NetApp 및 VMware를 활용한 NVIDIA AI Enterprise

NetApp artificial intelligence solutions

NetApp
August 18, 2025

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

목차

NetApp 및 VMware를 활용한 NVIDIA AI Enterprise . . . 1
NetApp 및 VMware를 활용한 NVIDIA AI Enterprise . . . 1
기술 개요 . . . 1
NVIDIA AI 엔터프라이즈 . . . 2
NVIDIA GPU 클라우드(NGC) . . . 2
VMware vSphere . . . 2
NetApp ONTAP . . . 2
NetApp DataOps 툴킷 . . . 3
아키텍처 . . . 3
초기 설정 . . . 4
필수 조건 . . . 4
NVIDIA AI Enterprise 호스트 소프트웨어 설치 . . . 5
NVIDIA NGC 소프트웨어를 사용하세요 . . . 5
설정 . . . 5
예시 사용 사례 - TensorFlow 학습 작업 . . . 7
추가 정보를 찾을 수 있는 곳 . . . 9
감사의 말 . . . 10

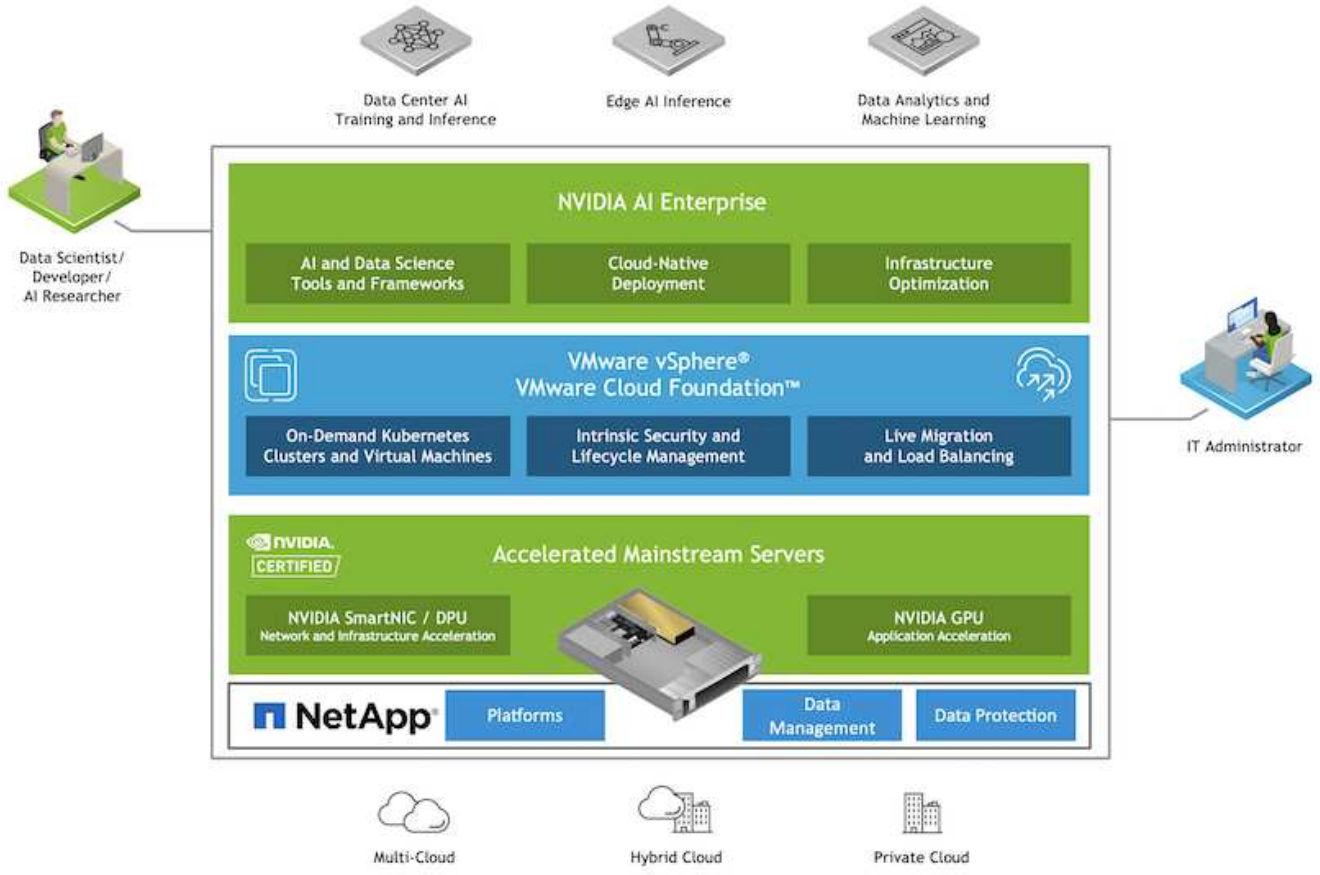
NetApp 및 VMware를 활용한 NVIDIA AI Enterprise

