



Manage Data Classification

NetApp Data Classification

NetApp

January 14, 2026

This PDF was generated from <https://docs.netapp.com/us-en/data-services-data-classification/task-exclude-scan-paths.html> on January 14, 2026. Always check docs.netapp.com for the latest.

Table of Contents

Manage Data Classification	1
Exclude specific directories from NetApp Data Classification scans	1
Supported data sources	1
Define the directories to exclude from scanning	1
Examples	2
Escaping special characters in folder names	3
View the current exclusion list	4
Define additional group IDs as open to organization in NetApp Data Classification	4
Add the "open to organization" permission to group IDs	4
View the current list of group IDs	5
Customize the stale data definition in NetApp Data Classification	5
Remove data sources from NetApp Data Classification	6
Deactivate scans for a system	6
Remove a database from Data Classification	6
Remove a group of file shares from Data Classification	6
Uninstall NetApp Data Classification	7
Uninstall Data Classification from a cloud provider	7
Uninstall Data Classification from an on-premises deployment	8

Manage Data Classification

Exclude specific directories from NetApp Data Classification scans

If you want NetApp Data Classification to exclude specific directories from scans, you can add these directory names to a configuration file. After you apply this change, the Data Classification engine excludes those directories from scans.



By default, Data Classification scans excludes volume snapshot data, which is identical to its source in the volume.

Supported data sources

Excluding specific directories from Data Classification scans is supported for NFS and CIFS shares in the following data sources:

- On-premises ONTAP
- Cloud Volumes ONTAP
- Amazon FSx for NetApp ONTAP
- Azure NetApp Files
- General file shares

Define the directories to exclude from scanning

Before you can exclude directories from classification scanning, you need to log into the Data Classification system so you can edit a configuration file and run a script. See how to [log in to the Data Classification system](#) depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Considerations

- You can exclude a maximum of 50 directory paths per Data Classification system.
- Excluding directory paths can affect scanning times.

Steps

1. On the Data Classification system, go to "/opt/netapp/config/custom_configuration" then open the file `data_provider.yaml`.
2. In the "data_providers" section under the line "exclude:", enter the directory paths to exclude. For example:

```
exclude:
  - "folder1"
  - "folder2"
```

Do not modify anything else in this file.

3. Save the changes to the file.
4. Go to "/opt/netapp/Datasense/tools/customer_configuration/data_providers" and run the following script:

```
update_data_providers_from_config_file.sh
```

+

This command commits the directories to be excluded from scanning to the classification engine.

Result

All subsequent scans of your data will exclude scanning of those specified directories.

You can add, edit, or delete items from the exclude list using these same steps. The revised exclude list will be updated after you run the script to commit your changes.

Examples

Configuration 1:

Every folder that contains "folder1" anywhere in the name will be excluded from all data sources.

```
data_providers:  
  exclude:  
    - "folder1"
```

Expected results for paths that will be excluded:

- /CVO1/folder1
- /CVO1/folder1name
- /CVO1/folder10
- /CVO1/*folder1
- /CVO1/+folder1name
- /CVO1/notfolder10
- /CVO22/folder1
- /CVO22/folder1name
- /CVO22/folder10

Examples for paths that will not be excluded:

- /CVO1/*folder
- /CVO1/foldername
- /CVO22/*folder20

Configuration 2:

Every folder that contains "*folder1" only at the start of the name will be excluded.

```
data_providers:  
  exclude:  
    - "\\\*folder1"
```

Expected results for paths that will be excluded:

- /CVO/*folder1
- /CVO/*folder1name
- /CVO/*folder10

Examples for paths that will not be excluded:

- /CVO/folder1
- /CVO/folder1name
- /CVO/not*folder10

Configuration 3:

Every folder in data source "CVO22" that contains "folder1" anywhere in the name will be excluded.

```
data_providers:  
  exclude:  
    - "CVO22/folder1"
```

Expected results for paths that will be excluded:

- /CVO22/folder1
- /CVO22/folder1name
- /CVO22/folder10

Examples for paths that will not be excluded:

- /CVO1/folder1
- /CVO1/folder1name
- /CVO1/folder10

Escaping special characters in folder names

If you have a folder name that contains one of the following special characters and you want to exclude data in that folder from being scanned, you'll need to use the escape sequence \\ before the folder name.

```
., +, *, ?, ^, $, (, ), [ , ], {, }, |
```

For example:

Path in source: /project/*not_to_scan

Syntax in exclude file: "*not_to_scan"

