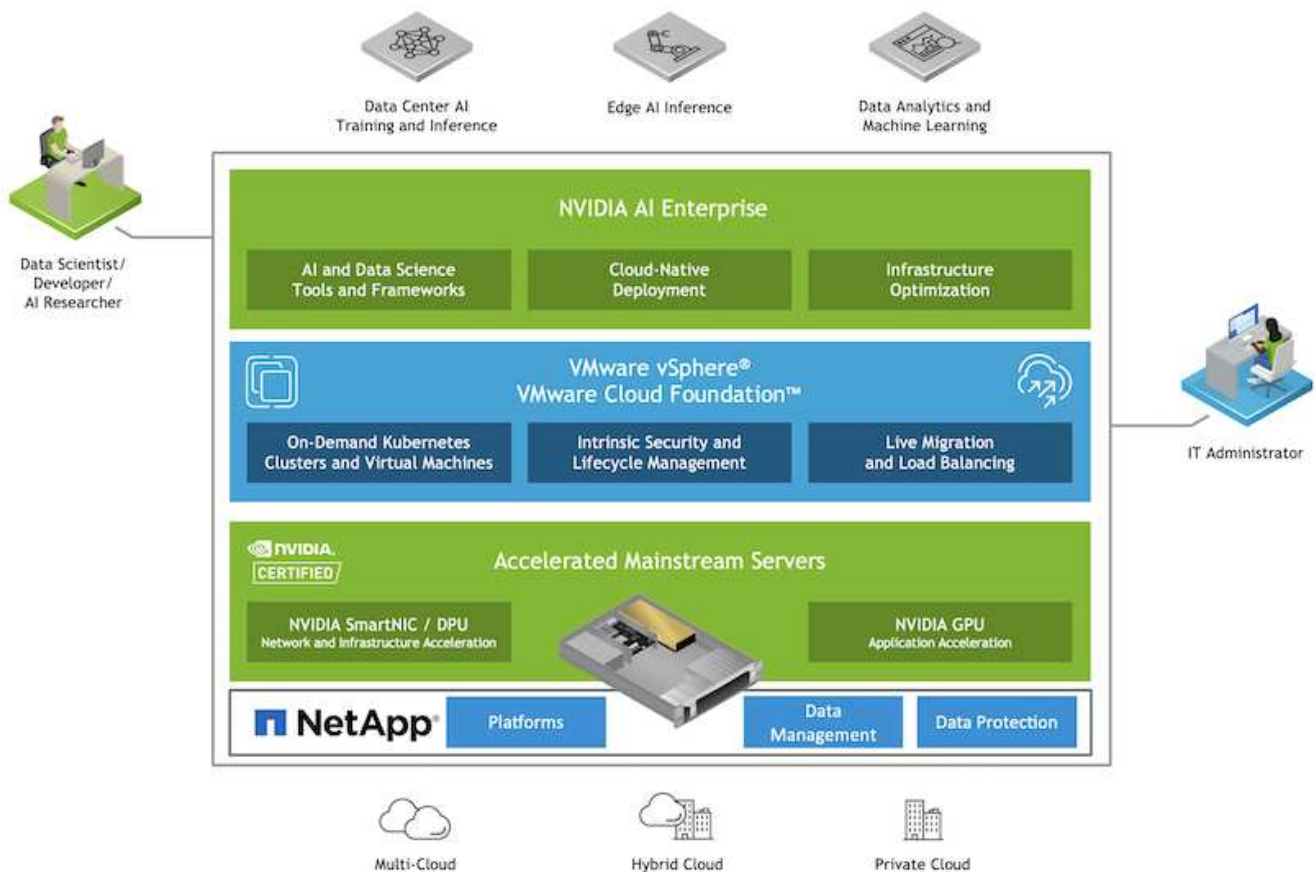
NVIDIA AI Enterprise 與NetApp和 VMware 合作

NVIDIA AI Enterprise 與NetApp和 VMware 合作

NetApp的 Mike Oglesby

對於 IT 架構師和管理員來說，AI 工具可能很複雜且陌生。此外，許多人工智慧平台尚未為企業做好準備。NVIDIA AI Enterprise 由NetApp和 VMware 提供支持，旨在提供精簡的企業級 AI 架構。

