



Kubeflow

NetApp Solutions

NetApp
May 10, 2024

This PDF was generated from https://docs.netapp.com/it-it/netapp-solutions/ai/aicp_kubeflow_deployment_overview.html on May 10, 2024. Always check docs.netapp.com for the latest.

Sommario

- Kubeflow 1
- Implementazione di Kubeflow 1
- Operazioni e attività Kubeflow di esempio 2

Kubeflow

Implementazione di Kubeflow

In questa sezione vengono descritte le attività da completare per implementare Kubeflow nel cluster Kubernetes.

Prerequisiti

Prima di eseguire l'esercizio di implementazione descritto in questa sezione, si presuppone che siano già state eseguite le seguenti attività:

1. Disponi già di un cluster Kubernetes funzionante e stai eseguendo una versione di Kubernetes supportata dalla versione di Kubeflow che intendi implementare. Per un elenco delle versioni di Kubernetes supportate, fare riferimento alle dipendenze per la versione di Kubeflow in ["Documentazione ufficiale del Kubeflow"](#).
2. Hai già installato e configurato NetApp Astra Trident nel tuo cluster Kubernetes. Per ulteriori informazioni su Astra Trident, fare riferimento alla ["Documentazione di Astra Trident"](#).

Impostare la classe di storage Kubernetes predefinita

Prima di implementare Kubeflow, ti consigliamo di indicare una classe storage predefinita all'interno del cluster Kubernetes. Il processo di implementazione di Kubeflow può tentare di eseguire il provisioning di nuovi volumi persistenti tramite StorageClass predefinito. Se nessuna StorageClass è designata come StorageClass predefinita, l'implementazione potrebbe non riuscire. Per designare una StorageClass predefinita all'interno del cluster, eseguire la seguente attività dall'host di distribuzione jump. Se è già stata designata una StorageClass predefinita all'interno del cluster, è possibile saltare questo passaggio.

1. Designare uno dei StorageClasses esistenti come StorageClass predefinito. I comandi di esempio che seguono mostrano la designazione di StorageClass denominata `ontap-ai-flexvols-retain` Come StorageClass di default.



Il `ontap-nas-flexgroup` Il tipo di backend Trident ha una dimensione minima del PVC che è abbastanza grande. Per impostazione predefinita, Kubeflow tenta di eseguire il provisioning di PVC di dimensioni limitate a poche GB. Pertanto, non è necessario designare un StorageClass che utilizzi `ontap-nas-flexgroup` Tipo di backend come StorageClass predefinito ai fini dell'implementazione di Kubeflow.

```

$ 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

```

Opzioni di implementazione di Kubeflow

Ci sono molte opzioni diverse per implementare Kubeflow. Fare riferimento a ["Documentazione ufficiale del Kubeflow"](#) per un elenco delle opzioni di distribuzione e scegliere l'opzione più adatta alle proprie esigenze.

