Prepare the hosts (Ubuntu or Debian)

StorageGRID 11.7

NetApp
January 09, 2024

This PDF was generated from https://docs.netapp.com/us-en/storagegrid-117/ubuntu/how-host-wide-settings-change.html on January 09, 2024. Always check docs.netapp.com for the latest.

Table of Contents

Prepare the hosts (Ubuntu or Debian)

How host-wide settings change during installation

On bare metal systems, StorageGRID makes some changes to host-wide `sysctl` settings.

The following changes are made:

```
# Recommended Cassandra setting: CASSANDRA-3563, CASSANDRA-13008, DataStax
documentation
vm.max_map_count = 1048575

# core file customization
# Note: for cores generated by binaries running inside containers, this
# path is interpreted relative to the container filesystem namespace.
# External cores will go nowhere, unless /var/local/core also exists on
# the host.
kernel.core_pattern = /var/local/core/%e.core.%p

# Set the kernel minimum free memory to the greater of the current value
or
# 512MiB if the host has 48GiB or less of RAM or 1.83GiB if the host has
more than 48GiB of RTAM
vm.min_free_kbytes = 524288

# Enforce current default swappiness value to ensure the VM system has
some
# flexibility to garbage collect behind anonymous mappings. Bump
watermark_scale_factor
# to help avoid OOM conditions in the kernel during memory allocation
bursts. Bump
# dirty_ratio to 90 because we explicitly fsync data that needs to be
persistent, and
# so do not require the dirty_ratio safety net. A low dirty_ratio combined
with a large
# working set (nr_active_pages) can cause us to enter synchronous I/O mode
unnecessarily,
# with deleterious effects on performance.
vm.swappiness = 60
vm.watermark_scale_factor = 200
vm.dirty_ratio = 90

# Turn off slow start after idle
net.ipv4.tcp_slow_start_after_idle = 0
```

```
# Tune TCP window settings to improve throughput
net.core.rmem_max = 8388608
net.core.wmem_max = 8388608
net.ipv4.tcp_rmem = 4096 524288 8388608
net.ipv4.tcp_wmem = 4096 262144 8388608
net.core.netdev_max_backlog = 2500

# Turn on MTU probing
net.ipv4.tcp_mtu_probing = 1

# Be more liberal with firewall connection tracking
net.ipv4.netfilter.ip_conntrack_tcp_be_liberal = 1

# Reduce TCP keepalive time to reasonable levels to terminate dead
connections
net.ipv4.tcp_keepalive_time = 270
net.ipv4.tcp_keepalive_probes = 3
net.ipv4.tcp_keepalive_intvl = 30

# Increase the ARP cache size to tolerate being in a /16 subnet
net.ipv4.neigh.default.gc_thresh1 = 8192
net.ipv4.neigh.default.gc_thresh2 = 32768
net.ipv4.neigh.default.gc_thresh3 = 65536
net.ipv6.neigh.default.gc_thresh1 = 8192
net.ipv6.neigh.default.gc_thresh2 = 32768
net.ipv6.neigh.default.gc_thresh3 = 65536

# Disable IP forwarding, we are not a router
net.ipv4.ip_forward = 0

# Follow security best practices for ignoring broadcast ping requests
net.ipv4.icmp_echo_ignore_broadcasts = 1

# Increase the pending connection and accept backlog to handle larger
connection bursts.
net.core.somaxconn=4096
net.ipv4.tcp_max_syn_backlog=4096
```

Install Linux

You must install Linux on all grid hosts. Use the NetApp Interoperability Matrix Tool (IMT) to get a list of supported versions.

Ensure that your operating system is upgraded to Linux kernel 4.15 or higher.

Steps

1. Install Linux on all physical or virtual grid hosts according to the distributor's instructions or your standard procedure.

   > Don't install any graphical desktop environments. When installing Ubuntu, you must select **standard system utilities**. Selecting **OpenSSH server** is recommended to enable ssh access to your Ubuntu hosts. All other options can remain cleared.

2. Ensure that all hosts have access to Ubuntu or Debian package repositories.
3. If swap is enabled:
   a. Run the following command: `$ sudo swapoff --all`
   b. Remove all swap entries from `/etc/fstab` to persist the settings.

      > Failing to disable swap entirely can severely lower performance.

