Beispiel Kubeflow Operations und Tasks

NetApp Solutions

NetApp
May 14, 2024

This PDF was generated from https://docs.netapp.com/de-de/netapp-solutions/ai/aicp_provision_a_jupyter_notebook_workspace_for_data_scientist_or_developer_use.html on May 14, 2024. Always check docs.netapp.com for the latest.

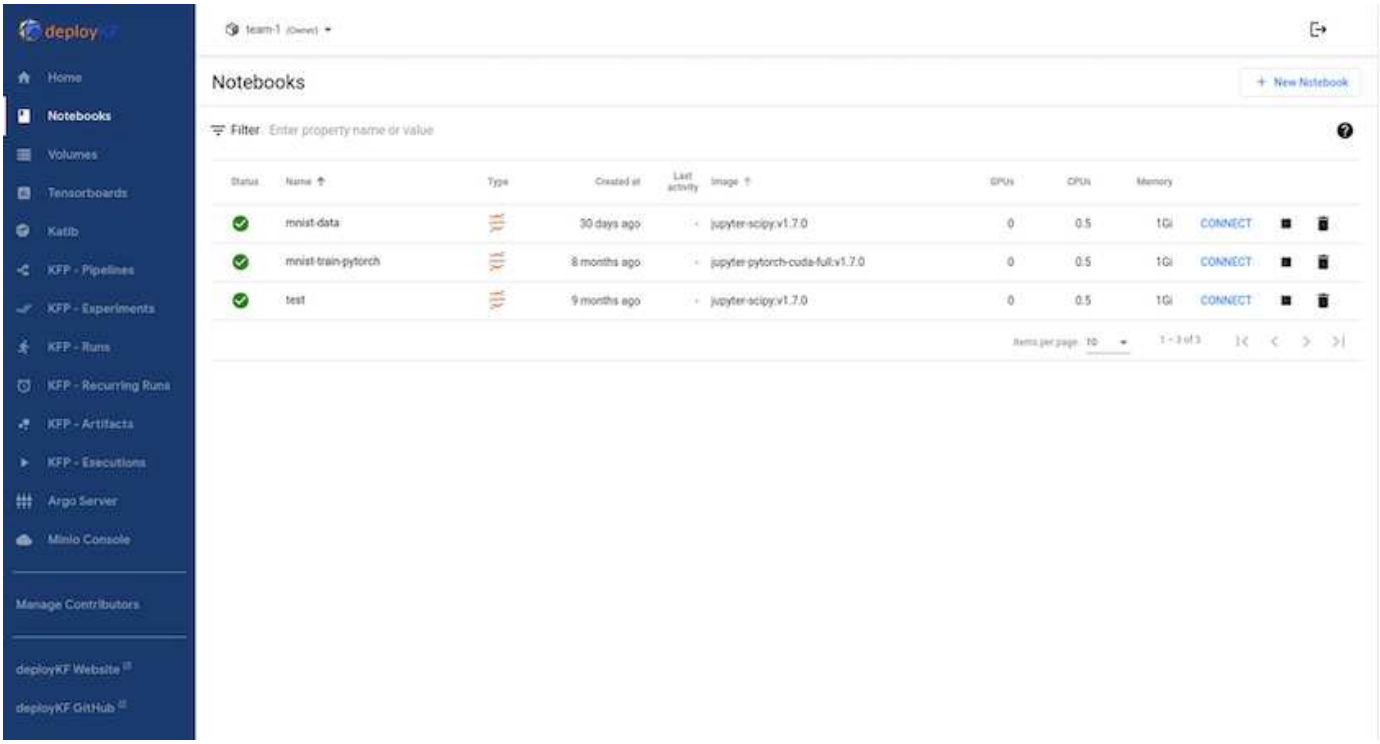
Inhalt

Beispiel Kubeflow Operations und Tasks . . . 1
Bereitstellen eines Jupyter Notebook Workspace für Data Scientist oder Entwickler . . . 1
Verwenden Sie das NetApp DataOps Toolkit mit Kubeflow . . . 1
Beispiel-Workflow – Trainieren eines Bilderkennungsmodells mit Kubeflow und dem NetApp DataOps Toolkit . . . 2

