



# **Provision a Jupyter Notebook Workspace for Data Scientist or Developer Use**

## **NetApp Solutions**

Dorian Henderson, Kevin Hoke, Michael Oglesby  
May 14, 2021

This PDF was generated from [https://docs.netapp.com/us-en/netapp-solutions/ai/aicp\\_provision\\_a\\_jupyter\\_notebook\\_workspace\\_for\\_data\\_scientist\\_or\\_developer\\_use.html](https://docs.netapp.com/us-en/netapp-solutions/ai/aicp_provision_a_jupyter_notebook_workspace_for_data_scientist_or_developer_use.html) on August 18, 2021. Always check docs.netapp.com for the latest.

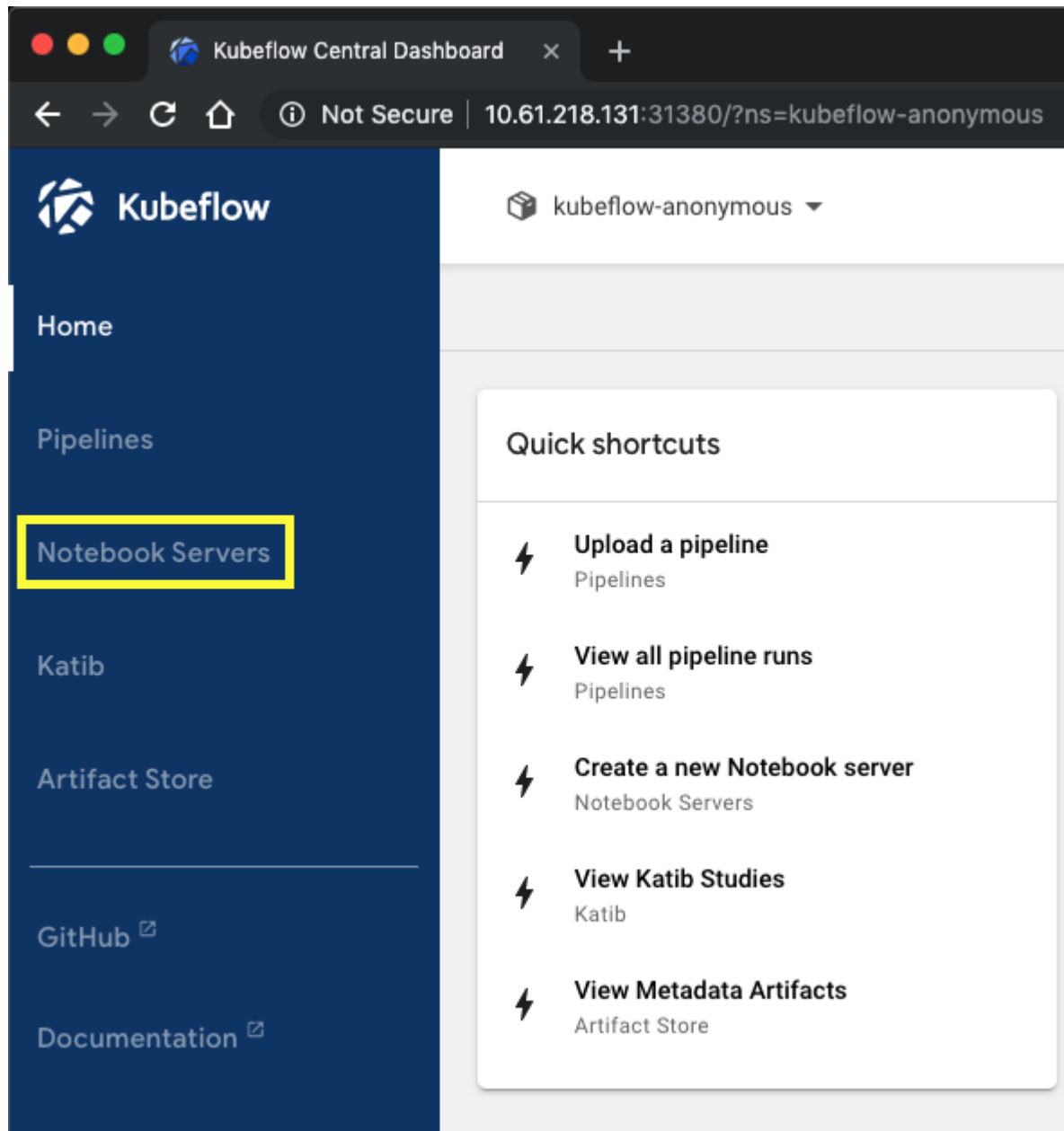
# Table of Contents

Provision a Jupyter Notebook Workspace for Data Scientist or Developer Use ..... 1