NetApp 및 VMware를 활용한 NVIDIA AI Enterprise

마이크 오글스비, NetApp

IT 설계자와 관리자에게 AI 툴은 복잡하고 익숙하지 않을 수 있습니다. 게다가 많은 AI 플랫폼은 기업에 적합하지 않습니다. NetApp 과 VMware 기반의 NVIDIA AI Enterprise는 간소화된 엔터프라이즈급 AI 아키텍처를 제공하기 위해 만들어졌습니다.

NVIDIA AI Enterprise는 NVIDIA 인증 시스템을 통해 VMware vSphere에서 실행되도록 NVIDIA 에서 최적화, 인증 및 지원하는 엔드 투 엔드 클라우드 기반 AI 및 데이터 분석 소프트웨어 제품군입니다. 이 소프트웨어는 최신 하이브리드 클라우드 환경에서 AI 워크로드를 간단하고 빠르게 배포, 관리, 확장할 수 있도록 지원합니다. NetApp 과 VMware 기반의 NVIDIA AI Enterprise는 간소화되고 친숙한 패키지로 엔터프라이즈급 AI 워크로드와 데이터 관리를 제공합니다.



기술 개요

이 섹션에서는 NetApp 및 VMware를 기반으로 한 NVIDIA AI Enterprise에 대한 기술 개요를 제공합니다.

NVIDIA AI 엔터프라이즈

NVIDIA AI Enterprise는 NVIDIA 인증 시스템을 통해 VMware vSphere에서 실행되도록 NVIDIA 에서 최적화, 인증 및 지원하는 엔드 투 엔드 클라우드 기반 AI 및 데이터 분석 소프트웨어 제품군입니다. 이 소프트웨어는 최신 하이브리드 클라우드 환경에서 AI 워크로드를 간단하고 빠르게 배포, 관리, 확장할 수 있도록 지원합니다.

NVIDIA GPU 클라우드(NGC)

NVIDIA NGC는 AI 실무자가 AI 솔루션을 개발할 수 있도록 GPU 최적화 소프트웨어 카탈로그를 호스팅합니다. 또한 모델 학습을 위한 NVIDIA Base Command, 모델 배포 및 모니터링을 위한 NVIDIA Fleet Command, 독점 AI 소프트웨어에 안전하게 액세스하고 관리하기 위한 NGC Private Registry 등 다양한 AI 서비스에 대한 액세스를 제공합니다. 또한 NVIDIA AI Enterprise 고객은 NGC 포털을 통해 지원을 요청할 수 있습니다.

VMware vSphere

VMware vSphere는 VMware의 가상화 플랫폼으로, 데이터 센터를 CPU, 스토리지 및 네트워킹 리소스를 포함하는 통합 컴퓨팅 인프라로 변환합니다. vSphere는 이러한 인프라를 통합 운영 환경으로 관리하고, 관리자에게 해당 환경에 참여하는 데이터 센터를 관리할 수 있는 도구를 제공합니다.

vSphere의 두 가지 핵심 구성 요소는 ESXi와 vCenter Server입니다. ESXi는 관리자가 가상 머신과 가상 어플라이언스를 생성하고 실행하는 가상화 플랫폼입니다. vCenter Server는 관리자가 네트워크에 연결된 여러 호스트를 관리하고 호스트 리소스를 풀링하는 서비스입니다.

NetApp ONTAP

NetApp 의 최신 스토리지 관리 소프트웨어인 ONTAP 9를 사용하면 기업이 인프라를 현대화하고 클라우드 지원 데이터 센터로 전환할 수 있습니다. ONTAP 업계 최고의 데이터 관리 역량을 활용하여 데이터가 어디에 있든 단일 도구 세트를 사용하여 데이터를 관리하고 보호할 수 있도록 지원합니다. 또한 필요한 곳, 즉 엣지, 코어, 클라우드로 데이터를 자유롭게 이동할 수 있습니다. ONTAP 9에는 데이터 관리를 간소화하고, 중요 데이터를 가속화하고 보호하며, 하이브리드 클라우드 아키텍처 전반에서 차세대 인프라 기능을 구현하는 다양한 기능이 포함되어 있습니다.