NVIDIA AI Enterprise 是一款端到端、雲端原生的 AI 和資料分析軟體套件，經過NVIDIA優化、認證和支持，可在具有NVIDIA認證系統的 VMware vSphere 上運行。該軟體有助於在現代混合雲環境中簡單、快速地部署、管理和擴展 AI 工作負載。由NetApp和 VMware 提供支援的NVIDIA AI Enterprise 以簡化、熟悉的軟體包提供企業級 AI 工作負載和資料管理。



技術概述

本節概述了NVIDIA AI Enterprise 與NetApp和 VMware 的技術。

NVIDIA AI 企業版

NVIDIA AI Enterprise 是一款端到端、雲端原生的 AI 和資料分析軟體套件，經過NVIDIA優化、認證和支持，可在具有NVIDIA認證系統的 VMware vSphere 上運行。該軟體有助於在現代混合雲環境中簡單、快速地部署、管理和擴展 AI 工作負載。

NVIDIA GPU 雲端 (NGC)

NVIDIA NGC 為 AI 從業者提供了一系列針對 GPU 最佳化的軟體，以開發他們的 AI 解決方案。它還提供對各種 AI 服務的訪問，包括用於模型訓練的NVIDIA Base Command、用於部署和監控模型的NVIDIA Fleet Command 以及用於安全存取和管理專有 AI 軟體的 NGC Private Registry。此外，NVIDIA AI Enterprise 客戶可以透過 NGC 入口網站請求支援。

VMware vSphere

VMware vSphere 是 VMware 的虛擬化平台，它將資料中心轉變為包含 CPU、儲存和網路資源的聚合運算基礎架構。vSphere 將這些基礎架構作為統一的操作環境進行管理，並為管理員提供管理該環境中資料中心的工具。

vSphere 的兩個核心元件是 ESXi 和 vCenter Server。ESXi 是管理員創建和運行虛擬機器及虛擬設備的虛擬化平台。vCenter Server 是管理員管理網路中連接的多台主機並池化主機資源的服務。

NetApp ONTAP

ONTAP 9 是NetApp最新一代儲存管理軟體，它支援企業實現基礎架構現代化並過渡到雲端就緒資料中心。ONTAP利用業界領先的數據管理功能，只需一套工具即可管理和保護數據，無論數據位於何處。您也可以將資料自由移動到任何需要的地方：邊緣、核心或雲端。ONTAP 9 包含眾多功能，可簡化資料管理、加速和保護關鍵數據，並支援跨混合雲架構的下一代基礎架構功能。

簡化資料管理

資料管理對於企業 IT 營運和資料科學家至關重要，以便將適當的資源用於 AI 應用程式和訓練 AI/ML 資料集。以下有關NetApp技術的附加資訊超出了本次驗證的範圍，但可能與您的部署相關。

ONTAP資料管理軟體包括以下功能，可簡化操作並降低總營運成本：

- 內聯資料壓縮和擴展重複資料刪除。資料壓縮減少了儲存區塊內部浪費的空間，重複資料刪除顯著增加了有效容量。這適用於本地儲存的資料和分層到雲端的資料。
- 最小、最大和自適應服務品質 (AQoS)。細粒度的服務品質 (QoS) 控制有助於維持高度共享環境中關鍵應用程式的效能水準。
- NetApp FabricPool。提供冷資料到公有和私有雲儲存選項的自動分層，包括 Amazon Web Services (AWS)、Azure 和NetApp StorageGRID儲存解決方案。有關FabricPool的更多信息，請參閱 ["TR-4598：FabricPool最佳實踐"](#)。

加速並保護數據

ONTAP提供卓越等級的效能和資料保護，並透過以下方式擴展這些功能：

- 性能和更低的延遲。ONTAP以盡可能低的延遲提供盡可能高的吞吐量。
- 資料保護。ONTAP提供內建資料保護功能，並在所有平台上提供通用管理。

- NetApp磁碟區加密 (NVE)。ONTAP提供原生磁碟區級加密，同時支援板載和外部金鑰管理。
- 多租戶和多因素身份驗證。ONTAP支援以最高等級的安全性共用基礎架構資源。