Understand AppArmor profile installation

If you are operating in a self-deployed Ubuntu environment and using the AppArmor mandatory access control system, the AppArmor profiles associated with packages you install on the base system might be blocked by the corresponding packages installed with StorageGRID.

By default, AppArmor profiles are installed for packages that you install on the base operating system. When you run these packages from the StorageGRID system container, the AppArmor profiles are blocked. The DHCP, MySQL, NTP, and tcdump base packages conflict with AppArmor, and other base packages might also conflict.

You have two choices for handling AppArmor profiles:

- Disable individual profiles for the packages installed on the base system that overlap with the packages in the StorageGRID system container. When you disable individual profiles, an entry appears in the StorageGRID log files indicating that AppArmor is enabled.

  Use the following commands:

  ```
  sudo ln -s /etc/apparmor.d/<profile.name> /etc/apparmor.d/disable/
  sudo apparmor_parser -R /etc/apparmor.d/<profile.name>
  ```

  Example:

  ```
  sudo ln -s /etc/apparmor.d/bin.ping /etc/apparmor.d/disable/
  sudo apparmor_parser -R /etc/apparmor.d/bin.ping
  ```

- Disable AppArmor altogether. For Ubuntu 9.10 or later, follow the instructions in the Ubuntu online community: Disable AppArmor. Disabling AppArmor altogether might not be possible on newer Ubuntu versions.

Once you disable AppArmor, no entries indicating that AppArmor is enabled will appear in the StorageGRID log files.

Configure the host network (Ubuntu or Debian)

After completing the Linux installation on your hosts, you might need to perform some additional configuration to prepare a set of network interfaces on each host that are suitable for mapping into the StorageGRID nodes you will deploy later.

Before you begin

- You have reviewed the StorageGRID networking guidelines.
- You have reviewed the information about node container migration requirements.
- If you are using virtual hosts, you have read the considerations and recommendations for MAC address cloning before configuring the host network.

> If you are using VMs as hosts, you should select VMXNET 3 as the virtual network adapter. The VMware E1000 network adapter has caused connectivity issues with StorageGRID containers deployed on certain distributions of Linux.

About this task

Grid nodes must be able to access the Grid Network and, optionally, the Admin and Client Networks. You provide this access by creating mappings that associate the host's physical interface to the virtual interfaces for each grid node. When creating host interfaces, use friendly names to facilitate deployment across all hosts, and to enable migration.

The same interface can be shared between the host and one or more nodes. For example, you might use the same interface for host access and node Admin Network access, to facilitate host and node maintenance. Although the same interface can be shared between the host and individual nodes, all must have different IP addresses. IP addresses can't be shared between nodes or between the host and any node.

You can use the same host network interface to provide the Grid Network interface for all StorageGRID nodes on the host; you can use a different host network interface for each node; or you can do something in between. However, you would not typically provide the same host network interface as both the Grid and Admin Network interfaces for a single node, or as the Grid Network interface for one node and the Client Network interface for another.

You can complete this task in many ways. For example, if your hosts are virtual machines and you are deploying one or two StorageGRID nodes for each host, you can create the correct number of network interfaces in the hypervisor, and use a 1-to-1 mapping. If you are deploying multiple nodes on bare metal hosts for production use, you can leverage the Linux networking stack's support for VLAN and LACP for fault tolerance and bandwidth sharing. The following sections provide detailed approaches for both of these examples. You don't need to use either of these examples; you can use any approach that meets your needs.

> Don't use bond or bridge devices directly as the container network interface. Doing so could prevent node start-up caused by a kernel issue with the use of MACVLAN with bond and bridge devices in the container namespace. Instead, use a non-bond device, such as a VLAN or virtual Ethernet (veth) pair. Specify this device as the network interface in the node configuration file.

Considerations and recommendations for MAC address cloning

MAC address cloning causes the container to use the MAC address of the host, and the host to use the MAC address of either an address you specify or a randomly generated one. You should use MAC address cloning to avoid the use of promiscuous mode network configurations.

Enabling MAC cloning

In certain environments, security can be enhanced through MAC address cloning because it enables you to use a dedicated virtual NIC for the Admin Network, Grid Network, and Client Network. Having the container use the MAC address of the dedicated NIC on the host allows you to avoid using promiscuous mode network configurations.