데이터 관리 간소화

적절한 리소스가 AI 애플리케이션과 AI/ML 데이터 세트 교육에 사용될 수 있도록 기업 IT 운영과 데이터 과학자에게 데이터 관리가 매우 중요합니다. NetApp 기술에 대한 다음 추가 정보는 이 검증 범위를 벗어나지만 배포에 따라 관련이 있을 수 있습니다.

ONTAP 데이터 관리 소프트웨어에는 다음과 같은 기능이 포함되어 있어 운영을 간소화하고 단순화하며 총 운영 비용을 절감할 수 있습니다.

- 인라인 데이터 압축 및 확장된 중복 제거. 데이터 압축은 저장 블록 내부의 낭비되는 공간을 줄이고, 중복 제거는 효과적인 용량을 크게 증가시킵니다. 이는 로컬에 저장된 데이터와 클라우드에 계층화된 데이터 모두에 적용됩니다.
- 최소, 최대 및 적응형 서비스 품질(AQoS). 세분화된 서비스 품질(QoS) 제어는 공유 빈도가 높은 환경에서 중요한 애플리케이션의 성능 수준을 유지하는 데 도움이 됩니다.
- NetApp FabricPool. Amazon Web Services(AWS), Azure, NetApp StorageGRID 스토리지 솔루션을 포함한 퍼블릭 및 프라이빗 클라우드 스토리지 옵션에 콜드 데이터의 자동 계층화를 제공합니다. FabricPool 에 대한 자세한 내용은 다음을 참조하세요. "TR-4598: FabricPool 모범 사례" .

데이터 가속화 및 보호

ONTAP 뛰어난 수준의 성능과 데이터 보호 기능을 제공하며 다음과 같은 방식으로 이러한 기능을 확장합니다.

- 성능과 낮은 지연 시간. ONTAP 가능한 가장 낮은 지연 시간으로 가능한 가장 높은 처리량을 제공합니다.
- 데이터 보호. ONTAP 모든 플랫폼에서 공통적으로 관리할 수 있는 내장형 데이터 보호 기능을 제공합니다.
- NetApp 볼륨 암호화(NVE). ONTAP 온보드 및 외부 키 관리 지원을 통해 기본 볼륨 수준 암호화를 제공합니다.
- 다중 테넌시 및 다중 요소 인증. ONTAP 최고 수준의 보안을 통해 인프라 리소스를 공유할 수 있도록 합니다.

미래 지향적 인프라

ONTAP 다음과 같은 기능을 통해 까다롭고 끊임없이 변화하는 비즈니스 요구 사항을 충족하는 데 도움이 됩니다.

- 원활한 확장과 중단 없는 운영. ONTAP 기존 컨트롤러와 확장형 클러스터에 중단 없이 용량을 추가할 수 있도록 지원합니다. 고객은 비용이 많이 드는 데이터 마이그레이션이나 중단 없이 NVMe 및 32Gb FC와 같은 최신 기술로 업그레이드할 수 있습니다.
- 클라우드 연결. ONTAP 모든 퍼블릭 클라우드에서 소프트웨어 정의 스토리지(ONTAP Select)와 클라우드 기반 인스턴스(Google Cloud NetApp Volumes)에 대한 옵션을 제공하는 가장 클라우드에 연결된 스토리지 관리 소프트웨어입니다.
- 새로운 애플리케이션과의 통합. ONTAP 기존 엔터프라이즈 앱을 지원하는 동일한 인프라를 사용하여 자율주행차, 스마트 시티, 산업 4.0과 같은 차세대 플랫폼과 애플리케이션을 위한 엔터프라이즈급 데이터 서비스를 제공합니다.