View the current exclusion list

It's possible for the contents of the `data_provider.yaml` configuration file to be different than what has actually been committed after running the `update_data_providers_from_config_file.sh` script. To view the current list of directories that you've excluded from Data Classification scanning, run the following command from `"/opt/netapp/Datasense/tools/customer_configuration/data_providers"`:

```
get_data_providers_configuration.sh
```

Define additional group IDs as open to organization in NetApp Data Classification

When group IDs (GIDs) are attached to files or folders in NFS file shares they define the permissions for the file or folder; such as whether they are "open to the organization". If some GIDs are not initially set up with the "Open to Organization" permission level, you can add that permission to the GID so that any files and folders that have that GID attached will be deemed "open to the organization".

After you make this change and NetApp Data Classification rescans your files and folders, any files and folders that have these group IDs attached will show this permission in the Investigation Details page, and they'll also appear in reports where you are displaying file permissions.

To activate this functionality, you need to log into the Data Classification system so you can edit a configuration file and run a script. See how to [log in to the Data Classification system](#) depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Add the "open to organization" permission to group IDs

You need to have the group ID numbers (GIDs) before starting this task.

Steps

1. On the Data Classification system, go to `"/opt/netapp/config/custom_configuration"` and open the file `data_provider.yaml`.
2. In the line `"organization_group_ids: []"` add the group IDs. For example:

```
organization_group_ids: [1014, 1015, 21, 2021, 1013, 2020, 1018, 1019]
```

Do not change anything else in this file.

3. Save the changes to the file.
4. Go to `"/opt/netapp/Datasense/tools/customer_configuration/data_providers"` and run the following script:

```
update_data_providers_from_config_file.sh
```

This command commits the revised group ID permissions to the classification engine.

Result

All subsequent scans of your data will identify files or folders that have these group IDs attached as "open to organization."

You can edit the list of group IDs and delete any group IDs you added in the past using these same steps. The revised list of group IDs will be updated after you run the script to commit your changes.

View the current list of group IDs

It's possible for the contents of the `data_provider.yaml` configuration file to differ from what has actually been committed after running the `update_data_providers_from_config_file.sh` script. To view the current list of group IDs that you've added to Data Classification, run the following command from `"/opt/netapp/Datasense/tools/customer_configuration/data_providers"`:

```
get_data_providers_configuration.sh
```

Customize the stale data definition in NetApp Data Classification

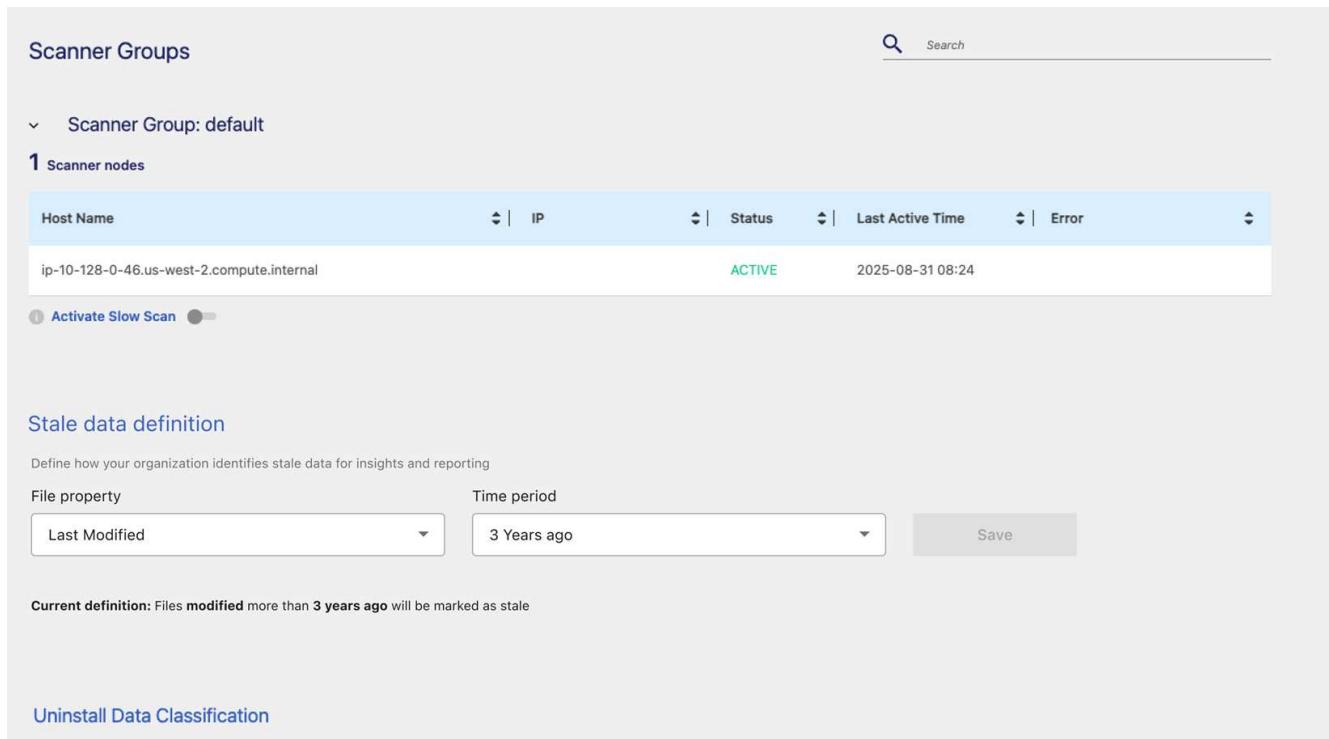
NetApp Data Classification identifies stale data to help you identify saving opportunities and governance risks. Because the definition of stale data can vary across different organizational contexts, you can customize how Data Classification defines stale data.

