



Kubeflow

NetApp Solutions

NetApp
May 10, 2024

목차

Kubeflow	1
Kubeflow 구축	1
Kubeflow 작업 및 작업 예	2

Kubeflow

Kubeflow 구축

이 섹션에서는 Kubernetes 클러스터에 Kubeflow를 구축하기 위해 완료해야 하는 작업에 대해 설명합니다.

필수 구성 요소

이 섹션에 요약된 배포 연습을 수행하기 전에 이미 다음 작업을 수행했다고 가정합니다.

1. 이미 작동 중인 Kubernetes 클러스터가 있으며, 구축하려는 Kubeflow 버전에서 지원하는 Kubernetes 버전을 실행하고 있습니다. 지원되는 Kubernetes 버전 목록은 [에서 Kubeflow 버전에 대한 종속성을 참조하십시오 "Kubeflow 공식 문서"](#).
2. Kubernetes 클러스터에 NetApp Astra Trident를 이미 설치하고 구성했습니다. Astra Trident에 대한 자세한 내용은 [를 참조하십시오 "Astra Trident 문서"](#).

기본 Kubernetes StorageClass를 설정합니다

Kubeflow를 구축하기 전에 Kubernetes 클러스터 내에 기본 StorageClass를 지정하는 것이 좋습니다. Kubeflow 구축 프로세스에서 기본 StorageClass를 사용하여 새 영구 볼륨 프로비저닝을 시도할 수 있습니다. 기본 StorageClass 로 지정된 StorageClass 가 없으면 배포가 실패할 수 있습니다. 클러스터 내에서 기본 StorageClass를 지정하려면 배포 점프 호스트에서 다음 작업을 수행합니다. 클러스터 내에서 기본 StorageClass를 이미 지정한 경우에는 이 단계를 건너뛸 수 있습니다.

1. 기존 StorageClasses 중 하나를 기본 StorageClass로 지정합니다. 다음 명령 예에서는 이름이 인 StorageClass 를 지정했습니다 `ontap-ai-flexvols-retain` 기본 StorageClass로 사용됩니다.



ONTAP-NAS-Flexgroup Trident 백엔드 유형은 PVC 크기가 매우 큼니다. 기본적으로 Kubeflow는 크기가 몇 GB인 PVC를 프로비저닝하려고 시도합니다. 따라서 Kubeflow 구축을 위해 "ONTAP-NAS-flexgroup" 백엔드 유형을 기본 StorageClass로 사용하는 StorageClass를 지정할 수 없습니다.

```

$ kubectl get sc
NAME                                PROVISIONER                        AGE
ontap-ai-flexgroups-retain          csi.trident.netapp.io             25h
ontap-ai-flexgroups-retain-iface1   csi.trident.netapp.io             25h
ontap-ai-flexgroups-retain-iface2   csi.trident.netapp.io             25h
ontap-ai-flexvols-retain             csi.trident.netapp.io             3s
$ kubectl patch storageclass ontap-ai-flexvols-retain -p '{"metadata": {"annotations":{"storageclass.kubernetes.io/is-default-class":"true"}}}'
storageclass.storage.k8s.io/ontap-ai-flexvols-retain patched
$ kubectl get sc
NAME                                PROVISIONER                        AGE
ontap-ai-flexgroups-retain          csi.trident.netapp.io             25h
ontap-ai-flexgroups-retain-iface1   csi.trident.netapp.io             25h
ontap-ai-flexgroups-retain-iface2   csi.trident.netapp.io             25h
ontap-ai-flexvols-retain (default)  csi.trident.netapp.io             54s

```

Kubeflow 구축 옵션

Kubeflow를 배포하는 다양한 옵션이 있습니다. 을 참조하십시오 ["Kubeflow 공식 문서"](#) 구축 옵션 목록을 확인하고 요구사항에 가장 적합한 옵션을 선택하십시오.