> MAC address cloning is intended to be used with virtual server installations and might not function properly with all physical appliance configurations.

> If a node fails to start due to a MAC cloning targeted interface being busy, you might need to set the link to "down" before starting node. Additionally, it is possible that the virtual environment might prevent MAC cloning on a network interface while the link is up. If a node fails to set the MAC address and start due to an interface being busy, setting the link to "down" before starting the node might fix the issue.

MAC address cloning is disabled by default and must be set by node configuration keys. You should enable it when you install StorageGRID.

There is one key for each network:

- `ADMIN_NETWORK_TARGET_TYPE_INTERFACE_CLONE_MAC`
- `GRID_NETWORK_TARGET_TYPE_INTERFACE_CLONE_MAC`
- `CLIENT_NETWORK_TARGET_TYPE_INTERFACE_CLONE_MAC`

Setting the key to "true" causes the container to use the MAC address of the host's NIC. Additionally, the host will then use the MAC address of the specified container network. By default, the container address is a randomly generated address, but if you have set one using the `_NETWORK_MAC` node configuration key, that address is used instead. The host and container will always have different MAC addresses.

> Enabling MAC cloning on a virtual host without also enabling promiscuous mode on the hypervisor might cause Linux host networking using the host's interface to stop working.

MAC cloning use cases

There are two use cases to consider with MAC cloning:

- MAC cloning not enabled: When the `_CLONE_MAC` key in the node configuration file is not set, or set to "false," the host will use the host NIC MAC and the container will have a StorageGRID-generated MAC unless a MAC is specified in the `_NETWORK_MAC` key. If an address is set in the `_NETWORK_MAC` key, the container will have the address specified in the `_NETWORK_MAC` key. This configuration of keys requires the use of promiscuous mode.
- MAC cloning enabled: When the `_CLONE_MAC` key in the node configuration file is set to "true," the container uses the host NIC MAC, and the host uses a StorageGRID-generated MAC unless a MAC is

specified in the `_NETWORK_MAC` key. If an address is set in the `_NETWORK_MAC` key, the host uses the specified address instead of a generated one. In this configuration of keys, you should not use promiscuous mode.

> If you don't want to use MAC address cloning and would rather allow all interfaces to receive and transmit data for MAC addresses other than the ones assigned by the hypervisor, ensure that the security properties at the virtual switch and port group levels are set to **Accept** for Promiscuous Mode, MAC Address Changes, and Forged Transmits. The values set on the virtual switch can be overridden by the values at the port group level, so ensure that settings are the same in both places.

To enable MAC cloning, see the instructions for creating node configuration files.

MAC cloning example

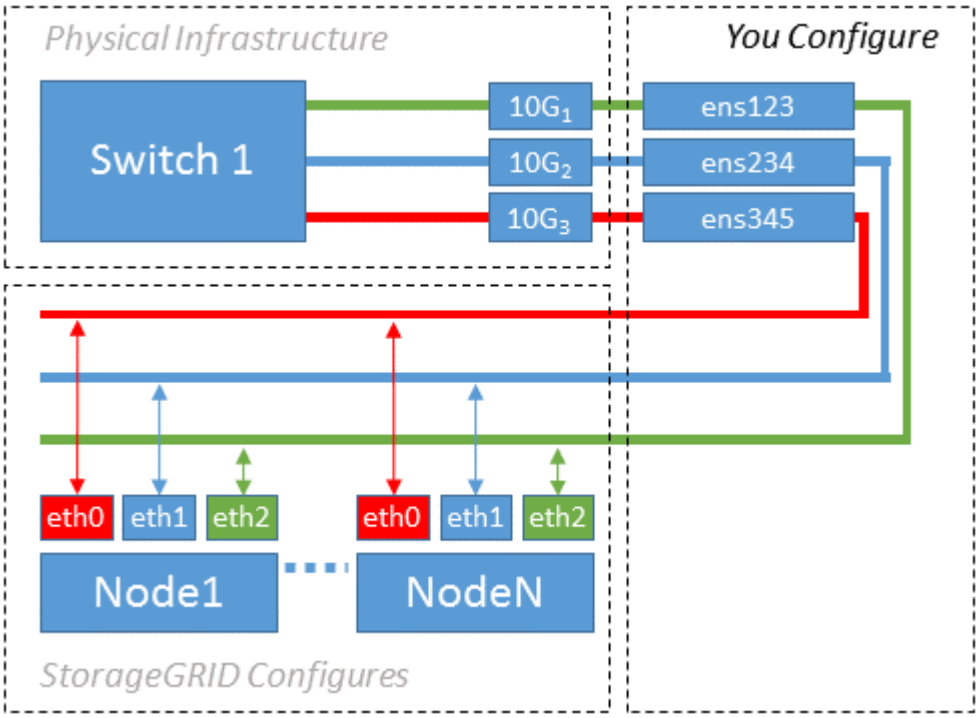
Example of MAC cloning enabled with a host having MAC address of 11:22:33:44:55:66 for the interface ens256 and the following keys in the node configuration file:

- `ADMIN_NETWORK_TARGET = ens256`
- `ADMIN_NETWORK_MAC = b2:9c:02:c2:27:10`
- `ADMIN_NETWORK_TARGET_TYPE_INTERFACE_CLONE_MAC = true`

Result: the host MAC for ens256 is b2:9c:02:c2:27:10 and the Admin Network MAC is 11:22:33:44:55:66

Example 1: 1-to-1 mapping to physical or virtual NICs

Example 1 describes a simple physical interface mapping that requires little or no host-side configuration.



The Linux operating system creates the ensXYZ interfaces automatically during installation or boot, or when the interfaces are hot-added. No configuration is required other than ensuring that the interfaces are set to

come up automatically after boot. You do have to determine which ensXYZ corresponds to which StorageGRID network (Grid, Admin, or Client) so you can provide the correct mappings later in the configuration process.

Note that the figure show multiple StorageGRID nodes; however, you would normally use this configuration for single-node VMs.

If Switch 1 is a physical switch, you should configure the ports connected to interfaces $10G_1$ through $10G_3$ for access mode, and place them on the appropriate VLANs.

Example 2: LACP bond carrying VLANs

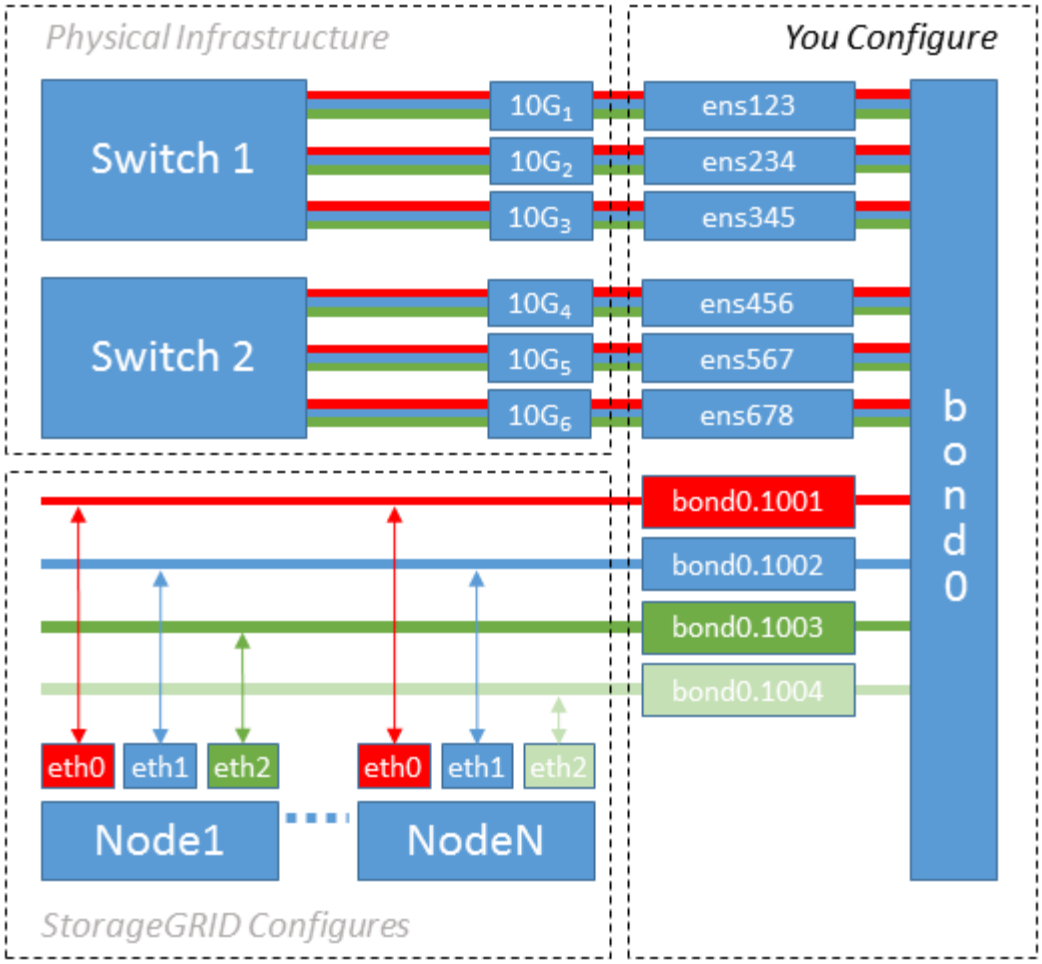
Example 2 assumes you are familiar with bonding network interfaces and with creating VLAN interfaces on the Linux distribution you are using.