面向未來的基礎設施

ONTAP具有以下功能，可協助滿足嚴苛且不斷變化的業務需求：

- 無縫擴展和無中斷運行。ONTAP支援無中斷地向現有控制器和橫向擴展叢集添加容量。客戶可以升級到最新技術，例如 NVMe 和 32Gb FC，而無需昂貴的資料遷移或中斷。
- 雲端連線。ONTAP是與雲端連接最緊密的儲存管理軟體，在所有公有雲中均提供軟體定義儲存（ONTAP Select）和Google Cloud NetApp Volumes Volumes）的選項。
- 與新興應用程式的整合。ONTAP使用支援現有企業應用的相同基礎架構，為下一代平台和應用（如自動駕駛汽車、智慧城市和工業 4.0）提供企業級資料服務。

NetApp DataOps 工具包

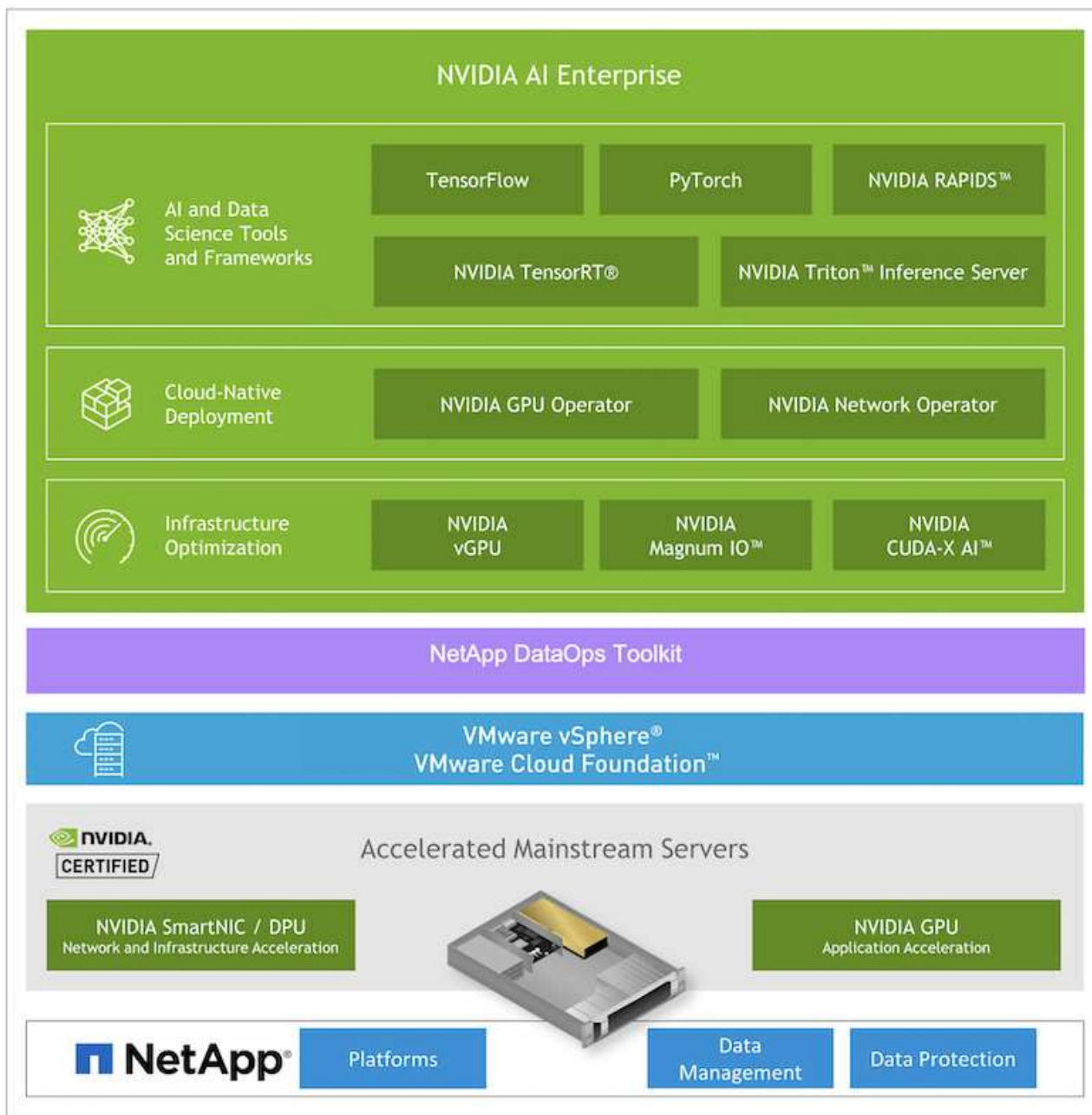
NetApp DataOps Toolkit 是一款基於 Python 的工具，可簡化由高效能、橫向擴展NetApp儲存支援的開發/培訓工作區和推理伺服器的管理。主要功能包括：