Beispiel Kubeflow Operations und Tasks

Bereitstellen eines Jupyter Notebook Workspace für Data Scientist oder Entwickler

Kubeflow ist in der Lage, neue Jupyter Notebook-Server schnell als Data Scientist-Workspaces bereitzustellen. Weitere Informationen zu Jupyter Notebooks im Kubeflow-Kontext finden Sie im "Offizielle Dokumentation von Kubeflow".



Verwenden Sie das NetApp DataOps Toolkit mit Kubeflow

Der "Das NetApp Data Science Toolkit für Kubernetes" Kann in Verbindung mit Kubeflow verwendet werden. Die Verwendung des NetApp Data Science Toolkit und Kubeflow bietet folgende Vorteile:

- Data Scientists können fortschrittliche NetApp-Datenmanagement-Vorgänge, wie das Erstellen von Snapshots und Klonen, direkt von einem Jupyter Notebook aus durchführen.
- Erweiterte NetApp Datenmanagement-Operationen wie das Erstellen von Snapshots und Klonen können mithilfe des Kubeflow Pipelines-Frameworks in automatisierte Workflows integriert werden.

Siehe "Kubeflow Beispiele" Im Abschnitt zum NetApp Data Science Toolkit GitHub Repository finden Sie weitere Informationen zur Verwendung des Toolkit mit Kubeflow.

Beispiel-Workflow – Trainieren eines Bilderkennungsmodells mit Kubeflow und dem NetApp DataOps Toolkit

In diesem Abschnitt werden die Schritte beschrieben, die bei der Schulung und Bereitstellung eines Neuronalen Netzwerks zur Bilderkennung mit Kubeflow und dem NetApp DataOps Toolkit erforderlich sind. Dies soll als Beispiel dienen, um eine Trainingsaufgabe zu zeigen, die NetApp Storage integriert.

