



Deploy BlueXP classification deprecations

BlueXP classification

NetApp
June 20, 2024

Table of Contents

- Deploy BlueXP classification deprecations 1
 - Install BlueXP classification on multiple hosts for large configurations with no internet access 1
 - Add scanner nodes to an existing deployment 2

Deploy BlueXP classification deprecations

Install BlueXP classification on multiple hosts for large configurations with no internet access

Complete a few steps to install BlueXP classification on multiple hosts in an on-premises site that doesn't have internet access - also known as *private mode*. This type of installation is perfect for your secure sites.

For very large configurations where you'll be scanning petabytes of data in sites without internet access, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the *Manager node* and the additional systems that provide extra processing power are called *Scanner nodes*.

Follow these steps when installing BlueXP classification software on multiple on-premises hosts in an offline environment.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

What you'll need

- Verify that all your Linux systems for the Manager and Scanner nodes meet the host requirements.
- Verify that you have installed the two prerequisite software packages (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your offline environment meets the required permissions and connectivity.
- You must have the IP addresses of the scanner node hosts that you plan to use.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

Steps

1. Follow steps 1 through 8 from the [Single-host installation](#) on the manager node.
2. As shown in step 9, when prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer.

In addition to the variables available for a single-host installation, a new option **-n <node_ip>** is used to specify the IP addresses of the scanner nodes. Multiple node IPs are separated by a comma.

For example, this command adds 3 scanner nodes:

```
sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host  
<ds_host> --manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --no  
-proxy --darksite
```

3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example: `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212`) and save it in a text file.
4. On **each** scanner node host:
 - a. Copy the Data Sense installer file (**cc_onprem_installer.tar.gz**) to the host machine.
 - b. Unzip the installer file.
 - c. Paste and run the command that you copied in step 3.

When the installation finishes on all scanner nodes and they have been joined to the manager node, the manager node installation finishes as well.

Result

The BlueXP classification installer finishes installing packages, and registers the installation. Installation can take 15 to 25 minutes.

What's Next

From the Configuration page you can select the local [on-prem ONTAP clusters](#) and local [databases](#) that you want to scan.

Add scanner nodes to an existing deployment

You can add scanner nodes to an existing deployment on a Linux host with internet access.

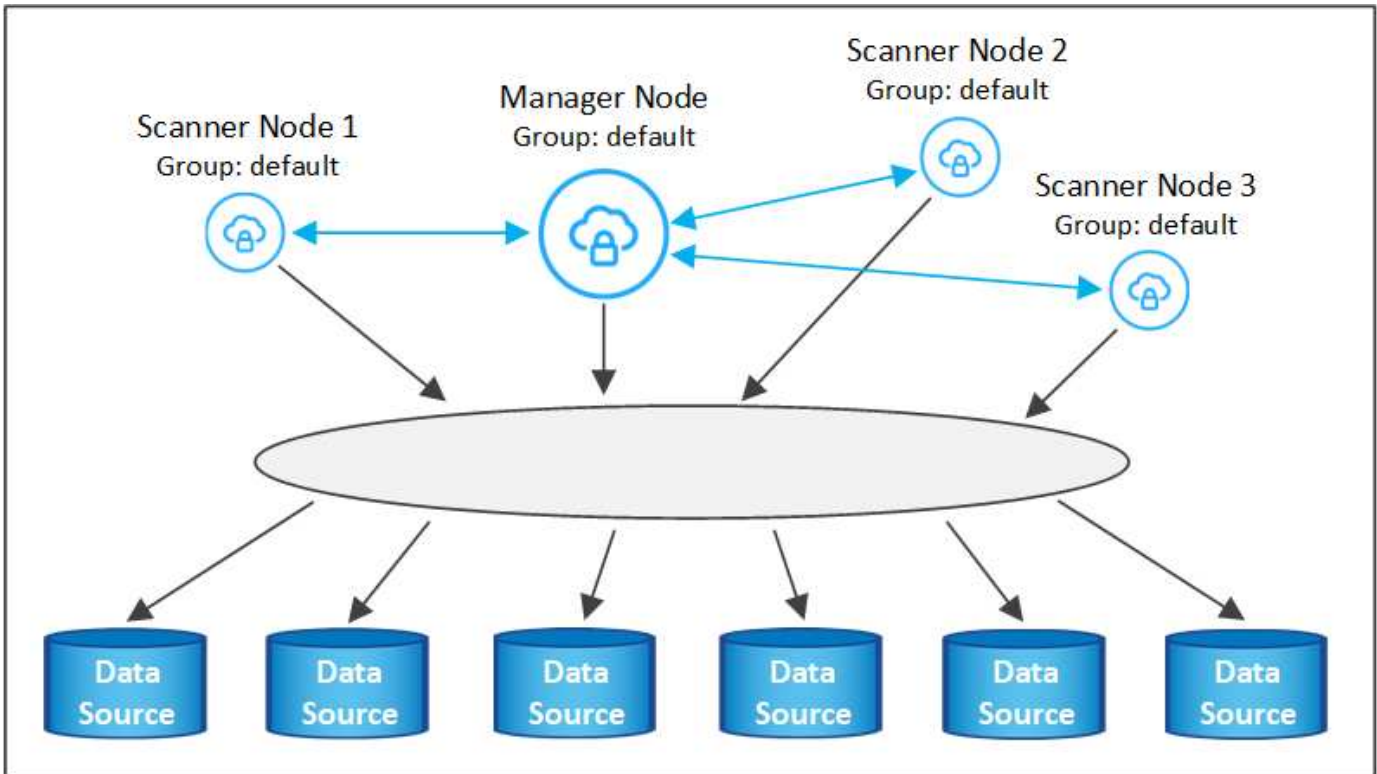
You can add more scanner nodes if you find that you need more scanning processing power to scan your data sources. You can add the scanner nodes immediately after installing the manager node, or you can add a scanner node later. For example, if you realize that the amount of data in one of your data sources has doubled or tripled in size after 6 months, you can add a new scanner node to assist with data scanning.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