- 快速配置由高效能、橫向擴充NetApp儲存支援的新的容量 JupyterLab 工作區。
- 快速配置由企業級NetApp儲存支援的全新NVIDIA Triton 推理伺服器實例。
- 近乎即時地克隆高容量的 JupyterLab 工作區，以實現實驗或快速迭代。
- 幾乎即時地保存高容量 JupyterLab 工作區的快照，以用於備份和/或可追溯性/基準測試。
- 近乎即時地提供、複製和快照大容量、高效能資料磁碟區。

架構

該解決方案建立在經過驗證且熟悉的架構之上，具有NetApp、VMware 和NVIDIA認證的系統。請參閱下表以了解詳情。

成分	細節
人工智慧和數據分析軟體	"適用於 VMware 的NVIDIA AI Enterprise"
虛擬化平台	"VMware vSphere"
計算平台	"NVIDIA認證系統"
資料管理平台	"NetApp ONTAP"



初始設定

本節介紹了將NVIDIA AI Enterprise 與NetApp和 VMware 結合使用所需執行的初始設定任務。

先決條件

在執行本節概述的步驟之前，我們假設您已經部署了 VMware vSphere 和NetApp ONTAP。請參閱["NVIDIA AI Enterprise 產品支援矩陣"](#)有關支援的 vSphere 版本的詳細資訊。請參閱["NetApp和 VMware 解決方案文檔"](#)有關使用NetApp ONTAP部署 VMware vSphere 的詳細資訊。

安裝NVIDIA AI Enterprise Host 軟體

若要安裝NVIDIA AI Enterprise 主機軟體，請依照["NVIDIA AI Enterprise 快速入門指南"](#)。

使用NVIDIA NGC 軟體

設定

本節介紹在NVIDIA AI Enterprise 環境中使用NVIDIA NGC 企業軟體所需執行的初始設定任務。

先決條件

在執行本節中概述的步驟之前，NVIDIA假設您已經按照["初始設定"](#)頁。

使用 vGPU 建立 Ubuntu Guest VM

首先，您必須建立一個具有 vGPU 的 Ubuntu 20.04 客戶虛擬機器。要建立具有 vGPU 的 Ubuntu 20.04 來賓虛擬機，請按照["NVIDIA AI 企業部署指南"](#)。

下載並安裝NVIDIA客戶軟體

接下來，您必須在上一步驟建立的客戶虛擬機器中安裝所需的NVIDIA客戶軟體。若要在客戶虛擬機器中下載並安裝所需的NVIDIA客戶軟體，請依照["NVIDIA AI Enterprise 快速入門指南"](#)。



執行第 5.4 節中概述的驗證任務時，您可能需要使用不同的 CUDA 容器映像版本標籤，因為自編寫指南以來 CUDA 容器映像已更新。在我們的驗證中，我們使用了「nvidia/cuda:11.0.3-base-ubuntu20.04」。

下載 AI/分析框架容器

接下來，您必須從NVIDIA NGC 下載所需的 AI 或分析框架容器映像，以便它們可以在您的客戶虛擬機器中使用。若要在客戶虛擬機器中下載框架容器，請依照["NVIDIA AI 企業部署指南"](#)。

安裝和設定NetApp DataOps 工具包

接下來，您必須在來賓虛擬機器中安裝適用於傳統環境的NetApp DataOps Toolkit。NetApp DataOps Toolkit 可用於直接從客戶虛擬機器內的終端機管理ONTAP系統上的橫向擴充資料磁碟區。若要在客戶虛擬機器中安裝NetApp DataOps Toolkit，請執行下列任務。

