Hewlett Packard
Enterprise

HPE Performance Cluster Manager Administration Guide

Abstract

This publication describes how to use the HPE Performance Cluster Manager 1.8 software to administer and maintain HPE cluster systems.

Part Number: 007-6499-013
Published: September 2022
Edition: 1

Notices

Acknowledgments

Product-specific acknowledgments

Revision history

Part number	Publication date	Edition	Summary of changes
007-6499-013	September 2022	1	Supports the HPE Performance Cluster Manager 1.8 release.
007-6499-012	April 2022	2	Supports the HPE Performance Cluster Manager 1.7 release. Edition 2 replaces edition 1 and includes information about how to deploy an HPE Superdome Flex server as a cluster node.
007-6499-012	March 2022	1	Supports the HPE Performance Cluster Manager 1.7 release.
007-6499-011	September 2021	1	Supports the HPE Performance Cluster Manager 1.6 release.
007-6499-010	March 2021	1	Supports the HPE Performance Cluster Manager 1.5 release.
007–6499-009	September 2020	1	Supports the HPE Performance Cluster Manager 1.4 release.
007–6499-008	April 2020	1	Supports the HPE Performance Cluster Manager 1.3.1 release.
007–6499-007	December 2019	1	Supports the HPE Performance Cluster Manager 1.3 release.
007–6499-006	June 2019	1	Supports the HPE Performance Cluster Manager 1.2 release.
007–6499-005	April 2019	1	Supports the HPE Performance Cluster Manager 1.1 release. Includes SLES 15 and CentOS 7.6 support.
007–6499-004	December 2018	1	Supports the HPE Performance Cluster Manager 1.1 release.
007–6499-003	September 2018	1	Supports the HPE Performance Cluster Manager 1.0 release, patch 3.
007–6499-002	June 2018	1	Supports the HPE Performance Cluster Manager 1.0 release, patch 1.
007–6499-001	June 2018	1	Original publication. Supports the HPE Performance Cluster Manager 1.0 release.

Contents

Administering a cluster with HPE Performance Cluster Manager

This guide is for system administrators of the HPE Performance Cluster Manager. Use the information in this guide to administer and maintain HPE clusters. The cluster manager release notes include a list of the specific hardware and operating systems that the cluster manager supports. This documentation typically uses the terms **cluster with leader nodes** and **clusters without leader nodes**. These clusters are as follows:

- A cluster with leader nodes includes the following types of nodes:
  - Admin node.
  - Leader nodes. The leader nodes add a communication hierarchy to a cluster system. A cluster with leader nodes includes one of the following types of leader nodes:
    - Scalable unit (SU) leader nodes. For example, HPE Cray EX clusters and HPE Apollo clusters can have SU leader nodes.
    - ICE leader nodes. All HPE SGI 8600 clusters have ICE leader nodes.
  - ICE compute nodes, which communicate to the admin node through an ICE leader node. Only clusters with ICE leader nodes have ICE compute nodes.
  - Compute nodes.
- A cluster without leader nodes includes the following types of nodes:
  - Admin node.
  - Compute nodes.

For more information about cluster terminology, see the glossary in the following:

HPE Performance Cluster Manager Getting Started Guide

NOTE: Unless otherwise noted, assume that information that pertains to RHEL platforms also pertains to Rocky Linux platforms and to CentOS platforms.

Cluster manager documentation

The following list shows the HPE Performance Cluster Manager documentation:

- The release notes contain feature information, platform requirements, and other release-specific guidance. To access the release notes, follow the links on the following website:

  https://www.hpe.com/software/hpcm

  On the product media, the release notes appear in a text file in the following directory:

  ```
  /docs
  ```

  Hewlett Packard Enterprise strongly recommends that you read the release notes, particularly the *Known Issues* section and the *Workarounds* section.