There are two ways in which you can add additional scanner nodes:

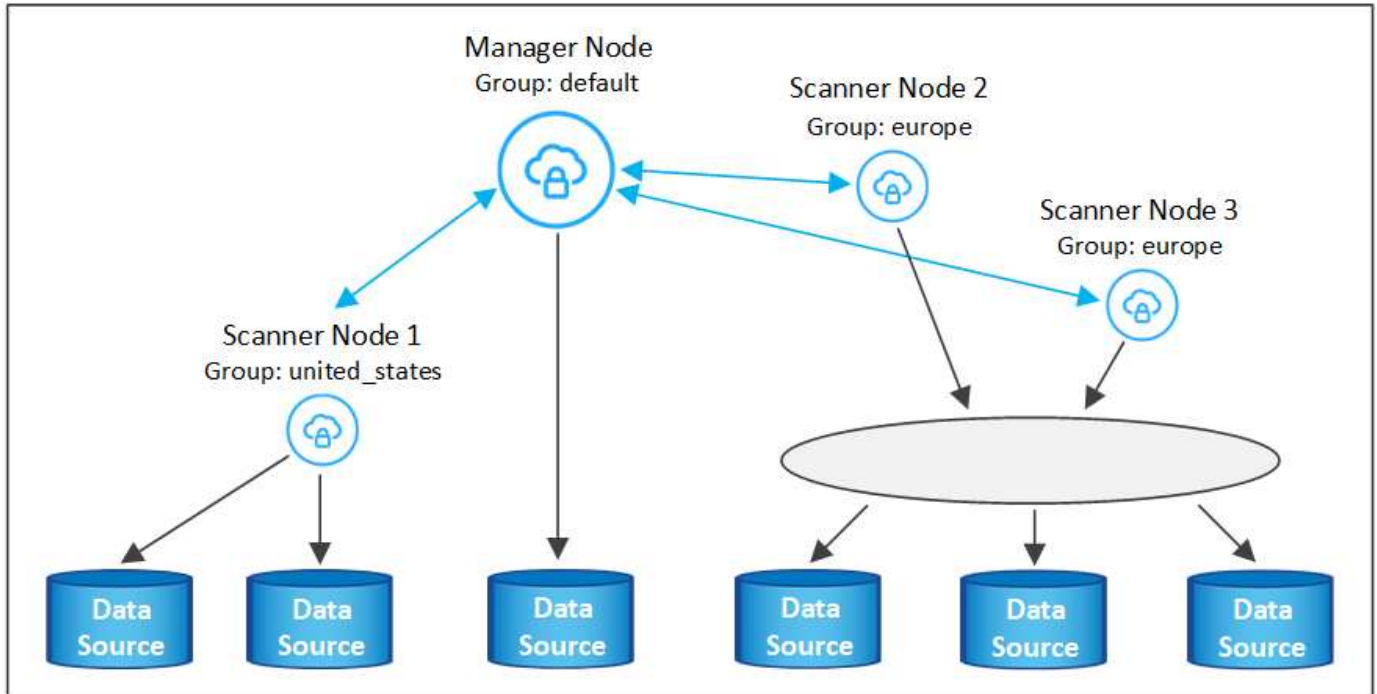
- add a node to assist with scanning all data sources
- add a node to assist with scanning a specific data source, or a specific group of data sources (typically based on location)