Voraussetzungen

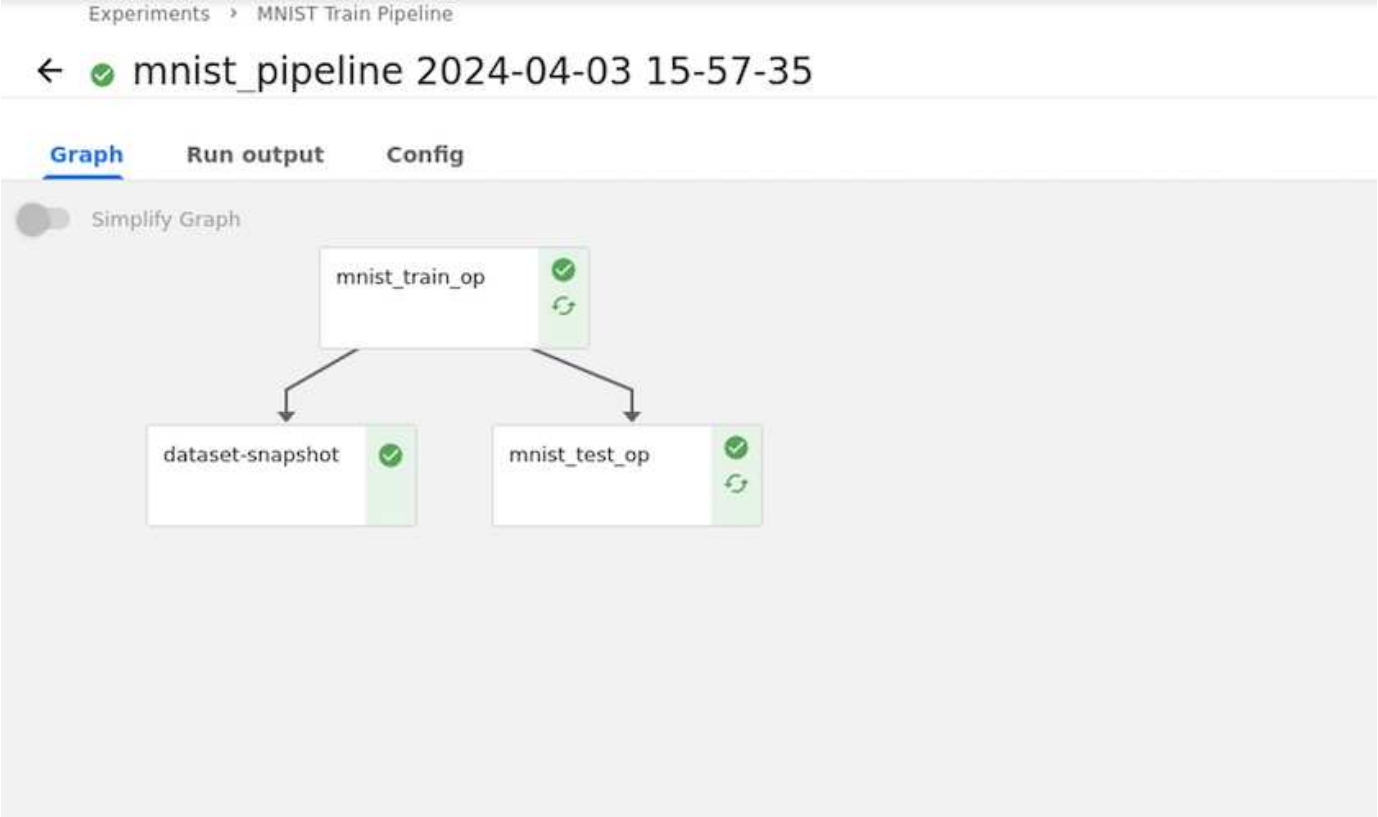
Erstellen Sie eine Dockdatei mit den erforderlichen Konfigurationen für die Zug- und Testschritte innerhalb der Kubeflow-Pipeline.
Hier ist ein Beispiel für eine Dockerdatei -

```
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"]
```

Installieren Sie je nach Ihren Anforderungen alle erforderlichen Bibliotheken und Pakete, die zum Ausführen des Programms erforderlich sind. Bevor Sie das Machine-Learning-Modell trainieren, wird davon ausgegangen, dass Sie bereits über eine funktionierende Kubeflow-Implementierung verfügen.