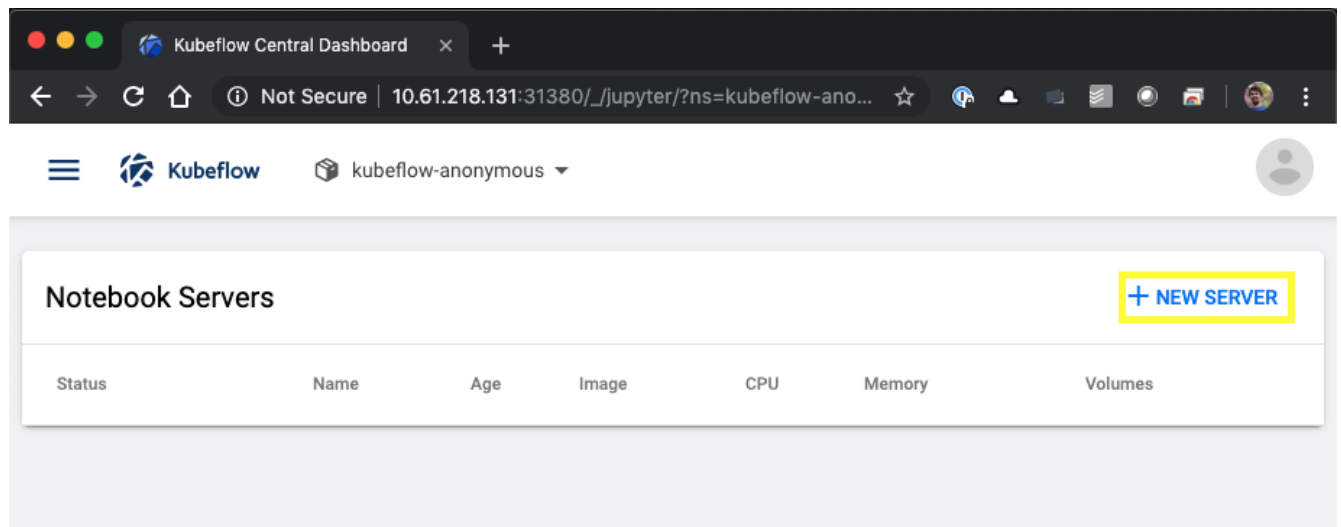
# Provision a Jupyter Notebook Workspace for Data Scientist or Developer Use

Kubeflow is capable of rapidly provisioning new Jupyter Notebook servers to act as data scientist workspaces. To provision a new Jupyter Notebook server with Kubeflow, perform the following tasks. For more information about Jupyter Notebooks within the Kubeflow context, see the [official Kubeflow documentation](#).

1. From the Kubeflow central dashboard, click Notebook Servers in the main menu to navigate to the Jupyter Notebook server administration page.



2. Click New Server to provision a new Jupyter Notebook server.



3. Give your new server a name, choose the Docker image that you want your server to be based on, and specify the amount of CPU and RAM to be reserved by your server. If the Namespace field is blank, use the Select Namespace menu in the page header to choose a namespace. The Namespace field is then auto-populated with the chosen namespace.

In the following example, the `kubeflow-anonymous` namespace is chosen. In addition, the default values for Docker image, CPU, and RAM are accepted.

**Name**

Specify the name of the Notebook Server and the Namespace it will belong to.

Name:  Namespace:

**Image**

A starter Jupyter Docker Image with a baseline deployment and typical ML packages.

☐ Custom Image

Image:

**CPU / RAM**

Specify the total amount of CPU and RAM reserved by your Notebook Server. For CPU-intensive workloads, you can choose more than 1 CPU (e.g. 1.5).

CPU:  Memory:

- Specify the workspace volume details. If you choose to create a new volume, then that volume or PVC is provisioned using the default StorageClass. Because a StorageClass utilizing Trident was designated as the default StorageClass in the section [Kubeflow Deployment](#), the volume or PVC is provisioned with Trident. This volume is automatically mounted as the default workspace within the Jupyter Notebook Server container. Any notebooks that a user creates on the server that are not saved to a separate data volume are automatically saved to this workspace volume. Therefore, the notebooks are persistent across reboots.