By default, any new scanner nodes you add are added to the general pool of scanning resources. This is called the "default scanner group". In the image below, there is 1 Manager node and 3 Scanner nodes in the "default" group that are all scanning data from all 6 data sources.



If you have certain data sources that you want to be scanned by scanner nodes that are physically closer to the data sources, you can define a scanner node, or group of scanner nodes, to scan a specific data source, or group of data sources. In the image below, there is 1 Manager node and 3 Scanner nodes.

- The Manager node is in the "default" group, and it is scanning 1 data source
- Scanner node 1 is in the "united_states" group, and it is scanning 2 data sources
- Scanner nodes 2 and 3 are in the "europe" group, and they share the scanning tasks for 3 data sources



BlueXP classification scanner groups can be defined as separate geographic areas where your data is stored.

You can deploy multiple BlueXP classification scanner nodes around the world and choose a scanner group for each node. In that way, each scanner node will scan the data that is the closest to it. The closer the scanner node is to the data, the better, because it reduces network latency as much as possible while scanning data.

You can choose which scanner groups to add to BlueXP classification and you can choose their names. BlueXP classification does not enforce that a node mapped to a scanner group named "europe" will be deployed in Europe.

You'll follow these steps to install additional BlueXP classification scanner nodes:

1. Prepare the Linux host systems that will act as the Scanner nodes
2. Download the Data Sense software to these Linux systems
3. Run a command on the Manager node to identify the Scanner nodes
4. Follow the steps to deploy the software on the Scanner nodes (and to optionally define a "scanner group" for certain Scanner nodes)
5. If you defined a scanner group, on the Manager node:
 - a. Open the file "working_environment_to_scanner_group_config.yml" and define the working environments that will be scanned by each scanner group
 - b. Run the following script to register this mapping information with all Scanner nodes:

```
update_we_scanner_group_from_config_file.sh
```

What you'll need

- Verify that all your Linux systems for Scanner nodes meet the host requirements.
- Verify that the systems have the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your environment meets the required permissions and connectivity.
- You must have the IP addresses of the Scanner node hosts that you are adding.
- You must have the IP address of the BlueXP classification Manager node host system
- You must have the IP address or host name of the Connector system, your NetApp Account ID, Connector Client ID, and user access token. If you're planning to use scanner groups, you'll need to know the Working Environment ID for each data source in your account. See **Prerequisite steps** below to get this information.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

- If you are using `firewalld` on your BlueXP classification machines, we recommend that you enable it before installing BlueXP classification. Run the following commands to configure `firewalld` so that it is compatible with BlueXP classification:

```

firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
firewall-cmd --reload

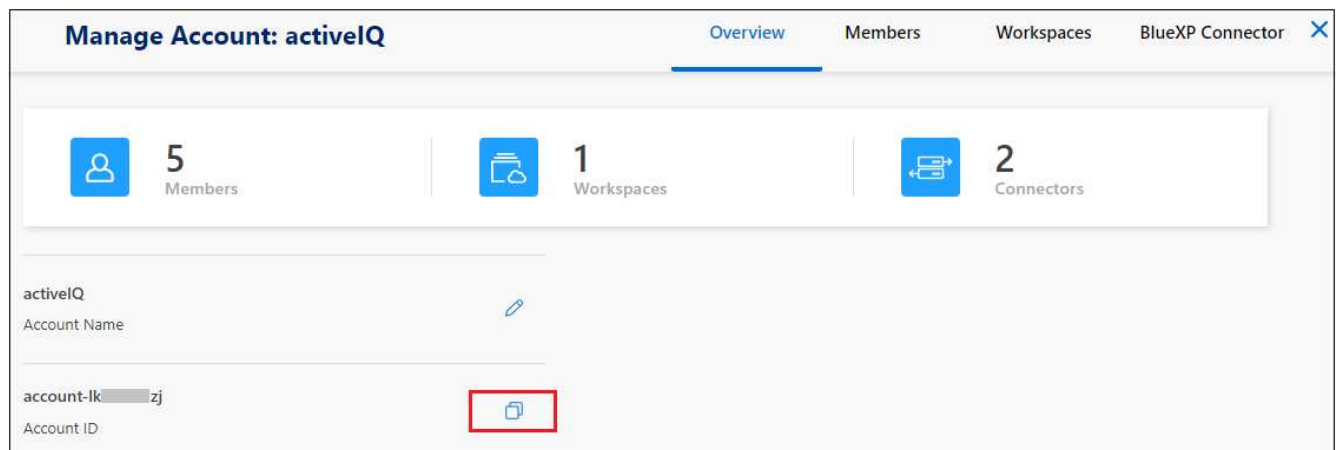
```