NetApp DataOps 툴킷

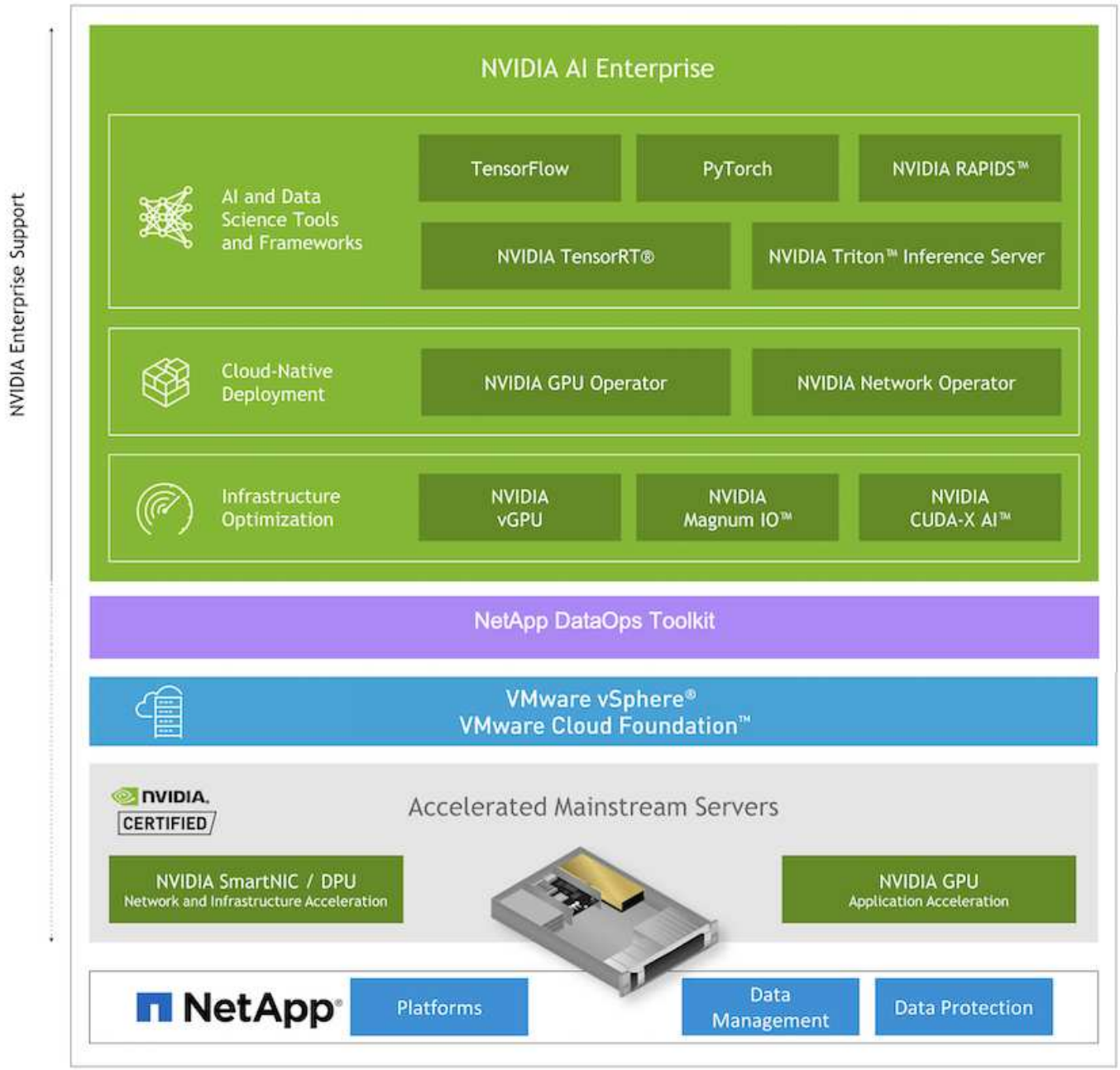
NetApp DataOps Toolkit은 고성능, 확장형 NetApp 스토리지에 의해 지원되는 개발/교육 작업 공간과 추론 서버의 관리를 간소화하는 Python 기반 도구입니다. 주요 기능은 다음과 같습니다.

- 고성능, 확장형 NetApp 스토리지로 지원되는 새로운 대용량 JupyterLab 작업 공간을 빠르게 프로비저닝하세요.
- 엔터프라이즈급 NetApp 스토리지로 지원되는 새로운 NVIDIA Triton Inference Server 인스턴스를 빠르게 프로비저닝하세요.
- 실험이나 빠른 반복을 가능하게 하기 위해 대용량 JupyterLab 작업 공간을 거의 즉각적으로 복제합니다.
- 대용량 JupyterLab 작업 공간의 스냅샷을 거의 즉각적으로 저장하여 백업 및/또는 추적/기준 설정을 할 수 있습니다.
- 거의 즉각적으로 대용량, 고성능 데이터 볼륨을 프로비저닝, 복제 및 스냅샷합니다.

아키텍처

이 솔루션은 NetApp, VMware, NVIDIA 인증 시스템을 갖춘 검증되고 친숙한 아키텍처를 기반으로 구축되었습니다. 자세한 내용은 다음 표를 참조하세요.