**Workspace Volume**

Configure the Volume to be mounted as your personal Workspace.

☐ Don't use Persistent Storage for User's home

Type:  Name:  Size:  Mode:  Mount Point:

- Add data volumes. The following example specifies an existing PVC named 'pb-fg-all' and accepts the default mount point.

## Data Volumes

Configure the Volumes to be mounted as your Datasets.

[+ ADD VOLUME](#)

Type	Name	Size	Mode	Mount Point
Existing	pb-fg-all	10Gi	ReadWriteOnce	/home/jovyan/data-vol-1

6. **Optional:** Request that the desired number of GPUs be allocated to your notebook server. In the following example, one GPU is requested.

## Configurations

Extra layers of configurations that will be applied to the new Notebook. (e.g. Insert credentials as Secrets, set Environment Variables.)

Configurations

## Extra Resources

Specify extra resources that might be needed in the Notebook Server.

☒ Enable Shared Memory

Extra Resources \*

`{"nvidia.com/gpu": 1}`

Extra Resources available in the cluster (ex. NVIDIA GPUs)

[LAUNCH](#)

[CANCEL](#)

7. Click Launch to provision your new notebook server.
8. Wait for your notebook server to be fully provisioned. This can take several minutes if you have never provisioned a server using the Docker image that you specified because the image needs to be downloaded. When your server has been fully provisioned, you see a green check mark in the Status column on the Jupyter Notebook server administration page.



Kubeflow Central Dashboard

Not Secure | 10.61.218.131:31380/\_jupyter/?ns=kubeflow-anonym...

Kubeflow kubeflow-anonymous

### Notebook Servers

[+ NEW SERVER](#)

Status	Name	Age	Image	CPU	Memory	Volumes	
	mike	12 mins ago	tensorflow-1.13.1-notebook-cpu:v0.5.0	0.5	1.0Gi		<a href="#">CONNECT</a> 

- Click Connect to connect to your new server web interface.
- Confirm that the dataset volume that was specified in step 6 is mounted on the server. Note that this volume is mounted within the default workspace by default. From the perspective of the user, this is just another folder within the workspace. The user, who is likely a data scientist and not an infrastructure expert, does not need to possess any storage expertise in order to use this volume.


Kubeflow Central Dashboard Home


Not Secure | 10.61.218.131:31380/notebook/kubeflow-anonymous/...

jupyter Quit

Files Running Clusters

Select items to perform actions on them.

Upload New 

<input type="checkbox"/>	0		Name	Last Modified	File size
<input type="checkbox"/>			data-vol-1	a day ago	

The screenshot shows the Jupyter web interface with the 'Files' tab selected. The breadcrumb path is '/ data-vol-1'. A table lists the contents of the volume:

	Name	Last Modified	File size
<input type="checkbox"/>	..	seconds ago	
<input type="checkbox"/>	blas_folder	2 months ago	
<input type="checkbox"/>	collected_trace	2 months ago	
<input type="checkbox"/>	container	3 months ago	
<input type="checkbox"/>	dataset	5 hours ago	
<input type="checkbox"/>	fio_test	3 months ago	
<input type="checkbox"/>	parabricks	7 months ago	
<input type="checkbox"/>	banking.csv	a month ago	4.88 MB

11. Open a Terminal and, assuming that a new volume was requested in step 5, execute `df -h` to confirm that a new Trident-provisioned persistent volume is mounted as the default workspace.