Note that you must restart Docker or Podman whenever you enable or update `firewalld` settings.

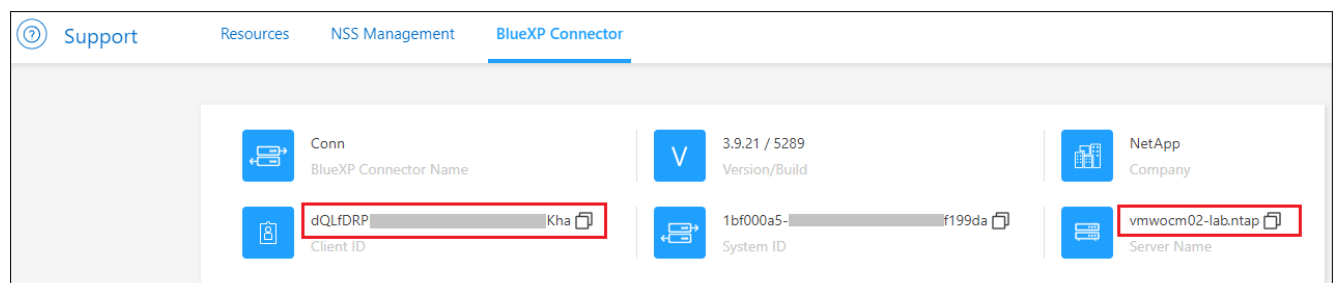
Prerequisite steps

Follow these steps to get the NetApp Account ID, Connector Client ID, Connector Server Name, and user access token that are required to add scanner nodes.

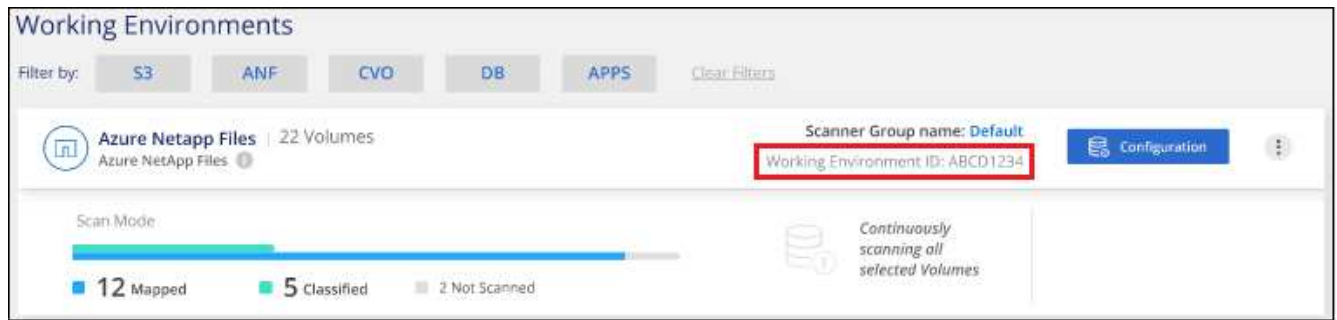
1. From the BlueXP menu bar, click **Account > Manage Accounts**.



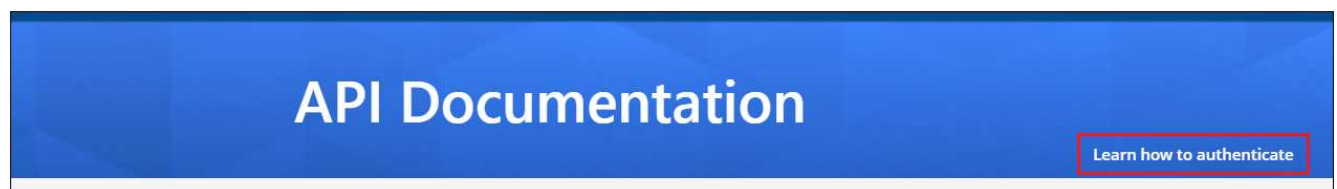
2. Copy the *Account ID*.
3. From the BlueXP menu bar, click **Help > Support > BlueXP Connector**.