요소	세부
AI 및 데이터 분석 소프트웨어	"VMware용 NVIDIA AI Enterprise"
가상화 플랫폼	"VMware vSphere"
컴퓨팅 플랫폼	"NVIDIA 인증 시스템"
데이터 관리 플랫폼	"NetApp ONTAP"



초기 설정

이 섹션에서는 NetApp 및 VMware에서 NVIDIA AI Enterprise를 활용하기 위해 수행해야 하는 초기 설정 작업에 대해 설명합니다.

필수 조건

이 섹션에 설명된 단계를 수행하기 전에 VMware vSphere와 NetApp ONTAP 이미 배포했다고 가정합니다. 를 참조하세요"NVIDIA AI Enterprise 제품 지원 매트릭스" 지원되는 vSphere 버전에 대한 자세한 내용은 다음을 참조하세요. 를 참조하세요"NetApp 및 VMware 솔루션 설명서" NetApp ONTAP 과 함께 VMware vSphere를 배포하는 방법에 대한 자세한 내용은 다음을 참조하세요.

NVIDIA AI Enterprise 호스트 소프트웨어 설치

NVIDIA AI Entrprise 호스트 소프트웨어를 설치하려면 섹션 1-4에 설명된 지침을 따르십시오."NVIDIA AI Enterprise 빠른 시작 가이드" .

NVIDIA NGC 소프트웨어를 사용하세요

설정

이 섹션에서는 NVIDIA AI Enterprise 환경 내에서 NVIDIA NGC Enterprise 소프트웨어를 활용하기 위해 수행해야 하는 초기 설정 작업에 대해 설명합니다.

필수 조건

이 섹션에 설명된 단계를 수행하기 전에 이미 NVIDIA AI Entrprise 호스트 소프트웨어를 배포했다고 가정합니다."초기 설정" 페이지.

vGPU를 사용하여 Ubuntu 게스트 VM 만들기

먼저, vGPU가 있는 Ubuntu 20.04 게스트 VM을 만들어야 합니다. vGPU를 사용하여 Ubuntu 20.04 게스트 VM을 생성하려면 다음 지침을 따르세요."NVIDIA AI 엔터프라이즈 배포 가이드" .

NVIDIA 게스트 소프트웨어 다운로드 및 설치

다음으로, 이전 단계에서 만든 게스트 VM 내에 필요한 NVIDIA 게스트 소프트웨어를 설치해야 합니다. 게스트 VM 내에서 필요한 NVIDIA 게스트 소프트웨어를 다운로드하고 설치하려면 섹션 5.1-5.4에 설명된 지침을 따르세요 ."NVIDIA AI Enterprise 빠른 시작 가이드" .

> 섹션 5.4에 설명된 검증 작업을 수행할 때, 가이드 작성 이후 CUDA 컨테이너 이미지가 업데이트되었으므로 다른 CUDA 컨테이너 이미지 버전 태그를 사용해야 할 수 있습니다. 우리는 검증 과정에서 'nvidia/cuda:11.0.3-base-ubuntu20.04'를 사용했습니다.

AI/분석 프레임워크 컨테이너 다운로드

다음으로, 게스트 VM에서 사용할 수 있도록 NVIDIA NGC에서 필요한 AI 또는 분석 프레임워크 컨테이너 이미지를 다운로드해야 합니다. 게스트 VM 내에서 프레임워크 컨테이너를 다운로드하려면 다음 지침을 따르세요."NVIDIA AI 엔터프라이즈 배포 가이드" .

NetApp DataOps 툴킷 설치 및 구성

다음으로, 게스트 VM 내에 NetApp DataOps Toolkit for Traditional Environment를 설치해야 합니다. NetApp DataOps Toolkit을 사용하면 게스트 VM 내의 터미널에서 직접 ONTAP 시스템의 확장형 데이터 볼륨을 관리할 수 있습니다. 게스트 VM 내에 NetApp DataOps Toolkit을 설치하려면 다음 작업을 수행하세요.

1. pip를 설치합니다.

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

2. 게스트 VM 터미널에서 로그아웃한 다음 다시 로그인합니다.
3. NetApp DataOps Toolkit을 구성합니다. 이 단계를 완료하려면 ONTAP 시스템에 대한 API 액세스 세부 정보가 필요합니다. 스토리지 관리자로부터 해당 정보를 얻어야 할 수도 있습니다.