- The following guide presents an overview of the cluster manager and explains how to attach a factory-installed cluster to your site network:

HPE Performance Cluster Manager Getting Started Guide

- The bare-metal installation documentation is specific to each platform. These guides are as follows:
  - **HPE Performance Cluster Manager Installation Guide for Clusters With ICE Leader Nodes**
  - **HPE Performance Cluster Manager Installation Guide for Clusters With Scalable Unit (SU) Leader Nodes**
  - **HPE Performance Cluster Manager Installation Guide for Clusters Without Leader Nodes**
- The following guide explains the power management features included in the cluster manager:

  HPE Performance Cluster Manager Power Management Guide
- The following guide includes procedures and information about system-wide administration features:

  HPE Performance Cluster Manager Administration Guide
- The following guide includes procedures and information about system monitoring features:

  HPE Performance Cluster Manager System Monitoring Guide
- The following quick-start guide presents an overview of the installation process:

  HPE Performance Cluster Manager Installation Quick Start
- The following command reference shows the cluster manager commands and compares them with the commands used in the SGI Management Suite and in the HPE Insight Cluster Manager Utility:

  HPE Performance Cluster Manager Command Reference
- The following guide explains how to upgrade a cluster from an HPE Performance Cluster Manager 1.X release:

  HPE Performance Cluster Manager Upgrade Guide

After installation, the documentation resides on the system in the following directories:

- Release notes and user guides: `/opt/clmgr/doc`
- Manpages: `/opt/clmgr/man`

Feature descriptions for the HPE SGI 8600 system also apply to SGI ICE XA and SGI ICE systems.

NOTE: The cluster manager documentation includes examples where appropriate. Make sure to substitute information that pertains to your cluster when following the examples.

Using the `cm` commands

The following topics explain how to use the cluster manager `cm` commands.

Formatting `cm` commands and using tab completion

Many cluster manager commands are of the following form:

```
cm topic [subtopic ...] action parameters
```

The `cm` commands support tab completion for each *topic*, each *subtopic*, each *action*, and many parameters.

The `cm` commands implement tab completion for the `-i` *`image`* and the `--image` *`image`* parameters by comparing command input against the image names stored in the HPE Performance Cluster Manager database.

Likewise, the `cm` commands implement tab completion for the `-n` *`nodes`* and the `--nodes` *`nodes`* parameters by comparing command input against the node names stored in the HPE Performance Cluster Manager database.

Using wildcard characters

You can use wildcard characters in the cluster manager cm commands. If you use wildcards in the cm commands, enclose your specification in apostrophes (' '). The following table shows the most commonly used wildcard characters.

Wildcard	Effect
*	Matches one or more characters. For example, the following specifies all nodes in rack 1, chassis 1, tray 1 on an HPE Apollo 9000 cluster: 'r1c1t1n*'
?	Matches exactly one character. For example, the following specifies all nodes in rack 1 that have a single-character chassis: • On an HPE Apollo 9000 cluster: 'r1c?t*n*' • On an HPE SGI 8600 cluster: 'r1i?n*'
[ ]	Matches any of the range of characters specified within brackets. For example, the following specifies racks 11, 12, 13, and 14: 'rack1[1-4]'

The cluster manager includes the --confirm parameter, which evaluates and then displays a hostname regular expression before it runs the command. These actions let you decide whether to run the command or to halt the command so you can rewrite the command. For example:

```
# cm node show -n x3000*
x3000c0s33b1n0
x3000c0s33b2n0
x3000c0s33b3n0
x3000c0s33b4n0
# cm node show -n x3000* --confirm

This command will include the following node(s): x3000c0s33b[1-4]n0

continue [y|n]: y

x3000c0s33b1n0
x3000c0s33b2n0
x3000c0s33b3n0
x3000c0s33b4n0
```

The --exclude parameter lets you specify nodes to be omitted from an operation. This parameter prevents the command from running on specified nodes. When specified, the command applies the exclusion after processing all inclusions. For example:

```
# cm node show -n x3000*
x3000c0s33b1n0
x3000c0s33b2n0
x3000c0s33b3n0
x3000c0s33b4n0

# cm node show -n x3000* --exclude *b2*
x3000c0s33b1n0
```

```
x3000c0s33b3n0
x3000c0s33b4n0
```

Using of the @ symbol to specify custom groups

If you configure custom groups of nodes, you can operate on these custom node groups in a collective way with a single command. To specify a custom group on a command line, specify @*custom_group_name* in place of the *node* argument.