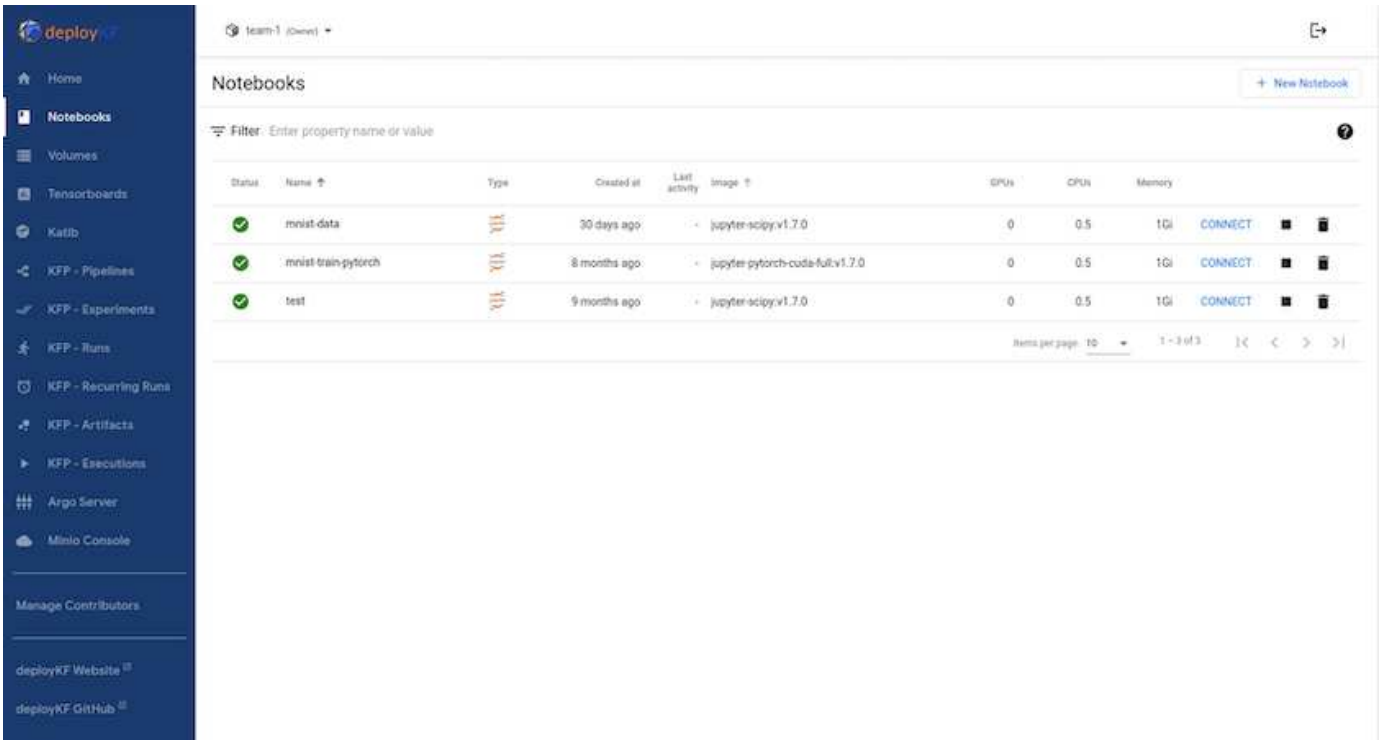


검증을 위해 를 사용하여 Kubeflow 1.7을 배포했습니다 ["구축KF" 0.1.1.](#)

Kubeflow 작업 및 작업 예

데이터 과학자 또는 개발자 사용을 위한 **Jupyter Notebook Workspace**를 제공합니다

Kubeflow는 새로운 Jupyter Notebook 서버를 신속하게 프로비저닝하여 데이터 과학자 작업 공간 역할을 할 수 있습니다. Kubeflow 컨텍스트 내의 Jupyter Notebooks에 대한 자세한 내용은 를 참조하십시오 ["Kubeflow 공식 문서"](#).



Kubeflow와 함께 NetApp DataOps 툴킷 사용

를 클릭합니다 "[Kubernetes용 NetApp 데이터 과학 툴킷](#)" Kubeflow와 함께 사용할 수 있습니다. Kubeflow와 함께 NetApp Data Science Toolkit을 사용하면 다음과 같은 이점이 있습니다.

- 데이터 과학자는 Jupyter Notebook 내에서 직접 스냅샷 및 클론 생성과 같은 고급 NetApp 데이터 관리 작업을 수행할 수 있습니다.
- 스냅샷 및 클론 생성과 같은 고급 NetApp 데이터 관리 작업은 Kubeflow 파이프라인 프레임워크를 사용하여 자동화된 워크플로에 통합할 수 있습니다.

을 참조하십시오 "[Kubeflow 예](#)" Kubeflow 기반 툴킷 사용에 대한 자세한 내용은 NetApp Data Science Toolkit GitHub 리포지토리를 참조하십시오.

워크플로우 예 - Kubeflow 및 NetApp DataOps 툴킷을 사용하여 이미지 인식 모델을 학습합니다

이 섹션에서는 Kubeflow 및 NetApp DataOps 툴킷을 사용하여 이미지 인식을 위한 신경망 교육 및 배포에 관련된 단계를 설명합니다. 이 슬라이드의 목적은 NetApp 스토리지를 통합하는 교육 작업을 보여주는 예입니다.