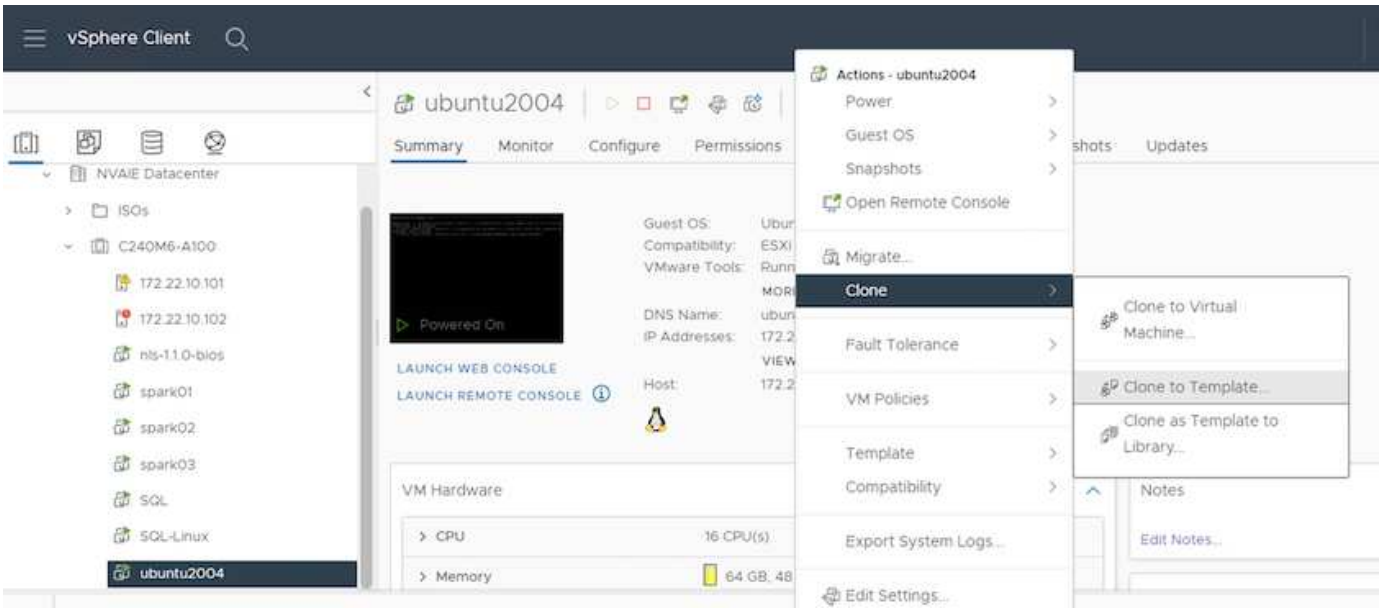
```
$ netapp_dataops_cli.py config

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'.
```

게스트 **VM** 템플릿 만들기

마지막으로 게스트 VM을 기반으로 VM 템플릿을 만들어야 합니다. 이 템플릿을 사용하면 NVIDIA NGC 소프트웨어를 활용하는 게스트 VM을 빠르게 만들 수 있습니다.

게스트 VM을 기반으로 VM 템플릿을 생성하려면 VMware vSphere에 로그인하고 게스트 VM 이름을 마우스 오른쪽 버튼으로 클릭한 다음 '복제'를 선택하고 '템플릿에 복제…'를 선택한 다음 마법사를 따르세요.