For example, the following command installs package zlib_devel on the SLES compute nodes in a custom group named comp:

```
# cm node zypper -n @comp install zlib_devel
```

Example *node* specifications

Many cm commands accept a -n *node* parameter. Generally, for *node*, you can specify one or more node hostnames. The following table shows example *node* specifications.

Specification	Nodes affected
admin	The admin node
n0	n0
n0,n34	n0 and n34
node?	All nodes that have node as the first four characters in the node name
node[13]	node13
node[10-14]	node10 through node14
node[001-022]	node001 through node022
node[2-6,20-26,36]	node2 through node6, node20 through node26, and node36
'node52*'	node520 through node529
@gpu-nodes	All nodes with graphics processing units (GPUs) that are configured into the custom group gpu-nodes

Node identification

The cluster manager recognizes distinct node hostnames for each type of cluster that it supports.

NOTE: The information in this topic shows the compute node names that the cluster manager assigns to nodes by default. This naming scheme identifies components by their location in the cluster. These names are assigned automatically when the compute nodes are configured into the cluster.

HPE Cray EX node identification

On HPE Cray EX supercomputers, the node name is in the following format:

`x`*CABINET*`c`*CHASSIS*`s`*SLOT*`b`*BLADE*`n`*NODE*

The variables are as follows:

Variable	Specification
CABINET	A 4-digit cabinet identifier in the range 1 <= *CABINET* <= 9999. Specific cabinet identifiers are as follows: • HPE Cray EX fluid-cooled compute: x1000 - x2999 • HPE Cray EX air-cooled I/O: x3000 - x4999 • HPE Cray EX air-cooled compute: x5000 - x5999 • HPE Cray EX TDS: x9000 • HPE Cray EX 2500: x8000 - x8999 Examples: `x1004`, `x3001`.
CHASSIS	A 1-digit chassis identifier in the range 0 <= *CHASSIS* <= 7. Examples: `c1`, `c7`.
SLOT	A 1-digit slot identifier in the range 0 <= *SLOT* <= 7. Examples: `s1`, `s4`.
BLADE	A 1-digit blade identifier in the range 0 <= *BLADE* <= 1. Examples: `b0`, `b1`.
NODE	A 1-digit node identifier in the range 0 <= *NODE* <= 1. Examples: `n0`, `n1`.

The following are node identification examples:

- `x9000c1s2b0n0` is a compute node.
- `fmn01` and `fmn02` are HPE Slingshot interconnect nodes.

HPE Cray EX switch identification

The default switch naming conventions are similar to the default node naming conventions. On HPE Cray EX supercomputers, the switch names are in the following format:

`x`*CABINET*`c`*CHASSIS*`r`*SWITCH*`b`*BMC*

The variables are as follows:

Variable	Specification
CABINET	A 4-digit rack identifier in the range 1 <= *CABINET* <= 9999. Examples: `x0046`, `x0178`.
CHASSIS	A 1-digit chassis identifier in the range 1 <= *CHASSIS* <= 4. Examples: `c1`, `c2`.
SWITCH	A 1-digit tray identifier in the range 0 <= *SWITCH* <= 7. Examples: `r5`, `r7`.
BMC	A 1-digit switch identifier in the range 0 <= *BMC* <= 1. Examples: `b0`, `b1`.

For example: `x1203c0r5b0` is a hostname for an HPE Cray EX switch controller.

HPE Apollo 9000 node identification

On HPE Apollo 9000 clusters, the node name is in one of the following formats:

`r`*`RACK`*`c`*`CHASSIS`*`t`*`TRAY`*`n`*`NODE`*

The variables are as follows:

Variable	Specification
RACK	A 3-digit rack identifier in the range 1 <= *RACK* <= 999. Examples: `r46`, `r178`.
CHASSIS	A 1-digit chassis identifier in the range 1 <= *CHASSIS* <= 4. Examples: `c1`, `c2`.
TRAY	A 1-digit tray identifier in the range 1 <= *TRAY* <= 8. Examples: `t5`, `t8`.
NODE	A 1-digit node identifier in the range 1 <= *NODE* <= 4. Examples: `n1`, `n4`.