필수 구성 요소

Kubeflow 파이프라인 내에서 기차 및 테스트 단계에 사용할 필수 구성으로 Dockerfile을 생성합니다. 다음은 Dockerfile의 예입니다.

```

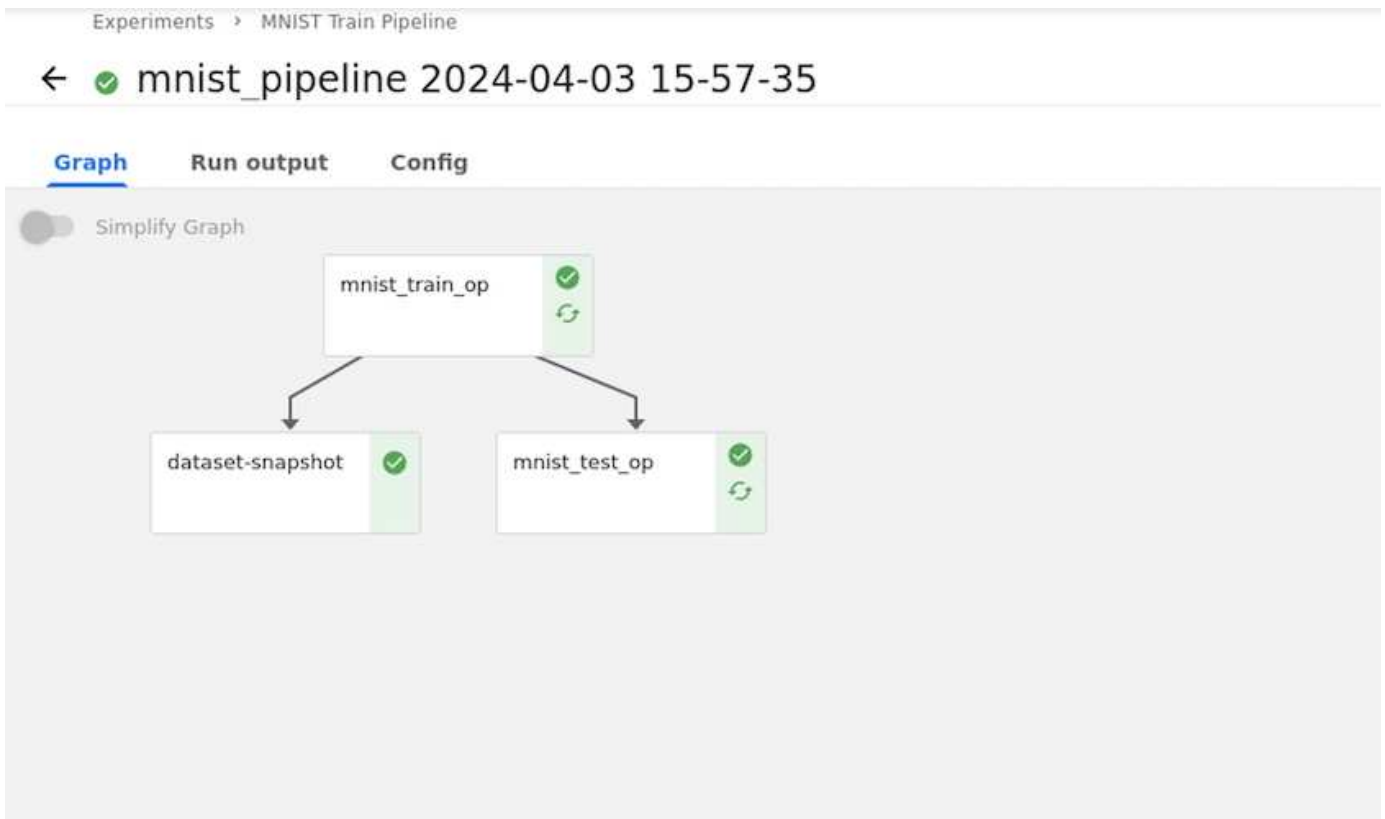
FROM pytorch/pytorch:latest
RUN pip install torchvision numpy scikit-learn matplotlib tensorboard
WORKDIR /app
COPY . /app
COPY train_mnist.py /app/train_mnist.py
CMD ["python", "train_mnist.py"]

```

요구 사항에 따라 프로그램을 실행하는 데 필요한 모든 필수 라이브러리와 패키지를 설치합니다. 기계 학습 모델을 교육하기 전에 이미 작동하는 KubeFlow 배포가 있다고 가정합니다.