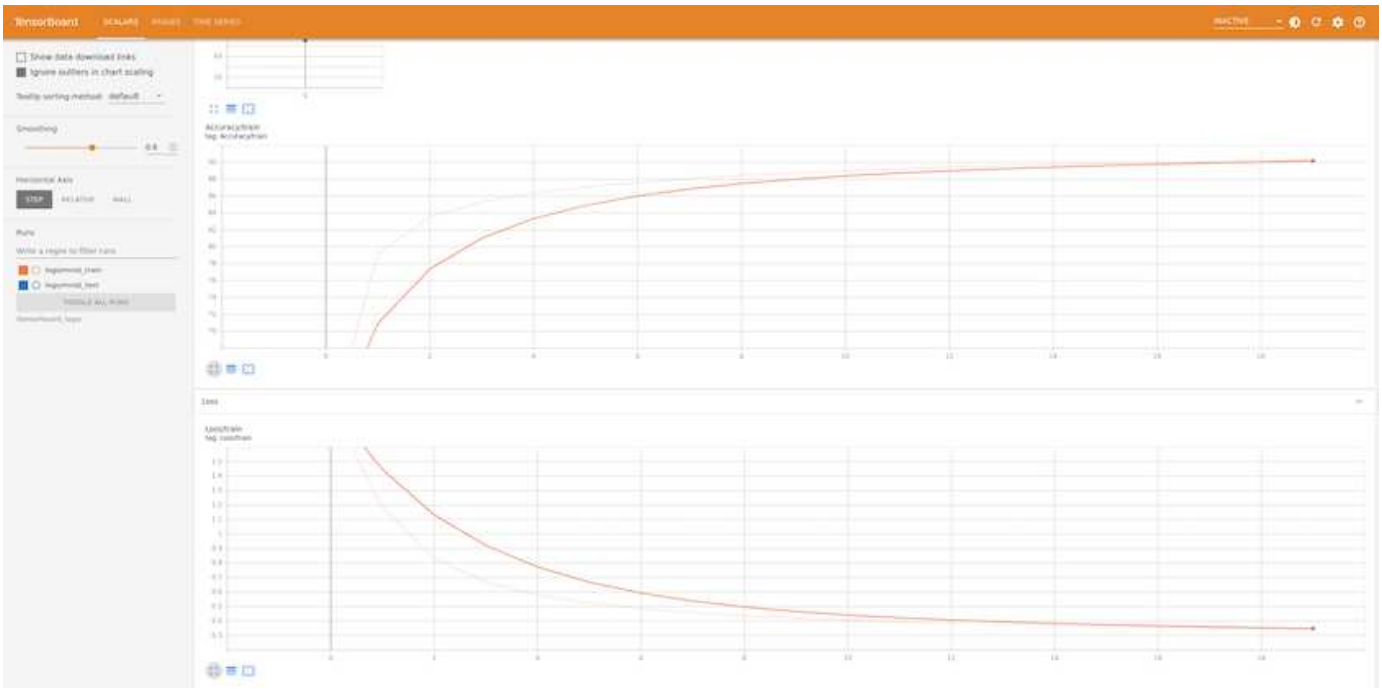
Trainieren Sie einen kleinen NN auf MNIST-Daten mit PyTorch- und Kubeflow-Pipelines

Wir verwenden das Beispiel eines kleinen neuronalen Netzwerks, das auf MNIST-Daten trainiert ist. Der MNIST-Datensatz besteht aus handschriftlichen Bildern von Ziffern von 0 bis 9. Die Bilder sind 28x28 Pixel groß. Der Datensatz ist in 60,000 Zug-Bilder und 10,000 Validierungsbilder unterteilt. Das für dieses Experiment verwendete Neuronale Netzwerk ist ein 2-schichtiges Feedforward-Netzwerk. Das Training wird mit Kubeflow Pipelines durchgeführt. Siehe Dokumentation "Hier" Finden Sie weitere Informationen. Unsere Kubeflow-Pipeline enthält das Docker-Image aus dem Abschnitt Voraussetzungen.