For example: `r100c3t5n1`

HPE Apollo 9000 swich identification

The default switch naming conventions are similar to the default node naming conventions. On HPE Apollo 9000 clusters, the switch names are in the following format:

`r`*`RACK`*`c`*`CHASSIS`*`t`*`TRAY`*`s`*`SWITCH`*

The variables are as follows:

Variable	Specification
RACK	A 3-digit rack identifier in the range 1 <= *RACK* <= 999. Examples: `r46`, `r178`.
CHASSIS	A 1-digit chassis identifier in the range 1 <= *CHASSIS* <= 4. Examples: `c1`, `i2`.
TRAY	A 1-digit tray identifier in the range 1 <= *TRAY* <= 8. Examples: `t5`, `t8`.
SWITCH	A 1-digit switch identifier in the range 1 <= *SWITCH* <= 4. Examples: `s2`, `s3`.

HPE SGI 8600 node identification

On HPE SGI 8600 clusters, the node name is in the following format:

`r`*`RACK`*`i`*`CHASSIS`*`n`*`NODE`*

The variables are as follows:

Variable	Specification
RACK	A 3-digit rack identifier in the range 1 <= *RACK* <= 999. Examples: `r46`, `r178`.
CHASSIS	A 1-digit chassis identifier in the range 0 <= *CHASSIS* <= 3. Examples: `i1`, `i2`.
NODE	A 2-digit node identifier in the range 0 <= *NODE* <= 35. Examples: `n2`, `n33`.

For example: `r10i0n31`

HPE SGI 8600 switch identification

The default switch naming conventions are similar to the default node naming conventions. On HPE SGI 8600 clusters, the switch names are in the following format:

`r`*`RACK`*`i`*`CHASSIS`*`s`*`SWITCH`*

The variables are as follows:

Variable	Specification
RACK	A 3-digit rack identifier in the range 1 <= *RACK* <= 999. Examples: `r46`, `r178`.
CHASSIS	A 1-digit chassis controller identifier in the range 0 <= *CHASSIS* <= 3. Examples: `i1`, `i2`.
SWITCH	A 2-digit switch identifier in the range 0 <= *SWITCH* <= 1. Examples: `s0`, `s1`.

Cluster manager videos

The following videos show HPE Performance Cluster Manager functionality:

- **Cluster manager overview**
- **Cluster manager integration with NVIDIA DCGM**
- **Workload management using the cluster manager and Altair PBS Professional**
- **Service Infrastructure Monitoring with Grafana**
- **AIOps in production**

Identifying the cluster manager release that is installed

Procedure

1. Log into the admin node as the root user.
2. Enter one of the following commands:

- # **cat /etc/*release**

  This command displays information about operation system distribution files and the cluster manager release.

- # **cat /etc/sgi*release**

- # **cadmin --version**

- # **cadmin --ver**

Configuring optional features on compute nodes

You can configure several features on compute nodes. The information in this manual supplements the documentation from the operating system distributions and where needed, explains extra steps needed for the cluster.

Configuring scratch disk space on an admin node

By default, the cluster manager does not allocate disk space for customer use on the system disk of the admin node.

To allocate scratch disk space on the system disk of an admin node, add the following parameters to the kernel parameter list of the admin node installer:

- `destroy_disk_label=yes`
- `root_disk_reserve=`*`size`*.

  For *size*, specify a size in GiB

As a result, the cluster manager creates the scratch disk space in partition 61. Notice that you must configure the scratch disk space. That is, you must create the file system, add the `fstab` entries, and complete other tasks.

Configuring scratch disk space on leader nodes and on compute nodes

By default, the cluster manager does not allocate disk space for customer use on the system disks of cluster nodes.

The following procedure uses the `cm node provision` command to allocate scratch disk space on leader nodes and compute nodes. This procedure uses existing disk drives to allocate space within the cluster for specific needs. No additional hardware is needed.