About this task

Example 2 describes a generic, flexible, VLAN-based scheme that facilitates the sharing of all available network bandwidth across all nodes on a single host. This example is particularly applicable to bare metal hosts.

To understand this example, suppose you have three separate subnets for the Grid, Admin, and Client Networks at each data center. The subnets are on separate VLANs (1001, 1002, and 1003) and are presented to the host on a LACP-bonded trunk port (bond0). You would configure three VLAN interfaces on the bond: bond0.1001, bond0.1002, and bond0.1003.

If you require separate VLANs and subnets for node networks on the same host, you can add VLAN interfaces on the bond and map them into the host (shown as bond0.1004 in the illustration).



Steps

1. Aggregate all physical network interfaces that will be used for StorageGRID network connectivity into a single LACP bond.

   Use the same name for the bond on every host, for example, bond0.

2. Create VLAN interfaces that use this bond as their associated “physical device,” using the standard VLAN interface naming convention `physdev-name.VLAN ID`.

   Note that steps 1 and 2 require appropriate configuration on the edge switches terminating the other ends of the network links. The edge switch ports must also be aggregated into a LACP port channel, configured as a trunk, and allowed to pass all required VLANs.

   Sample interface configuration files for this per-host networking configuration scheme are provided.

Related information

Example /etc/network/interfaces

Configure host storage

You must allocate block storage volumes to each host.

Before you begin

You have reviewed the following topics, which provide information you need to accomplish this task:

Storage and performance requirements

Node container migration requirements

About this task

When allocating block storage volumes (LUNs) to hosts, use the tables in “Storage requirements” to determine the following:

- Number of volumes required for each host (based on the number and types of nodes that will be deployed on that host)
- Storage category for each volume (that is, System Data or Object Data)
- Size of each volume

You will use this information as well as the persistent name assigned by Linux to each physical volume when you deploy StorageGRID nodes on the host.

> ⓘ You don’t need to partition, format, or mount any of these volumes; you just need to ensure they are visible to the hosts.

Avoid using “raw” special device files (`/dev/sdb`, for example) as you compose your list of volume names. These files can change across reboots of the host, which will impact proper operation of the system. If you are using iSCSI LUNs and Device Mapper Multipathing, consider using multipath aliases in the `/dev/mapper` directory, especially if your SAN topology includes redundant network paths to the shared storage. Alternatively, you can use the system-created softlinks under `/dev/disk/by-path/` for your persistent device names.

For example:

```
ls -l
$ ls -l /dev/disk/by-path/
total 0
lrwxrwxrwx 1 root root  9 Sep 19 18:53 pci-0000:00:07.1-ata-2 -> ../../sr0
lrwxrwxrwx 1 root root  9 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:0:0 ->
../../sda
lrwxrwxrwx 1 root root 10 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:0:0-part1
-> ../../sda1
lrwxrwxrwx 1 root root 10 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:0:0-part2
-> ../../sda2
lrwxrwxrwx 1 root root  9 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:1:0 ->
../../sdb
lrwxrwxrwx 1 root root  9 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:2:0 ->
../../sdc
lrwxrwxrwx 1 root root  9 Sep 19 18:53 pci-0000:03:00.0-scsi-0:0:3:0 ->
../../sdd
```

Results will differ for each installation.