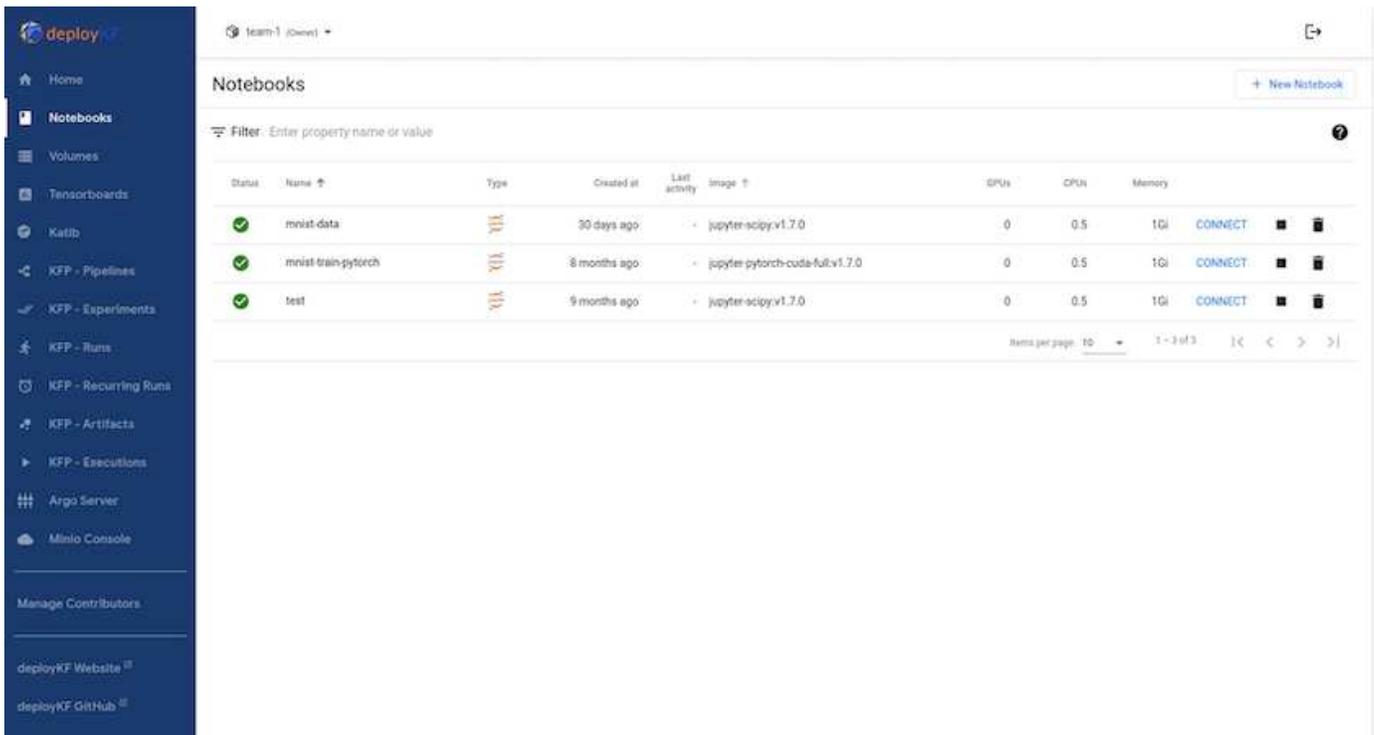


A scopo di convalida, abbiamo implementato Kubeflow 1,7 utilizzando ["DeployKF" 0,1.1](#).

Operazioni e attività Kubeflow di esempio

Provisioning di un'area di lavoro Jupyter notebook per l'utilizzo da parte di Data Scientist o Developer

Kubeflow è in grado di eseguire rapidamente il provisioning dei nuovi server Jupyter notebook per agire come aree di lavoro per scienziati dei dati. Per ulteriori informazioni sui notebook Jupyter all'interno del contesto Kubeflow, vedere ["Documentazione ufficiale del Kubeflow"](#).



USA il toolkit DataOps di NetApp con Kubeflow

Il ["NetApp Data Science Toolkit per Kubernetes"](#) Utilizzabile in combinazione con Kubeflow. L'utilizzo del NetApp Data Science Toolkit con Kubeflow offre i seguenti vantaggi:

- I data scientist possono eseguire operazioni avanzate di gestione dei dati NetApp, come la creazione di snapshot e cloni, direttamente dall'interno di un notebook Jupyter.
- Le operazioni avanzate di gestione dei dati di NetApp, come la creazione di snapshot e cloni, possono essere incorporate nei workflow automatizzati utilizzando il framework Kubeflow Pipelines.

Fare riferimento a ["Esempi di Kubeflow"](#) Sezione all'interno del repository GitHub del NetApp Data Science Toolkit per informazioni dettagliate sull'utilizzo del toolkit con Kubeflow.

Esempio di flusso di lavoro - Traduci un modello di riconoscimento delle immagini utilizzando Kubeflow e il toolkit DataOps di NetApp

In questa sezione vengono descritte le fasi della formazione e dell'implementazione di una rete neurale per il riconoscimento delle immagini utilizzando Kubeflow e il toolkit NetApp DataOps. Lo scopo è quello di mostrare un lavoro di formazione che incorpora lo storage NetApp.

Prerequisiti

Creare un Dockerfile con le configurazioni necessarie da utilizzare per i passaggi di treno e test all'interno della pipeline Kubeflow.