1. 安裝 pip。

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

2. 登出客戶 VM 終端，然後重新登入。

3. 配置NetApp DataOps 工具包。為了完成此步驟，您將需要ONTAP系統的 API 存取詳細資訊。您可能需要從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
```

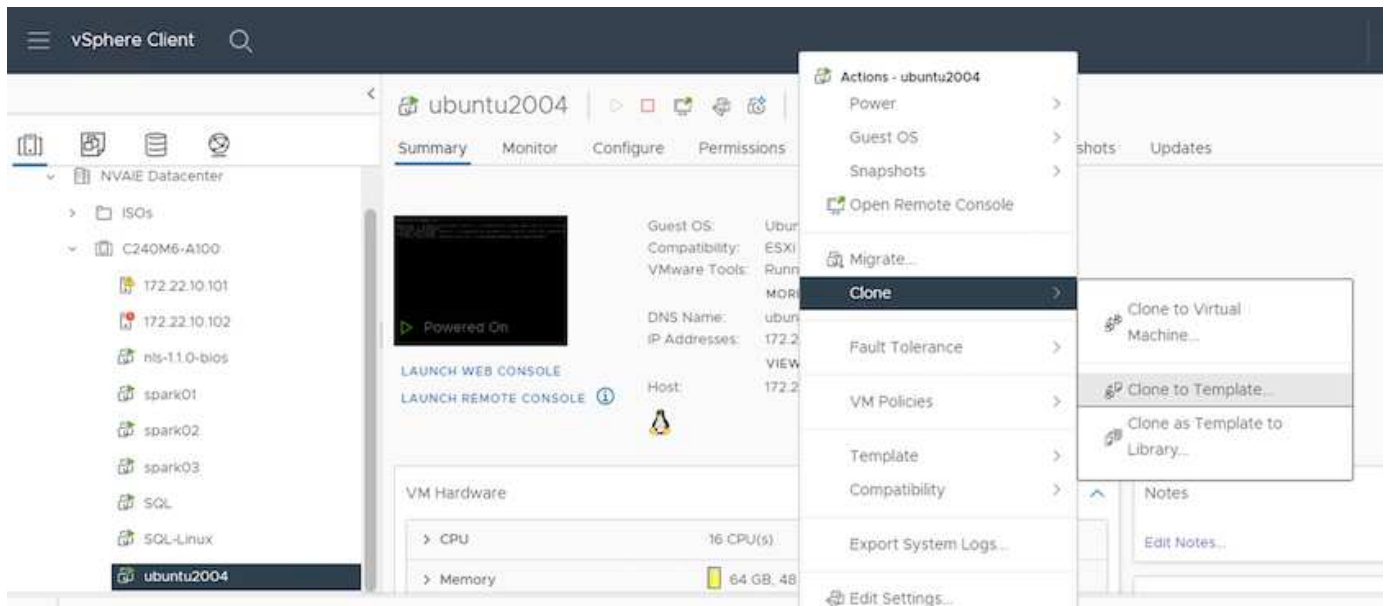
```
Do you intend to use this toolkit to push/pull from S3? (yes/no): no
```