예시 사용 사례 - **TensorFlow** 학습 작업

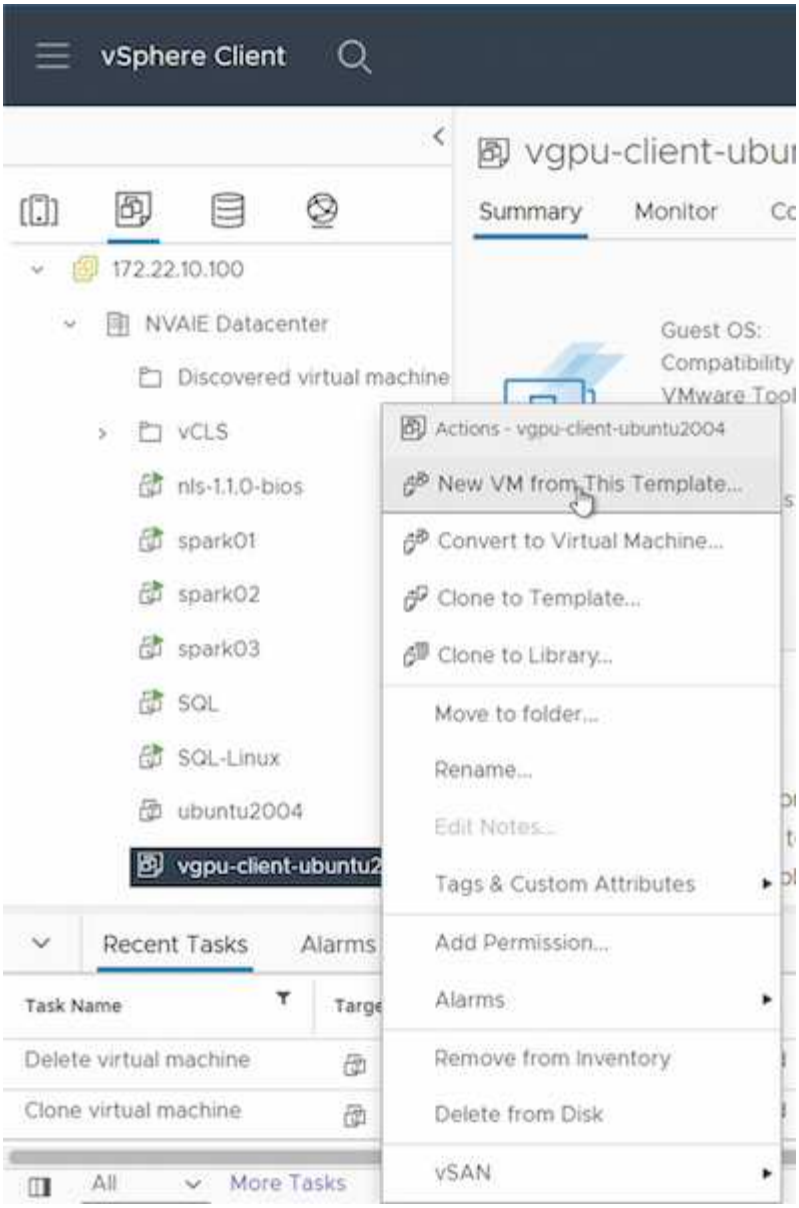
이 섹션에서는 NVIDIA AI Enterprise 환경 내에서 TensorFlow 학습 작업을 실행하기 위해 수행해야 하는 작업에 대해 설명합니다.

필수 조건

이 섹션에 설명된 단계를 수행하기 전에 이미 다음 지침에 따라 게스트 VM 템플릿을 생성했다고 가정합니다."설정" 페이지.

템플릿에서 게스트 **VM** 만들기

먼저, 이전 섹션에서 만든 템플릿에서 새 게스트 VM을 만들어야 합니다. 템플릿에서 새 게스트 VM을 만들려면 VMware vSphere에 로그인하고 템플릿 이름을 마우스 오른쪽 버튼으로 클릭한 다음 '이 템플릿에서 새 VM...'을 선택하고 마법사를 따르세요.



데이터 볼륨 생성 및 마운트

다음으로, 훈련 데이터 세트를 저장할 새로운 데이터 볼륨을 만들어야 합니다. NetApp DataOps Toolkit을 사용하면 새로운 데이터 볼륨을 빠르게 만들 수 있습니다. 다음 예제 명령은 2TB 용량의 'imagenet'이라는 볼륨을 생성하는 것을 보여줍니다.

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

데이터 볼륨에 데이터를 채우려면 먼저 게스트 VM 내에서 해당 볼륨을 마운트해야 합니다. NetApp DataOps Toolkit을 사용하면 데이터 볼륨을 빠르게 마운트할 수 있습니다. 다음 예제 명령은 이전 단계에서 생성된 볼륨의 마운트를 보여줍니다.

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

데이터 볼륨 채우기

새 볼륨이 프로비저닝되고 마운트된 후에는 소스 위치에서 교육 데이터 세트를 검색하여 새 볼륨에 배치할 수 있습니다. 일반적으로 이 작업에는 S3 또는 Hadoop 데이터 레이크에서 데이터를 가져오는 작업이 포함되며, 때로는 데이터 엔지니어의 도움이 필요하기도 합니다.

TensorFlow 학습 작업 실행

이제 TensorFlow 학습 작업을 실행할 준비가 되었습니다. TensorFlow 학습 작업을 실행하려면 다음 작업을 수행하세요.