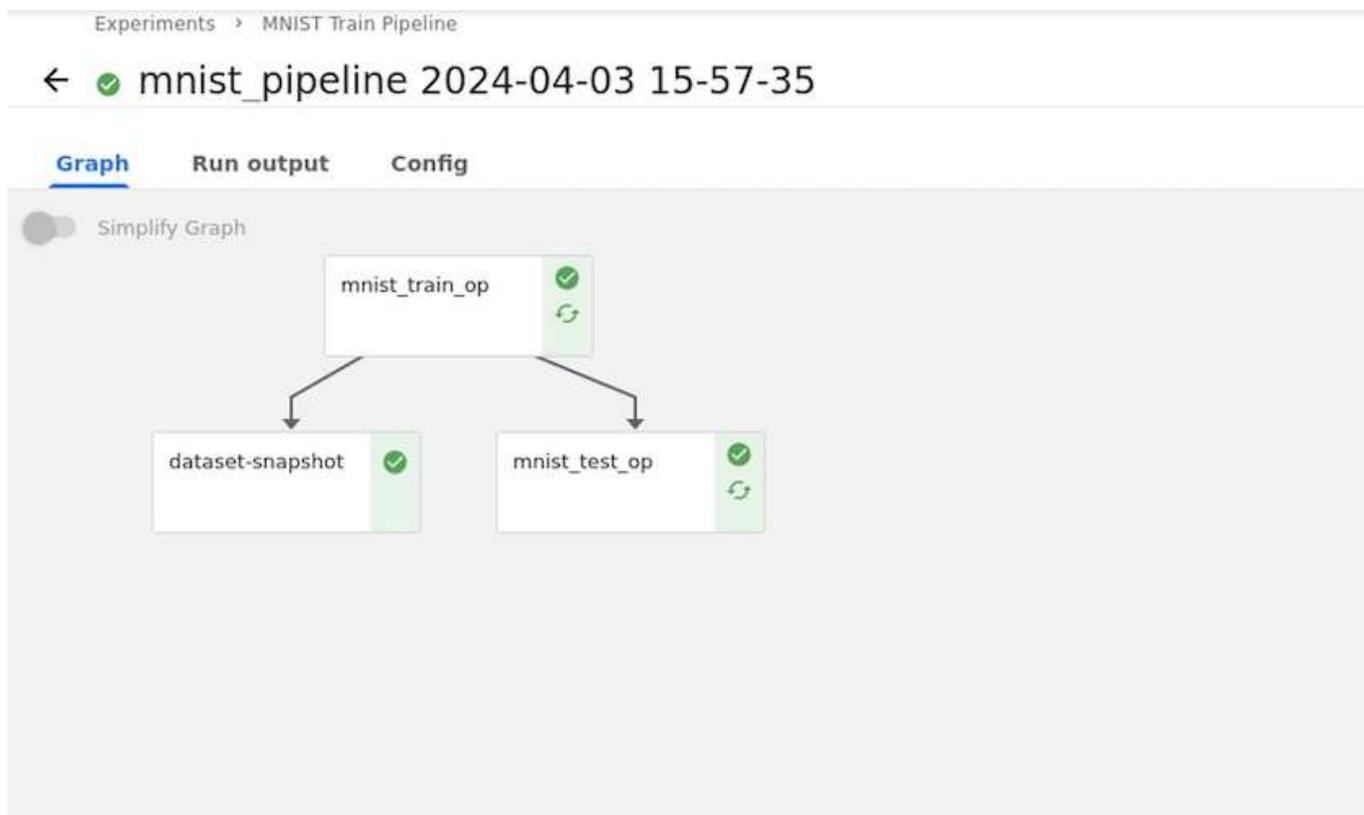
Ecco un esempio di un Dockerfile -

```
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"]
```

A seconda dei requisiti, installare tutte le librerie e i pacchetti necessari per eseguire il programma. Prima di addestrare il modello di apprendimento automatico, si presuppone che si disponga già di una distribuzione Kubeflow funzionante.

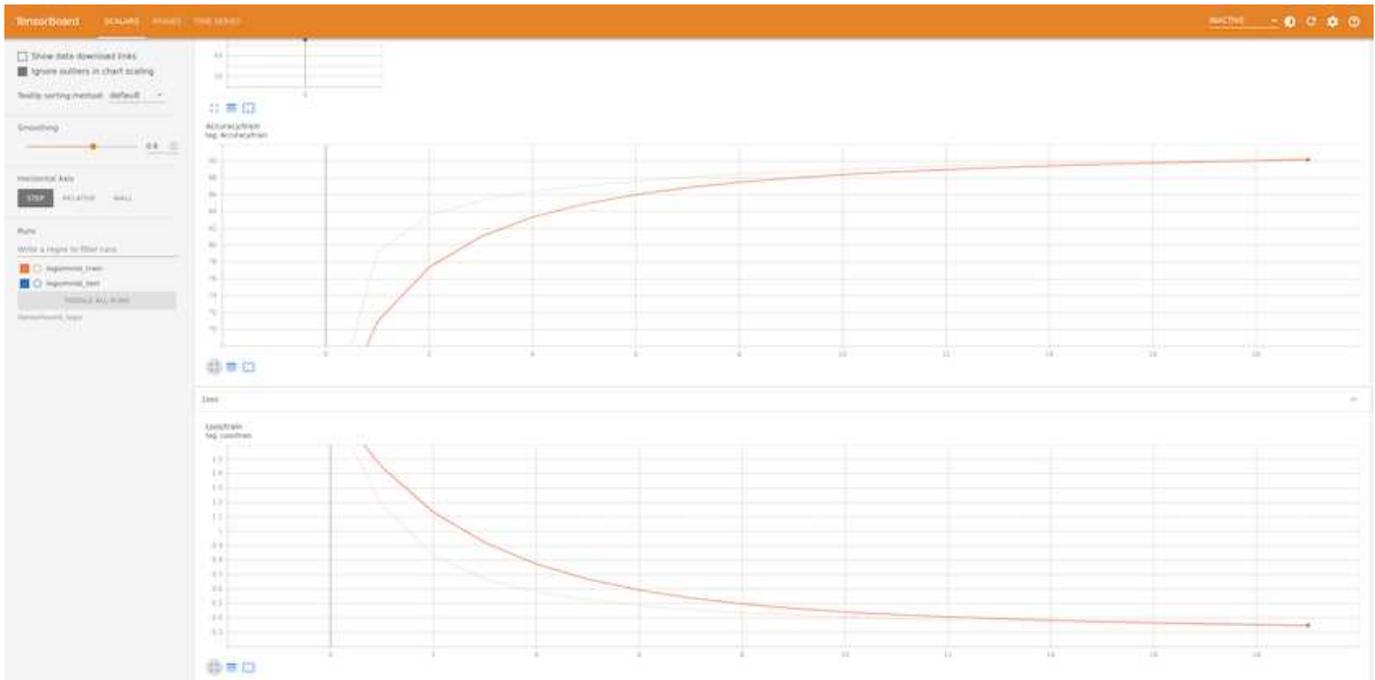
Addestrare un NN piccolo sui dati MNIST utilizzando le tubazioni PyTorch e Kubeflow