```
Created config file: '/home/user/.netapp_dataops/config.json'.
```

建立來賓虛擬機器模板

最後，您必須根據您的客戶虛擬機器建立虛擬機器範本。您將能夠使用此範本快速建立用於利用NVIDIA NGC軟體的客戶虛擬機器。

若要根據您的來賓 VM 建立 VM 模板，請登入 VMware vSphere，右鍵單擊來賓 VM 名稱，選擇“複製”，選擇“複製到模板...”，然後按照精靈進行操作。



範例用例 - TensorFlow 訓練作業

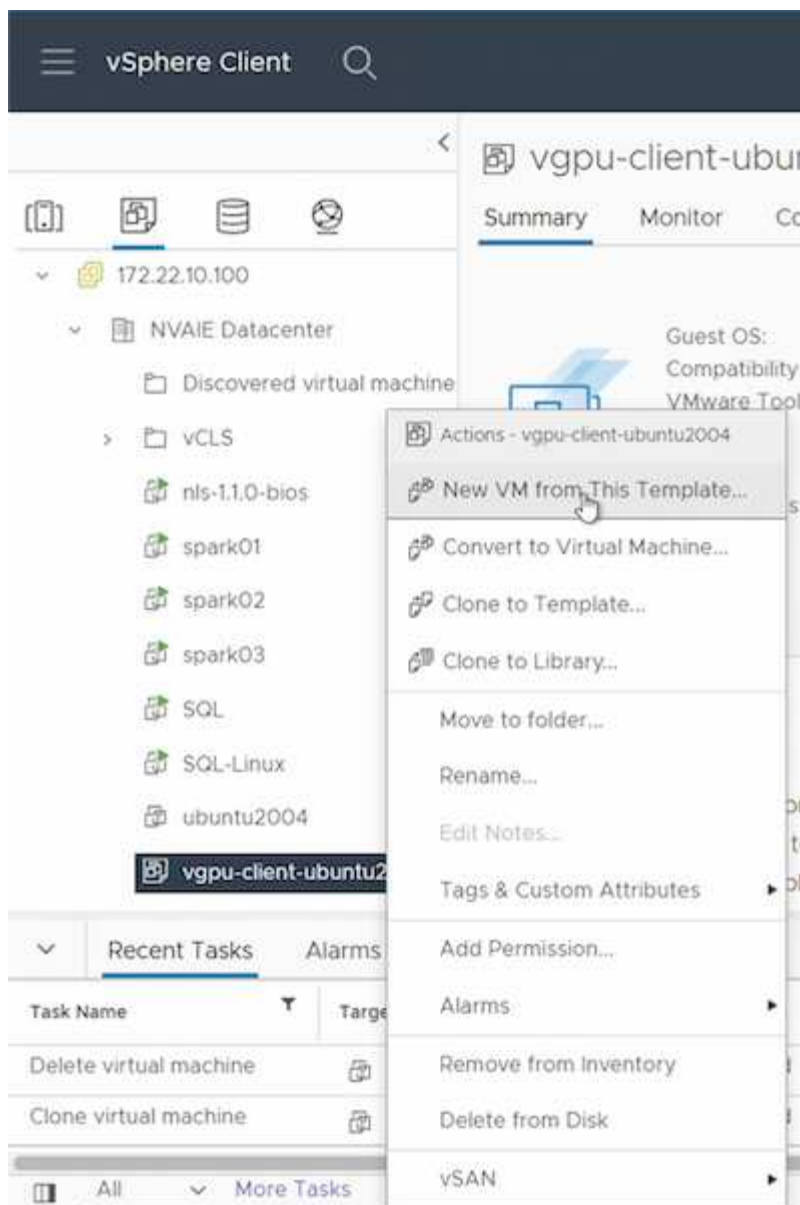
本節介紹在NVIDIA AI Enterprise 環境中執行 TensorFlow 訓練作業所需執行的任務。

先決條件

在執行本節中概述的步驟之前，我們假設您已經按照["設定"](#)頁。

從模板建立來賓虛擬機

首先，您必須根據上一節中建立的範本建立新的客戶虛擬機器。若要從範本建立新的客戶虛擬機，請登入 VMware vSphere，右鍵單擊範本名稱，選擇“從此範本新虛擬機器...”，然後按照精靈操作。



建立並掛載資料卷

接下來，您必須建立一個新的資料卷來儲存您的訓練資料集。您可以使用NetApp DataOps Toolkit 快速建立新的資料卷。下面的範例指令顯示建立一個名為「imagenet」、容量為 2 TB 的磁碟區。

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

在您用資料填入資料磁碟區之前，您必須將其安裝在來賓虛擬機器中。您可以使用NetApp DataOps Toolkit 快速安裝資料卷。下面的範例指令顯示了上一個步驟中所建立的磁碟區的安裝。

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

填充數據量

在配置並安裝新磁碟區後，可以從來源位置擷取訓練資料集並將其放置在新磁碟區上。這通常涉及從 S3 或 Hadoop 數據湖中提取數據，有時還需要數據工程師的幫助。

執行 TensorFlow 訓練作業

現在，您已準備好執行 TensorFlow 訓練作業。若要執行 TensorFlow 訓練作業，請執行下列任務。

1. 拉取NVIDIA NGC 企業 TensorFlow 容器鏡像。

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

2. 啟動NVIDIA NGC 企業 TensorFlow 容器的執行個體。使用“-v”選項將資料卷附加到容器。

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

3. 在容器內執行您的 TensorFlow 訓練程序。下面的範例指令展示了容器鏡像中所包含的範例 ResNet-50 訓練程式的執行。

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

如何查找更多信息

要了解有關本文檔中描述的信息的更多信息，請參閱以下文檔和/或網站：

- NetApp ONTAP資料管理軟體 — ONTAP資訊庫

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

- NetApp DataOps 工具包

<https://github.com/NetApp/netapp-dataops-toolkit>

- NVIDIA AI Enterprise 與 VMware

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

致謝

- Bobby Oommen，NetApp資深經理
- Ramesh Isaac，NetApp系統管理員

- NetApp技術行銷工程師 Roney Daniel

版權資訊

Copyright © 2026 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。