- Copy the connector *Client ID* and the *Server Name*.
- If you're planning to use scanner groups, from the BlueXP classification Configuration tab, copy the Working Environment ID for each working environment that you plan to add to a scanner group.



- Go to the [API Documentation Developer Hub](#) and click **Learn how to authenticate**.



- Follow the authentication instructions, using the username and password of the account admin in the "username" and "password" parameters.
- Then copy the *access token* from the response.

Steps

- On the BlueXP classification Manager node, run the script "add_scanner_node.sh". For example, this command adds 2 scanner nodes:

```
sudo ./add_scanner_node.sh -a <account_id> -c <client_id> -m <cm_host> -h
<ds_manager_ip> -n <node_private_ip_1,node_private_ip_2> -t <user_token>
```

Variable values:

- account_id* = NetApp Account ID
 - client_id* = Connector Client ID (add the suffix "clients" to the client ID that you copied in the Prerequisite steps)
 - cm_host* = IP address or host name of the Connector system
 - ds_manager_ip* = Private IP address of the BlueXP classification Manager node system
 - node_private_ip* = IP addresses of the BlueXP classification Scanner node systems (multiple scanner node IPs are separated by a comma)
 - user_token* = JWT user access token
- Before the add_scanner_node script completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example: `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF1s35212 -u red95467j`) and save it in a text file.
 - On **each** scanner node host:
 - Copy the Data Sense installer file (**DATASENSE-INSTALLER-<version>.tar.gz**) to the host machine (using `scp` or some other method).

- b. Unzip the installer file.
- c. Paste and execute the command that you copied in step 2.
- d. If you want to add a scanner node into a "scanner group", add the parameter **-r <scanner_group_name>** to the command. Otherwise, the scanner node is added to the "default" group.

When the installation finishes on all scanner nodes and they have been joined to the manager node, the "add_scanner_node.sh" script finishes as well. The installation can take 10 to 20 minutes.

4. If you added any scanner nodes into a scanner group, return to the Manager node and perform the following 2 tasks:

- a. Open the file `/opt/netapp/config/custom_configuration/working_environment_to_scanner_group_config.yml` and enter the mapping for which scanner groups will scan specific working environments. You'll need to have the *Working Environment ID* for each data source. For example, the following entries add 2 working environments to the "europe" scanner group and 2 to the "united_states" scanner group:

```
scanner_groups:
  europe:
    working_environments:
      - "working_environment_id1"
      - "working_environment_id2"
  united_states:
    working_environments:
      - "working_environment_id3"
      - "working_environment_id4"
```

Any working environment that is not added to the list is scanned by the "default" group - you must have at least one manager or scanner node in the "default" group.

- b. Run the following script to register this mapping information with all Scanner nodes:
`/opt/netapp/Datasense/tools/update_we_scanner_group_from_config_file.sh`

Result

BlueXP classification is set up with Manager and Scanner nodes to scan all your data sources.

What's Next

From the Configuration page you can select the data sources that you want to scan - if you haven't already done that. If you created scanner groups, each data source is scanned by the Scanner nodes in the respective group.

You can see the Scanner Group name for each working environment in the Configuration page.

Working Environments

Filter by: S3 ANF CVO DB APPS [Clear Filters](#)

Azure Netapp Files | 22 Volumes
 Azure NetApp Files ⓘ

Scanner Group name: **Default**
 Working Environment ID: **ABCD1234** [Configuration](#)

Scan Mode

Continuously scanning all selected Volumes

12 Mapped 5 Classified 2 Not Scanned

You can also see the list of all scanner groups along with the IP address and status for each scanner node in the group in the bottom of the Configuration page.

Scanner Groups

Scanner Group: **Default** Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-██████.us-west-2.compute	172-██████	23/09/2022 14:32	Active	
ip-172-██████.us-west-2.compute	172-██████	23/09/2022 14:32	Active	

Scanner Group: **United_States** Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-██████.us-west-2.compute	172-██████	23/09/2022 14:32	Active	
ip-172-██████.us-west-2.compute	172-██████	23/09/2022 14:32	Active	

Scanner Group: **Europe** Scanner nodes

Copyright information

Copyright © 2024 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.