Stale data can be defined based on when it was *last accessed* or *last modified*. The time period selections range from 6 months ago to 10 years ago.

By default, data is considered stale if it was last modified three years ago.

Define stale data

1. In Ransomware Resilience, select **Configuration**.
2. In the Configuration page, scroll to the **Stale data definition** heading.
3. In the **File properties** dropdown menu, choose if you want to define stale data based on when it was **Last accessed** or **Last modified**.
4. Choose the time period for the stale data definition.



Scanner Groups

Scanner Group: default

1 scanner nodes

Host Name	IP	Status	Last Active Time	Error
ip-10-128-0-46.us-west-2.compute.internal		ACTIVE	2025-08-31 08:24	

Activate Slow Scan

Stale data definition

Define how your organization identifies stale data for insights and reporting

File property

Last Modified

Time period

3 Years ago

Save

Current definition: Files modified more than 3 years ago will be marked as stale

Uninstall Data Classification

5. Select **Save**.

Remove data sources from NetApp Data Classification

If you need to, you can stop NetApp Data Classification from scanning one or more systems, databases, or file share groups.

Deactivate scans for a system

When you deactivate scans, Data Classification no longer scans the data on the system and it removes the indexed insights from the Data Classification instance. The data from the system itself isn't deleted.

1. From the *Configuration* page, select the  button in the row for the system then **Deactivate Data Classification**.



You can also disable scans for a system from the Services panel when you select the system.

Remove a database from Data Classification

If you no longer need to scan a certain database, you can delete it from the Data Classification interface and stop all scans.

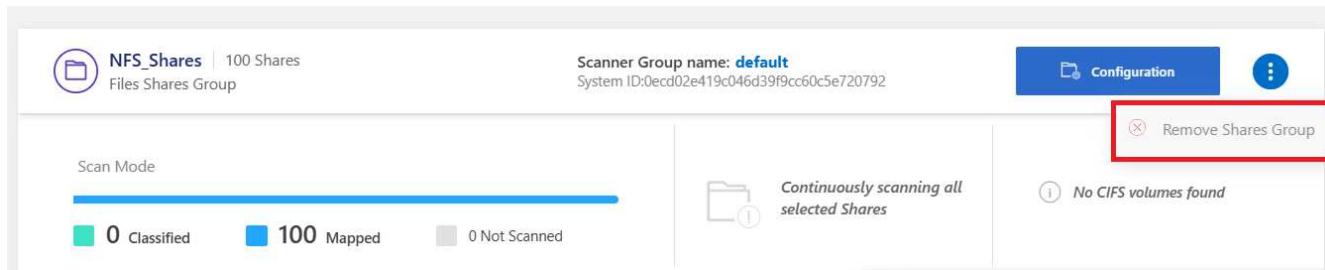
1. From the *Configuration* page, select the  button in the row for the database then **Remove DB Server**.

Remove a group of file shares from Data Classification

If you no longer want to scan user files from a file share group, you can delete the File Shares Group from the Data Classification interface and stop all scans.

Steps

1. From the **Configuration** page, select the  button in the row for the File Shares Group then **Remove File Shares Group**.