Utilizziamo l'esempio di una piccola rete neurale formata su dati MNIST. Il set di dati MNIST è composto da immagini scritte a mano di cifre da 0 a 9. Le immagini sono di dimensioni 28x28 pixel. Il set di dati è diviso in 60.000 immagini del treno e 10.000 immagini di convalida. La rete neurale utilizzata per questo esperimento è una rete feedforward a 2 strati. La formazione viene eseguita utilizzando le pipeline Kubeflow. Consultare la documentazione "[qui](#)" per ulteriori informazioni. La nostra pipeline Kubeflow incorpora l'immagine docker della sezione Prerequisiti.



Visualizzare i risultati utilizzando Tensorboard

Una volta addestrato il modello, possiamo visualizzare i risultati utilizzando Tensorboard. "[Tensorboard](#)" È disponibile come funzione nella dashboard Kubeflow. È possibile creare una scheda tensoriale personalizzata per il proprio lavoro. Un esempio riportato di seguito mostra il grafico della precisione della formazione rispetto al numero di epoche e perdita di formazione vs numero di epoche.



Sperimenta con Hyperparameters usando Katib

"Katib" È uno strumento all'interno di Kubeflow che può essere utilizzato per sperimentare gli iperparametri del modello. Per creare un esperimento, definire prima una metrica/obiettivo desiderato. Questa è solitamente la precisione del test. Una volta definita la metrica, scegliete gli iperparametri con cui volete giocare (optimizer/learning_rate/number of layers). Katib esegue una scansione iperparametrica con i valori definiti dall'utente per trovare la migliore combinazione di parametri che soddisfano la metrica desiderata. È possibile definire questi parametri in ciascuna sezione dell'interfaccia utente. In alternativa, è possibile definire un file **YAML** con le specifiche necessarie. Qui sotto è un'illustrazione di un esperimento di Katib -

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	
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, ltri

Algorithm	
Name	grid

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, and Katib. The main area displays 'Experiment details' for 'mnist-pytorch'. A message at the top states 'Couldn't find any successful Trial'. Below this is a table with columns: OVERVIEW, TRIALS, DETAILS, and YAML. The table lists various metrics such as Name, Status (Experiment is running), Best trial, Best trial's params, Best trial performance, User defined goal (Validation-accuracy > 0.9), Running trials (3), Failed trials (0), and Succeeded trials (0). At the bottom, there is a section for 'Experiment Conditions' and a filter input field.

Utilizzare le istantanee NetApp per salvare i dati per la tracciabilità

Durante il training sui modelli, potremmo voler salvare un'istantanea del set di dati di training per la tracciabilità. A tale scopo, possiamo aggiungere un passo snapshot alla pipeline, come illustrato di seguito. Per creare l'istantanea, è possibile utilizzare ["NetApp DataOps Toolkit per Kubernetes"](#).

```
@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=["\
            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={workflow.namespace}"],
        file_outputs={'volume_snapshot_name': '/volume_snapshot_name.txt'}
    )
    mnist_test_task.after(mnist_train_task)
    dataset_snapshot.after(mnist_train_task)
```

Fare riferimento a ["Esempio di toolkit DataOps NetApp per Kubeflow"](#) per ulteriori informazioni.

Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALIZZABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.