1. NVIDIA NGC 엔터프라이즈 TensorFlow 컨테이너 이미지를 가져옵니다.

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

2. NVIDIA NGC 엔터프라이즈 TensorFlow 컨테이너 인스턴스를 실행합니다. 컨테이너에 데이터 볼륨을 연결하려면 '-v' 옵션을 사용하세요.

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

3. 컨테이너 내에서 TensorFlow 학습 프로그램을 실행합니다. 다음 예제 명령은 컨테이너 이미지에 포함된 ResNet-50 학습 프로그램의 실행을 보여줍니다.

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

추가 정보를 찾을 수 있는 곳

이 문서에 설명된 정보에 대해 자세히 알아보려면 다음 문서 및/또는 웹사이트를 참조하세요.

- NetApp ONTAP 데이터 관리 소프트웨어 - ONTAP 정보 라이브러리

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

- NetApp DataOps 툴킷

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

- VMware를 활용한 NVIDIA AI Enterprise

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

감사의 말

- Bobby Oommen, NetApp 선임 관리자
- Ramesh Isaac, NetApp 시스템 관리자
- Roney Daniel, NetApp 기술 마케팅 엔지니어

저작권 정보

Copyright © 2025 NetApp, Inc. All Rights Reserved. 미국에서 인쇄된 본 문서의 어떠한 부분도 저작권 소유자의 사전 서면 승인 없이는 어떠한 형식이나 수단(복사, 녹음, 녹화 또는 전자 검색 시스템에 저장하는 것을 비롯한 그래픽, 전자적 또는 기계적 방법)으로도 복제될 수 없습니다.

NetApp이 저작권을 가진 자료에 있는 소프트웨어에는 아래의 라이센스와 고지사항이 적용됩니다.

본 소프트웨어는 NetApp에 의해 '있는 그대로' 제공되며 상품성 및 특정 목적에의 적합성에 대한 명시적 또는 묵시적 보증을 포함하여(이에 제한되지 않음) 어떠한 보증도 하지 않습니다. NetApp은 대체품 또는 대체 서비스의 조달, 사용 불능, 데이터 손실, 이익 손실, 영업 중단을 포함하여(이에 국한되지 않음), 이 소프트웨어의 사용으로 인해 발생하는 모든 직접 및 간접 손해, 우발적 손해, 특별 손해, 징벌적 손해, 결과적 손해의 발생에 대하여 그 발생 이유, 책임론, 계약 여부, 엄격한 책임, 불법 행위(과실 또는 그렇지 않은 경우)와 관계없이 어떠한 책임도 지지 않으며, 이와 같은 손실의 발생 가능성이 통지되었다 하더라도 마찬가지입니다.

NetApp은 본 문서에 설명된 제품을 언제든지 예고 없이 변경할 권리를 보유합니다. NetApp은 NetApp의 명시적인 서면 동의를 받은 경우를 제외하고 본 문서에 설명된 제품을 사용하여 발생하는 어떠한 문제에도 책임을 지지 않습니다. 본 제품의 사용 또는 구매의 경우 NetApp에서는 어떠한 특허권, 상표권 또는 기타 지적 재산권이 적용되는 라이센스도 제공하지 않습니다.

본 설명서에 설명된 제품은 하나 이상의 미국 특허, 해외 특허 또는 출원 중인 특허로 보호됩니다.

제한적 권리 표시: 정부에 의한 사용, 복제 또는 공개에는 DFARS 252.227-7013(2014년 2월) 및 FAR 52.227-19(2007년 12월)의 기술 데이터-비상업적 품목에 대한 권리(Rights in Technical Data -Noncommercial Items) 조항의 하위 조항 (b)(3)에 설명된 제한사항이 적용됩니다.

여기에 포함된 데이터는 상업용 제품 및/또는 상업용 서비스(FAR 2.101에 정의)에 해당하며 NetApp, Inc.의 독점 자산입니다. 본 계약에 따라 제공되는 모든 NetApp 기술 데이터 및 컴퓨터 소프트웨어는 본질적으로 상업용이며 개인 비용만으로 개발되었습니다. 미국 정부는 데이터가 제공된 미국 계약과 관련하여 해당 계약을 지원하는 데에만 데이터에 대한 전 세계적으로 비독점적이고 양도할 수 없으며 재사용이 불가능하며 취소 불가능한 라이센스를 제한적으로 가집니다. 여기에 제공된 경우를 제외하고 NetApp, Inc.의 사전 서면 승인 없이는 이 데이터를 사용, 공개, 재생산, 수정, 수행 또는 표시할 수 없습니다. 미국 국방부에 대한 정부 라이센스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

NETAPP, NETAPP 로고 및 http://www.netapp.com/TM에 나열된 마크는 NetApp, Inc.의 상표입니다. 기타 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다.