The screenshot shows the 'File Shares Group' section of the Configuration page. It displays 'NFS_Shares' with 100 Shares, 'Scanner Group name: default', and 'System ID: 0ecd02e419c046d39f9cc60c5e720792'. Below this, 'Scan Mode' is set to 'Continuously scanning all selected Shares'. A progress bar shows 100 Mapped shares. A message indicates 'No CIFS volumes found'. On the right, a red box highlights the 'Remove Shares Group' button.

2. Select **Delete Group of Shares** from the confirmation dialog.

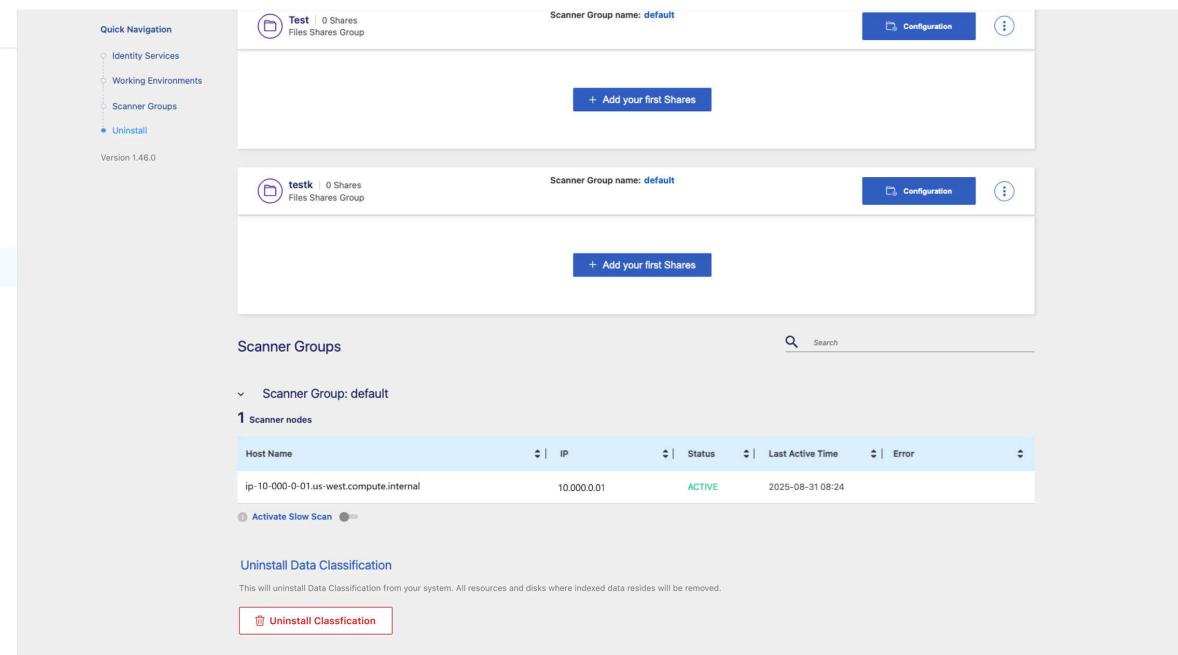
Uninstall NetApp Data Classification

You can uninstall NetApp Data Classification to troubleshoot issues or to permanently remove the software from the host. Deleting the instance also deletes the associated disks where the indexed data resides, meaning all the information Data Classification has scanned will be permanently deleted.

The steps that you need to use depend on whether you deployed Data Classification in the cloud or on an on-premises host.

Uninstall Data Classification from a cloud provider

1. From Data Classification, select **Configuration**.
2. At the bottom of the configuration page, select **Uninstall Classification**.



The screenshot shows the Configuration page with the 'Scanner Groups' section. It lists 'Scanner Group: default' with 1 Scanner node. The node details are: Host Name: ip-10-000-0-01.us-west.compute.internal, IP: 10.000.0.01, Status: ACTIVE, Last Active Time: 2025-08-31 08:24. Below this, an 'Uninstall Data Classification' dialog is open, stating: 'This will uninstall Data Classification from your system. All resources and disks where indexed data resides will be removed.' A red box highlights the 'Uninstall Classification' button.