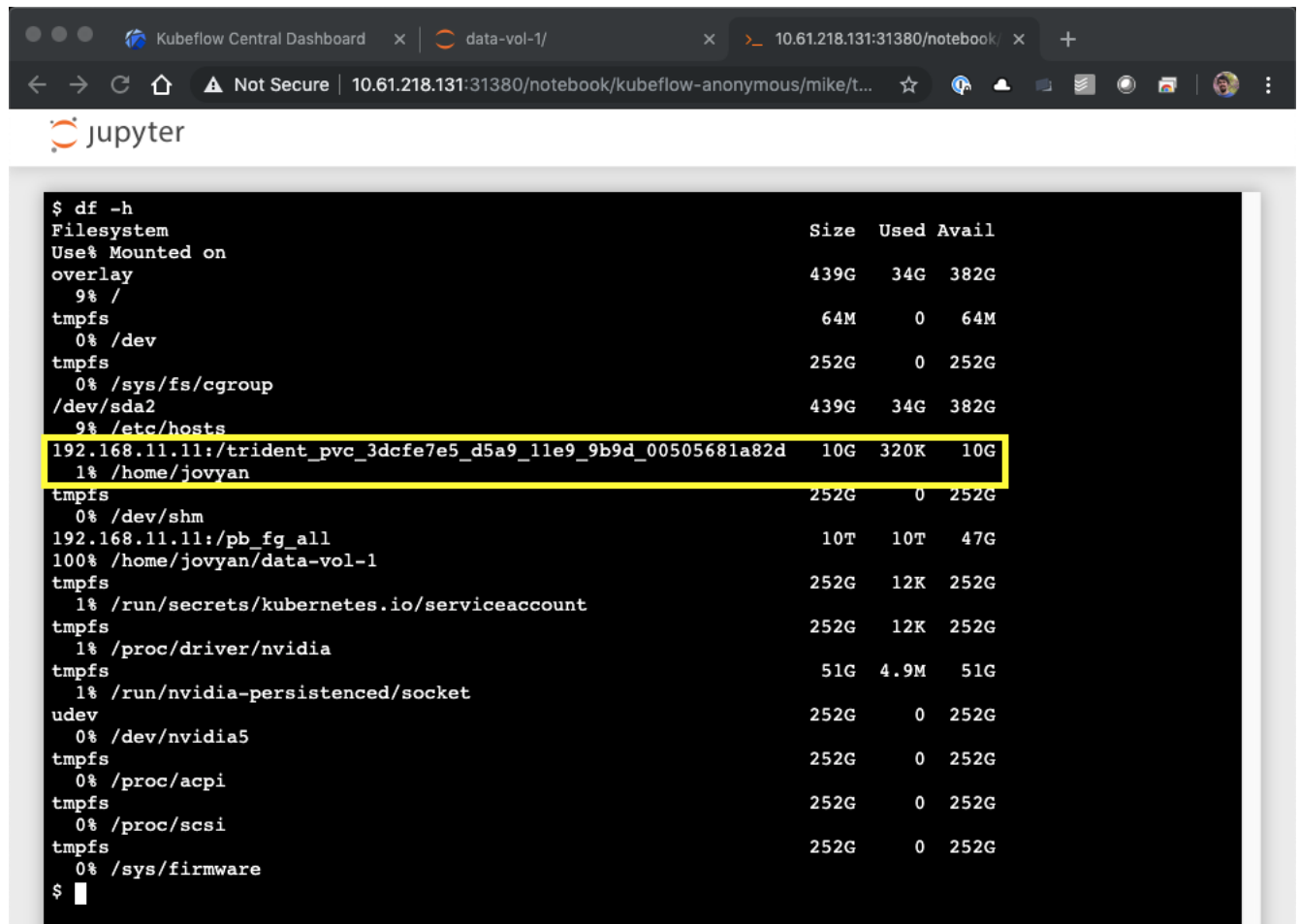
The default workspace directory is the base directory that you are presented with when you first access the server's web interface. Therefore, any artifacts that you create by using the web interface are stored on this Trident-provisioned persistent volume.

The screenshot shows the Jupyter web interface with the 'New' dropdown menu open. The menu options are:

- Notebook:
- Python 2
- Python 3
- Other:
- Text File
- Folder
- Terminal

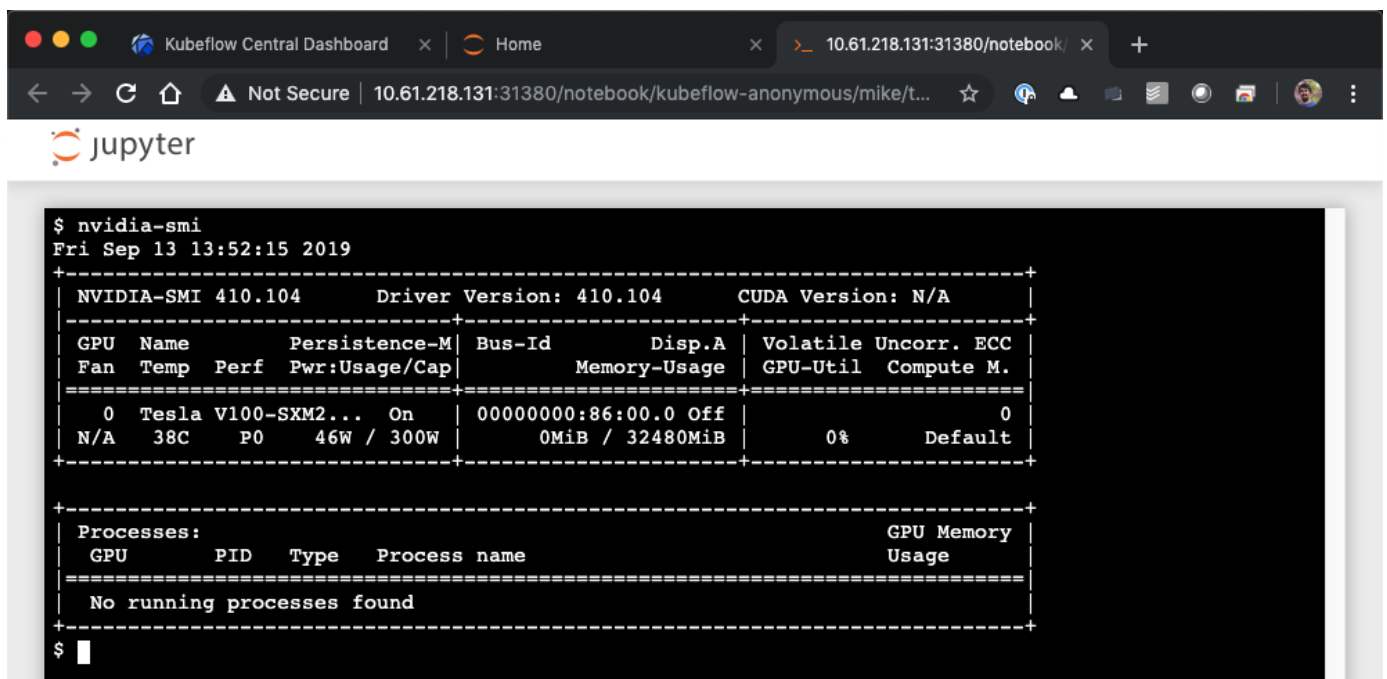
The 'Terminal' option is highlighted with a yellow box. The background shows the same file browser as the previous screenshot.





```
$ df -h
Filesystem                                Size  Used Avail
Use% Mounted on
overlay                                  439G   34G  382G
9% /
tmpfs                                     64M    0   64M
0% /dev
tmpfs                                     252G    0  252G
0% /sys/fs/cgroup
/dev/sda2                                439G   34G  382G
9% /etc/hosts
192.168.11.11:/trident_pvc_3dcfe7e5_d5a9_11e9_9b9d_00505681a82d 10G 320K 10G
1% /home/jovyan
tmpfs                                     252G    0  252G
0% /dev/shm
192.168.11.11:/pb_fg_all                  10T   10T   47G
100% /home/jovyan/data-vol-1
tmpfs                                     252G   12K  252G
1% /run/secrets/kubernetes.io/serviceaccount
tmpfs                                     252G   12K  252G
1% /proc/driver/nvidia
tmpfs                                     51G   4.9M   51G
1% /run/nvidia-persistenced/socket
udev                                     252G    0  252G
0% /dev/nvidia5
tmpfs                                     252G    0  252G
0% /proc/acpi
tmpfs                                     252G    0  252G
0% /proc/scsi
tmpfs                                     252G    0  252G
0% /sys/firmware
$
```

12. Using the terminal, run `nvidia-smi` to confirm that the correct number of GPUs were allocated to the notebook server. In the following example, one GPU has been allocated to the notebook server as requested in step 7.



```
$ nvidia-smi
Fri Sep 13 13:52:15 2019
+-----+
| NVIDIA-SMI 410.104      Driver Version: 410.104      CUDA Version: N/A      |
+-----+
| GPU   Name               Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|  0  Tesla V100-SXM2...    On         | 00000000:86:00:0 Off  |          0%         |
| N/A   38C    P0      46W / 300W | 0MiB / 32480MiB |           | Default |
+-----+-----+

+-----+
| Processes:                                                       GPU Memory |
|  GPU       PID    Type   Process name                     Usage      |
+-----+-----+
| No running processes found                                         |
+-----+
$
```

Next: [Example Notebooks and Pipelines](#)

## Copyright Information

Copyright © 2021 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

## Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.