Assign friendly names to each of these block storage volumes to simplify the initial StorageGRID installation and future maintenance procedures. If you are using the device mapper multipath driver for redundant access to shared storage volumes, you can use the `alias` field in your `/etc/multipath.conf` file.

For example:

```
multipaths {
     multipath {
          wwid 3600a09800059d6df00005df2573c2c30
          alias docker-storage-volume-hostA
     }
     multipath {
          wwid 3600a09800059d6df00005df3573c2c30
          alias sgws-adm1-var-local
     }
     multipath {
          wwid 3600a09800059d6df00005df4573c2c30
          alias sgws-adm1-audit-logs
     }
     multipath {
          wwid 3600a09800059d6df00005df5573c2c30
          alias sgws-adm1-tables
     }
     multipath {
          wwid 3600a09800059d6df00005df6573c2c30
          alias sgws-gw1-var-local
     }
     multipath {
          wwid 3600a09800059d6df00005df7573c2c30
          alias sgws-sn1-var-local
     }
     multipath {
          wwid 3600a09800059d6df00005df7573c2c30
          alias sgws-sn1-rangedb-0
     }
     …
```

This will cause the aliases to appear as block devices in the `/dev/mapper` directory on the host, allowing you to specify a friendly, easily-validated name whenever a configuration or maintenance operation requires specifying a block storage volume.

> If you are setting up shared storage to support StorageGRID node migration and using Device Mapper Multipathing, you can create and install a common `/etc/multipath.conf` on all co-located hosts. Just make sure to use a different Docker storage volume on each host. Using aliases and including the target hostname in the alias for each Docker storage volume LUN will make this easy to remember and is recommended.

Related information

Storage and performance requirements

Node container migration requirements

Configure the Docker storage volume

Before installing Docker, you might need to format the Docker storage volume and mount it on `/var/lib/docker`.

About this task

You can skip these steps if you plan to use local storage for the Docker storage volume and have sufficient space available on the host partition containing `/var/lib`.

Steps

1. Create a file system on the Docker storage volume:

```
sudo mkfs.ext4 docker-storage-volume-device
```

2. Mount the Docker storage volume:

```
sudo mkdir -p /var/lib/docker
sudo mount docker-storage-volume-device /var/lib/docker
```

3. Add an entry for docker-storage-volume-device to /etc/fstab.

   This step ensures that the storage volume will remount automatically after host reboots.

Install Docker

The StorageGRID system runs on Linux as a collection of Docker containers. Before you can install StorageGRID, you must install Docker.

Steps

1. Install Docker by following the instructions for your Linux distribution.

   If Docker is not included with your Linux distribution, you can download it from the Docker website.

2. Ensure Docker has been enabled and started by running the following two commands:

```
sudo systemctl enable docker
```

```
sudo systemctl start docker
```

3. Confirm you have installed the expected version of Docker by entering the following:

```
sudo docker version
```

The Client and Server versions must be 1.11.0 or later.

Related information

Configure host storage

Install StorageGRID host services

You use the StorageGRID DEB package to install the StorageGRID host services.

About this task

These instructions describe how to install the host services from the DEB packages. As an alternative, you can use the APT repository metadata included in the installation archive to install the DEB packages remotely. See the APT repository instructions for your Linux operating system.

Steps

1. Copy the StorageGRID DEB packages to each of your hosts, or make them available on shared storage.

   For example, place them in the `/tmp` directory, so you can use the example command in the next step.

2. Log in to each host as root or using an account with sudo permission, and run the following commands.

   You must install the `images` package first, and the `service` package second. If you placed the packages in a directory other than `/tmp`, modify the command to reflect the path you used.

```
sudo dpkg --install /tmp/storagegrid-webscale-images-version-SHA.deb
```

```
sudo dpkg --install /tmp/storagegrid-webscale-service-version-SHA.deb
```

> Python 2.7 must already be installed before the StorageGRID packages can be installed. The `sudo dpkg --install /tmp/storagegrid-webscale-images-version-SHA.deb` command will fail until you have done so.

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.