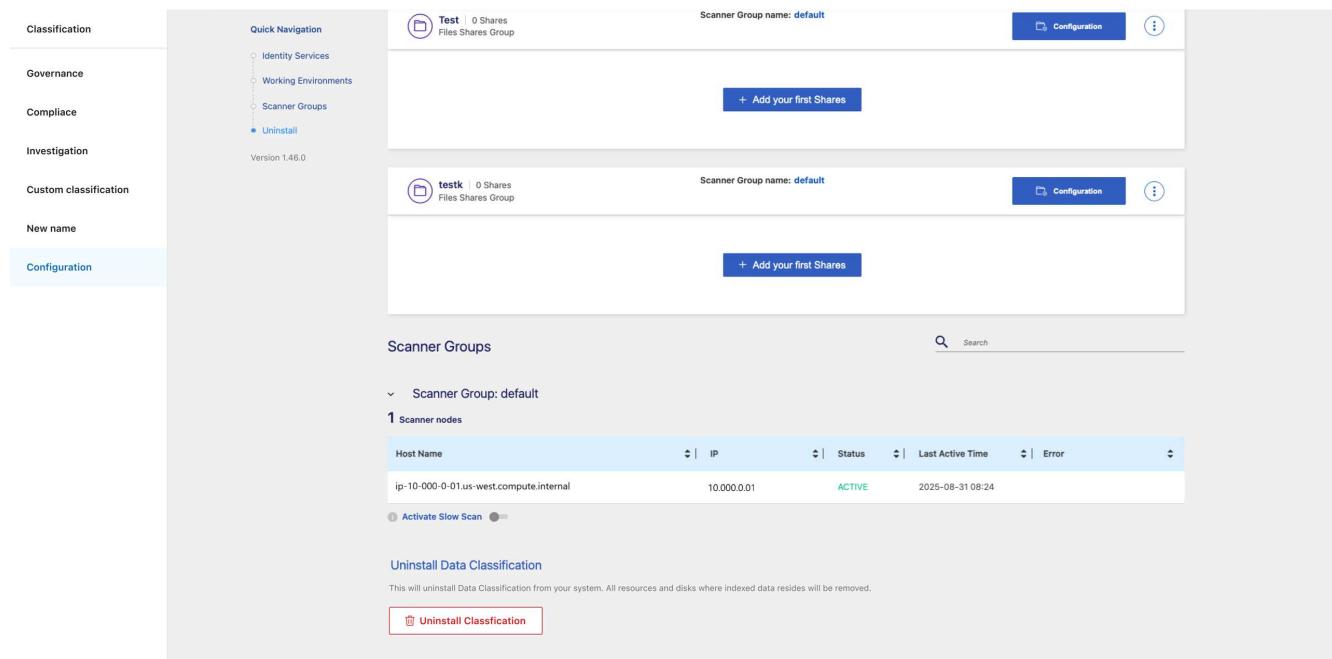
3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Classification instance from the

Console agent. Select **Uninstall** to confirm.

4. In the *Uninstall Classification* dialog, type **uninstall** to confirm that you want to disconnect the Data Classification instance from the Console agent then select **Uninstall**.
5. To finalize the uninstall process, go to your cloud provider's console and delete the Data Classification instance. The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

Uninstall Data Classification from an on-premises deployment

1. From Data Classification, select **Configuration**.
2. At the bottom of the configuration page, select **Uninstall Classification**.



The screenshot shows the Data Classification configuration interface. On the left, a sidebar lists categories: Classification, Governance, Compliance, Investigation, Custom classification, New name, and Configuration (which is selected). The main area is titled 'Scanner Groups' and shows two entries: 'Test' (0 Shares) and 'testk' (0 Shares). Each entry has a 'Configuration' button and a three-dot menu. Below the entries is a table titled 'Scanner Groups' with one row for 'Scanner Group: default'. The table has columns for Host Name, IP, Status, Last Active Time, and Error. A single node is listed: 'ip-10-000-0-01.us-west.compute.internal' with IP '10.000.0.01', Status 'ACTIVE', and Last Active Time '2025-08-31 08:24'. A button 'Activate Slow Scan' is shown. At the bottom, a 'Uninstall Data Classification' dialog box is open, containing the text: 'This will uninstall Data Classification from your system. All resources and disks where indexed data resides will be removed.' and a red 'Uninstall Classification' button.

3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Classification instance from the Console agent. Select **Uninstall** to confirm.
4. To uninstall the software from the host, run the `cleanup.sh` script on the Data Classification host machine, for example:

```
cleanup.sh
```

The script is located in the `/install/light_probe/onprem_installer/cleanup.sh` directory. See how to [log in to the Data Classification host machine](#).

Copyright information

Copyright © 2026 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.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

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.