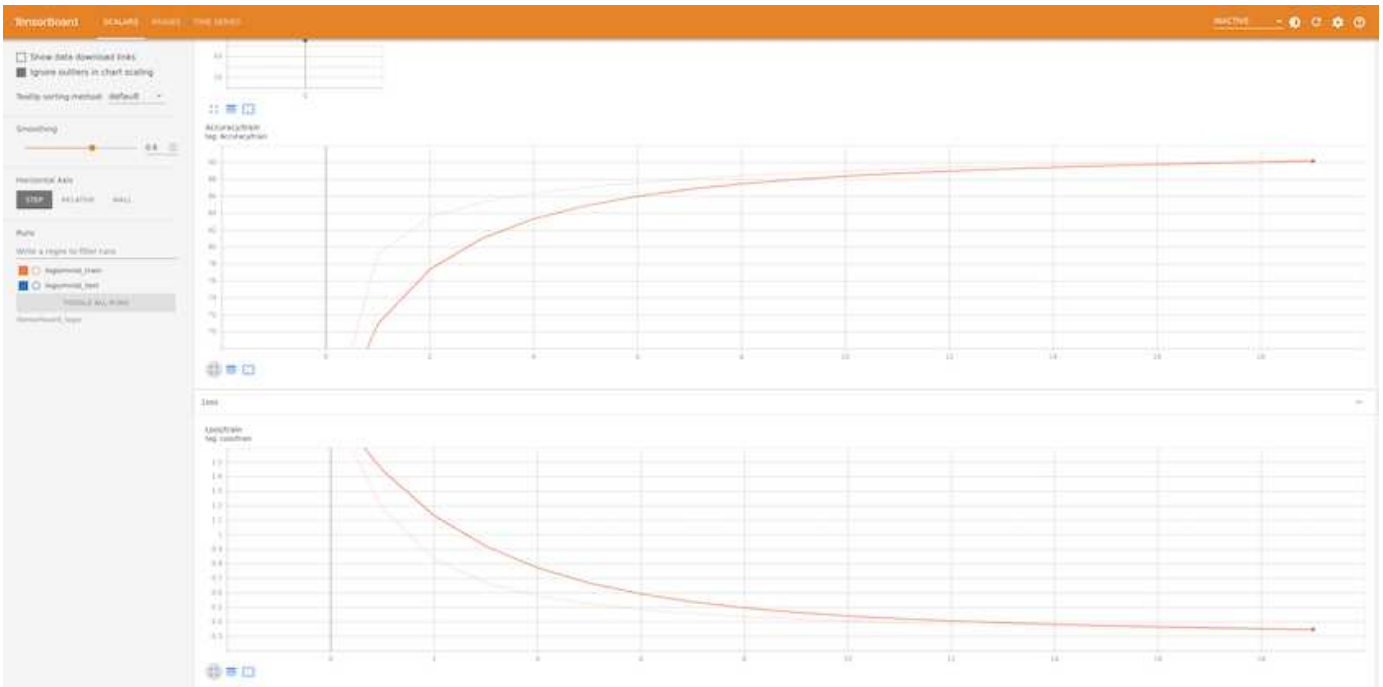
PyTorch 및 **KubeFlow** 파이프라인을 사용하여 **MNIST** 데이터에 대한 소규모 **NN**을 훈련하십시오

MNIST 데이터에 대해 훈련된 소규모 Neural Network의 예를 사용합니다. MNIST 데이터 세트는 0-9의 숫자 자필 이미지로 구성됩니다. 이미지 크기는 28x28픽셀입니다. 데이터 세트는 60,000개의 기차 이미지와 10,000개의 검증 이미지로 구분됩니다. 이 실험에 사용되는 신경망은 2계층 피드포워드 네트워크입니다. 교육은 KubeFlow 파이프라인을 사용하여 실행됩니다. 설명서를 참조하십시오 ["여기"](#) 를 참조하십시오. KubeFlow 파이프라인은 사전 요구 사항 섹션의 Docker 이미지를 통합합니다.



Tensorboard를 사용하여 결과를 시각화합니다

모델이 훈련되면 Tensorboard를 사용하여 결과를 시각화할 수 있습니다. ["Tensorboard를 사용합니다"](#) 은 KubeFlow Dashboard에서 기능으로 제공됩니다. 작업에 대한 사용자 정의 텐서보드를 작성할 수 있습니다. 아래 예는 과(와) 비교 트레이닝 정확도의 플롯을 보여줍니다 기 수 및 교육 손실 대 Epoch 수.