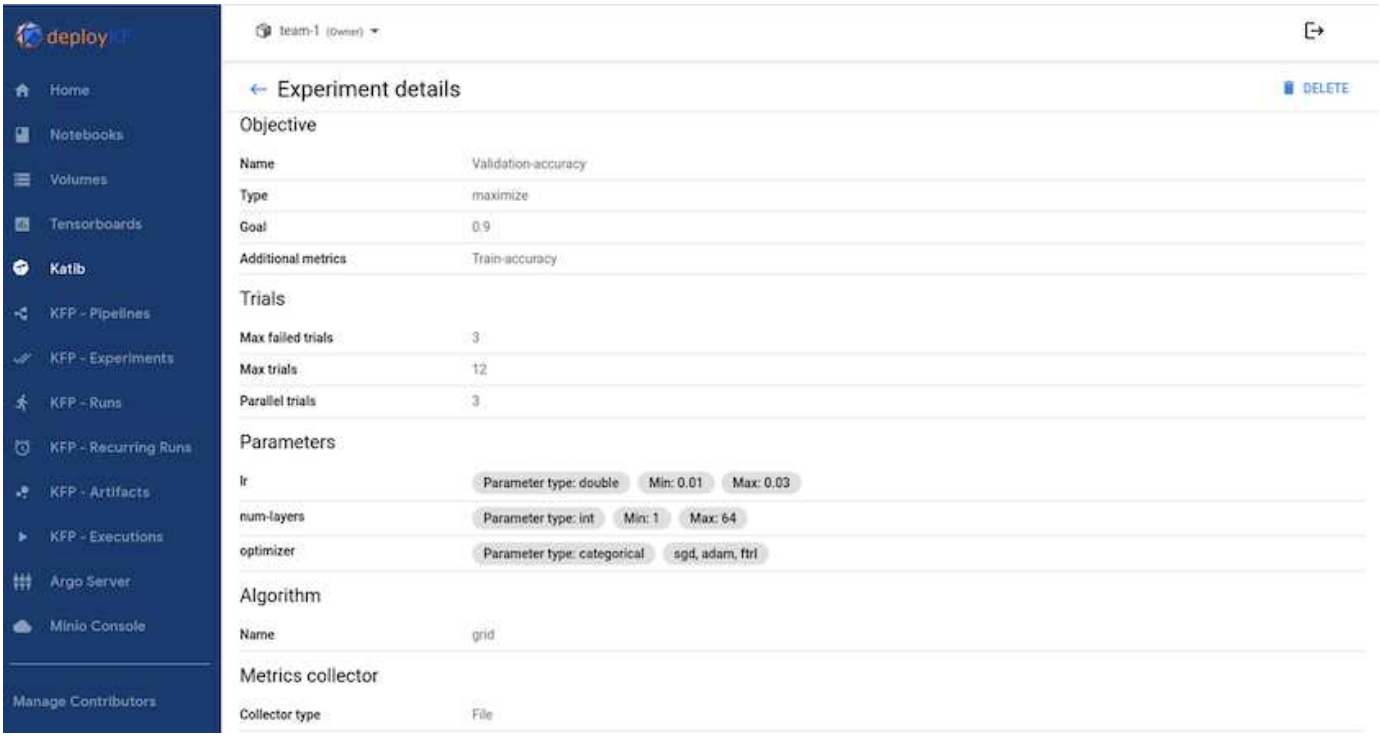
Visualisieren Sie Die Ergebnisse Mit Tensorboard

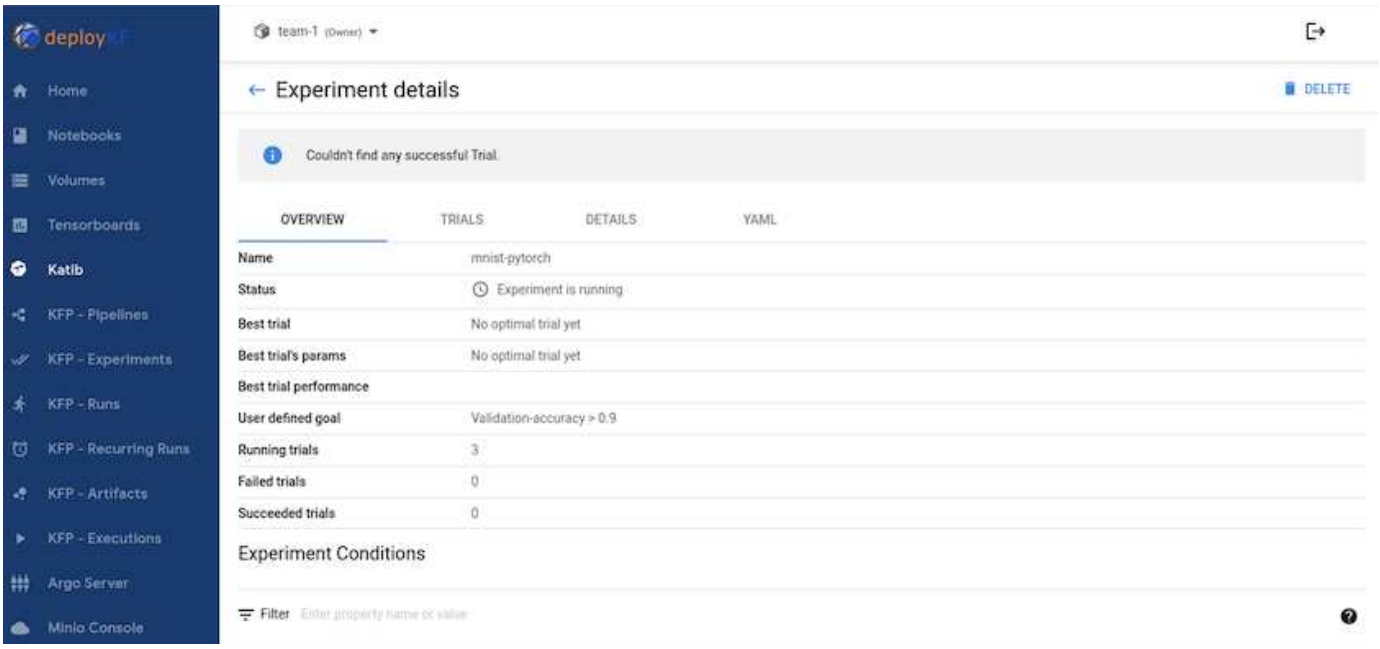
Sobald das Modell trainiert ist, können wir die Ergebnisse mit Tensorboard visualisieren. "Tensorboard" Ist als Funktion im Kubeflow Dashboard verfügbar. Sie können ein individuelles Tensorboard für Ihren Job erstellen. Ein Beispiel unten zeigt die Darstellung der Trainingsgenauigkeit vs. Anzahl der Epochen und Trainingsverluste vs Anzahl der Epochen.



Experimentieren Sie mit Hyperparametern mit Katib

"Katib" Ist ein Werkzeug innerhalb von Kubeflow, das verwendet werden kann, um mit den Modell-Hyperparametern zu experimentieren. Um ein Experiment zu erstellen, definieren Sie zunächst eine gewünschte Metrik/ein Ziel. Dies ist in der Regel die Prüfgenauigkeit. Nachdem die Metrik definiert wurde, wählen Sie Hyperparameter aus, mit denen Sie spielen möchten (Optimizer/Learning_Rate/Anzahl der Ebenen). Katib führt einen Hyperparameter-Sweep mit den benutzerdefinierten Werten durch, um die beste Kombination von Parametern zu finden, die die gewünschte Metrik erfüllen. Sie können diese Parameter in jedem Abschnitt der Benutzeroberfläche definieren. Alternativ können Sie eine **YAML**-Datei mit den erforderlichen Spezifikationen definieren. Unten ist eine Illustration eines Katib-Experiments -





Verwenden Sie NetApp-Snapshots, um Daten für die Rückverfolgbarkeit zu speichern

Während des Modelltrainings möchten wir eventuell eine Momentaufnahme des Trainingsdatensatzes zur Rückverfolgbarkeit speichern. Zu diesem Zweck können wir der Pipeline wie unten gezeigt einen Snapshot-Schritt hinzufügen. Zum Erstellen des Snapshots können wir den verwenden "NetApp DataOps Toolkit für 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)
```

Siehe "Beispiel für das NetApp DataOps Toolkit für Kubeflow" Finden Sie weitere Informationen.

Copyright-Informationen

Copyright © 2024 NetApp. Alle Rechte vorbehalten. Gedruckt in den USA. Dieses urheberrechtlich geschützte Dokument darf ohne die vorherige schriftliche Genehmigung des Urheberrechtsinhabers in keiner Form und durch keine Mittel – weder grafische noch elektronische oder mechanische, einschließlich Fotokopieren, Aufnehmen oder Speichern in einem elektronischen Abrufsystem – auch nicht in Teilen, vervielfältigt werden.

Software, die von urheberrechtlich geschütztem NetApp Material abgeleitet wird, unterliegt der folgenden Lizenz und dem folgenden Haftungsausschluss:

DIE VORLIEGENDE SOFTWARE WIRD IN DER VORLIEGENDEN FORM VON NETAPP ZUR VERFÜGUNG GESTELLT, D. H. OHNE JEGLICHE EXPLIZITE ODER IMPLIZITE GEWÄHRLEISTUNG, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDE GEWÄHRLEISTUNG DER MARKTGÄNGIGKEIT UND EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, DIE HIERMIT AUSGESCHLOSSEN WERDEN. NETAPP ÜBERNIMMT KEINERLEI HAFTUNG FÜR DIREKTE, INDIREKTE, ZUFÄLLIGE, BESONDERE, BEISPIELHAFTE SCHÄDEN ODER FOLGESCHÄDEN (EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE BESCHAFFUNG VON ERSATZWAREN ODER -DIENSTLEISTUNGEN, NUTZUNGS-, DATEN- ODER GEWINNVERLUSTE ODER UNTERBRECHUNG DES GESCHÄFTSBETRIEBS), UNABHÄNGIG DAVON, WIE SIE VERURSACHT WURDEN UND AUF WELCHER HAFTUNGSTHEORIE SIE BERUHEN, OB AUS VERTRAGLICH FESTGELEGTER HAFTUNG, VERSCHULDENSUNABHÄNGIGER HAFTUNG ODER DELIKTSHAFTUNG (EINSCHLIESSLICH FAHRLÄSSIGKEIT ODER AUF ANDEREM WEGE), DIE IN IRGENDEINER WEISE AUS DER NUTZUNG DIESER SOFTWARE RESULTIEREN, SELBST WENN AUF DIE MÖGLICHKEIT DERARTIGER SCHÄDEN HINGEWIESEN WURDE.

NetApp behält sich das Recht vor, die hierin beschriebenen Produkte jederzeit und ohne Vorankündigung zu ändern. NetApp übernimmt keine Verantwortung oder Haftung, die sich aus der Verwendung der hier beschriebenen Produkte ergibt, es sei denn, NetApp hat dem ausdrücklich in schriftlicher Form zugestimmt. Die Verwendung oder der Erwerb dieses Produkts stellt keine Lizenzierung im Rahmen eines Patentrechts, Markenrechts oder eines anderen Rechts an geistigem Eigentum von NetApp dar.

Das in diesem Dokument beschriebene Produkt kann durch ein oder mehrere US-amerikanische Patente, ausländische Patente oder anhängige Patentanmeldungen geschützt sein.

ERLÄUTERUNG ZU „RESTRICTED RIGHTS“: Nutzung, Vervielfältigung oder Offenlegung durch die US-Regierung unterliegt den Einschränkungen gemäß Unterabschnitt (b)(3) der Klausel „Rights in Technical Data – Noncommercial Items“ in DFARS 252.227-7013 (Februar 2014) und FAR 52.227-19 (Dezember 2007).

Die hierin enthaltenen Daten beziehen sich auf ein kommerzielles Produkt und/oder einen kommerziellen Service (wie in FAR 2.101 definiert) und sind Eigentum von NetApp, Inc. Alle technischen Daten und die Computersoftware von NetApp, die unter diesem Vertrag bereitgestellt werden, sind gewerblicher Natur und wurden ausschließlich unter Verwendung privater Mittel entwickelt. Die US-Regierung besitzt eine nicht ausschließliche, nicht übertragbare, nicht unterlizenzierbare, weltweite, limitierte unwiderrufliche Lizenz zur Nutzung der Daten nur in Verbindung mit und zur Unterstützung des Vertrags der US-Regierung, unter dem die Daten bereitgestellt wurden. Sofern in den vorliegenden Bedingungen nicht anders angegeben, dürfen die Daten ohne vorherige schriftliche Genehmigung von NetApp, Inc. nicht verwendet, offengelegt, vervielfältigt, geändert, aufgeführt oder angezeigt werden. Die Lizenzrechte der US-Regierung für das US-Verteidigungsministerium sind auf die in DFARS-Klausel 252.227-7015(b) (Februar 2014) genannten Rechte beschränkt.

Markeninformationen

NETAPP, das NETAPP Logo und die unter http://www.netapp.com/TM aufgeführten Marken sind Marken von NetApp, Inc. Andere Firmen und Produktnamen können Marken der jeweiligen Eigentümer sein.