Katib를 사용하여 Hyperparameters를 실험합니다

"케이티비주식회사"은 KubeFlow 내의 도구로서 모델 hyperparameters를 실험하는 데 사용할 수 있습니다. 실험을 생성하려면 먼저 원하는 메트릭/목표를 정의합니다. 이것은 일반적으로 테스트 정확도입니다. 메트릭이 정의되면 재생할 하이퍼파라미터(optimizer/learning_rate/layer 수)를 선택합니다. Katib는 사용자 정의 값으로 hyperparameter sweep을 수행하여 원하는 메트릭을 만족하는 최적의 매개 변수 조합을 찾습니다. UI의 각 섹션에서 이러한 매개 변수를 정의할 수 있습니다. 또는 필요한 사양으로 *YAML* 파일을 정의할 수도 있습니다. 다음은 Katib 실험의 그림입니다.

The screenshot shows the 'Experiment details' page in the Kubeflow UI. The page is titled 'Experiment details' and includes a 'DELETE' button. The configuration is as follows:

Section	Parameter	Value
Objective	Name	Validation-accuracy
	Type	maximize
	Goal	0.9
	Additional metrics	Train-accuracy
Trials	Max failed trials	3
	Max trials	12
	Parallel trials	3
	Parameters	
Parameters	lr	Parameter type: double, Min: 0.01, Max: 0.03
	num-layers	Parameter type: int, Min: 1, Max: 64
	optimizer	Parameter type: categorical, sgd, adam, trl
Algorithm	Name	grid
	Metrics collector	
Metrics collector	Collector type	File

The screenshot shows the Kubeflow Katib interface. On the left is a navigation sidebar with options like Home, Notebooks, Volumes, Tensorboards, Katib, KFP - Pipelines, KFP - Experiments, KFP - Runs, KFP - Recurring Runs, KFP - Artifacts, KFP - Executions, Argo Server, and Minio Console. The main panel is titled 'Experiment details' and shows a message: 'Couldn't find any successful Trial'. Below this is a table with columns: OVERVIEW, TRIALS, DETAILS, and YAML. The table contains the following information:

OVERVIEW	TRIALS	DETAILS	YAML
Name	mnist-pytorch		
Status	🕒 Experiment is running		
Best trial	No optimal trial yet		
Best trial's params	No optimal trial yet		
Best trial performance			
User defined goal	Validation-accuracy > 0.9		
Running trials	3		
Failed trials	0		
Succeeded trials	0		

Below the table is a section for 'Experiment Conditions' and a 'Filter' input field.

NetApp 스�냅샷을 사용하여 추적을 위해 데이터를 저장합니다

모델 훈련 중에 추적 기능을 위해 훈련 데이터 세트의 스�냅샷을 저장하고자 할 수 있습니다. 이를 위해 아래와 같이 파이프라인에 스�냅샷 단계를 추가할 수 있습니다. 스�냅샷을 생성하기 위해 [NetApp DataOps 툴킷](#)을 사용할 수 있습니다.

```
@dsl.pipeline(
  name = 'MNIST Classification Pipeline',
  description = 'Train a simple NN for classification'
)
def mnist_pipeline():
  mnist_train_task = mnist_train_op()
  mnist_train_task.apply(
    kfp.onprem.mount_pvc('mnist-data', 'mnist-data-vol', '/mnt/data/')
  )

  mnist_test_task = mnist_test_op()
  mnist_test_task.apply(
    kfp.onprem.mount_pvc('mnist-data', 'mnist-data-vol', '/mnt/data/')
  )

  volume_snapshot_name = "mnist-pytorch-snapshot"
  dataset_snapshot = dsl.ContainerOp(
    name="dataset-snapshot",
    image="python:3.9",
    command=["/bin/bash", "-c"],
    arguments=["\n
      python3 -m pip install netapp-dataops-k8s && \
      echo "" + volume_snapshot_name + "" > /volume_snapshot_name.txt && \
      netapp_dataops_k8s_cli.py create volume-snapshot --pvc-name=" + "mnist-data" + " --snapshot-name=" + str(volume_snapshot_name) + " --namespace={work[0,namespace]}",
      file_outputs["volume_snapshot_name": "/volume_snapshot_name.txt"]
    ]
  )
  mnist_test_task.after(mnist_train_task)
  dataset_snapshot.after(mnist_train_task)
```

을 참조하십시오 ["Kubeflow에 대한 NetApp DataOps 툴킷의 예"](#) 를 참조하십시오.

저작권 정보

Copyright © 2024 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 라이선스와 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 라이선스도 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 라이선스를 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 라이선스는 DFARS 조항 252.227-7015(b)(2014년 2월)에 명시된 권한으로 제한됩니다.

상표 정보

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