Procedure

1. Log into the admin node as the root user.
2. Open the following script file with a text editor, edit the script to customize the attributes of the disk space, and save the file:

   ```
   /opt/clmgr/image/scripts/post-install/50all.create_filesystem_for_reserved_partition
   ```

   The following notes and restrictions apply:

   - You must modify this script. For example, you can change the `ext4` file system to an XFS file system.
   - If you add your own set of partitions, ensure that they start at 61.

   You might have to repeat this procedure, experimenting with the settings in this file, until the result matches your expectations.
3. (Conditional) Synchronize the compute node image with the scalable unit (SU) leader node image.

   Complete this step if the cluster has SU leader nodes, and `rsync` is the transport method.

   Use the `su-sync-image` command in the following format:

   ```
   su-sync-image image_name
   ```

For *image_name*, specify the image name. For example:

```
# su-sync-image rhel8.X
```

4. Use the `cm node provision` command in the following format to provide the primary specifications:

   ```
   cm node provision -n node --force-disk dev --wipe-disk --root-disk-reserve nn -s
   ```

   The variables are as follows:

Variable	Specification
node	The hostname name of one or more cluster nodes.
dev	The disk device name. For example, `/dev/sdz`
nn	The disk space in GB.

   When you reboot the affected nodes, a post-install script creates scratch disk space on the nodes with the following default attributes:

   - A single partition, partition number 61.
   - Disk space of *nn*GB.
   - An `ext4` file system.
   - Mount point `/scratch` in the image.
   - Persistent across installs.

Configuring additional storage for scalable unit (SU) leader nodes or for compute nodes

The procedure in this topic explains how to add additional, external storage to SU leader nodes and to compute nodes. The added storage is external to the node, separate from the root disk. Complete the procedure in this topic if you want to use additional, non-system disk drives and mount them on your system.

If the nodes are configured to boot from disk, complete this procedure before you install an image on the nodes.

This feature is compatible with multiple slots.

When you implement this feature, you can separate your monitoring data, logging data, and other data from the system disk.

NOTE: The cluster manager does not support this feature on ICE leader nodes or ICE compute nodes. You cannot add additional storage to ICE leader nodes or ICE compute nodes.

Procedure

1. Verify the hardware path to the external storage device.

   This procedure requires you to specify the path using the hardware path used to access the device. For example:

   ```
   /dev/disk/by-path/pci-0000:00:05.0-scsi-0:0:0:1
   ```

Do not use device names of the form `/dev/sdb` to specify the storage device because that naming format is not boot persistent on all platforms.

2. Open the following file in a text editor, use the documentation in the file to customize the specifications, and save the file:

   ```
   /opt/clmgr/image/scripts/pre-install/55all.setup_additional_storage_before_provision-example
   ```

   You might have to repeat this procedure, experimenting with the settings in this file, until the result matches your expectations.

3. Use the `cm node set` command in the following format to specify the extra storage:

   ```
   cm node set -n node --kernel-extra-params "additional_storage=/path_to_storage_device"
   ```

   The variables are as follows:

Variable	Specification
node	One or more node hostnames.
path_to_storage_device	For example: `/dev/disk/by-path/pci-0000:00:05.0-scsi-0:0:0:1`

   The `cm node set` command in this step runs the `55all.setup_additional_storage_before_provision-example` script.

4. (Conditional) Synchronize the compute node image with the scalable unit (SU) leader node image.

   Complete this step if the cluster has SU leader nodes, and `rsync` is the transport method.

   Use the `su-sync-image` command in the following format:

   ```
   su-sync-image image_name
   ```

   For *image_name*, specify the image name. For example:

   ```
   # su-sync-image rhel8.X
   ```

Configuring the system dump utility (SDU) and HPE remote device access (RDA)

The SDU and HPE RDA tools facilitate remote support and diagnostics for the cluster manager. These are optional tools. SDU enables the copying of files to HPE via RDA using the outbox feature.

To access the RDA documentation, see the following:

https://midway.ext.hpe.com/home

Procedure

1. Contact your HPE support representative before you complete the steps that follow.

2. Verify that your site has HTTPS outbound enabled.

   HTTPS outbound is required when using HPE RDA to upload an SDU `tar` file.

3. Log into the admin node as the root user.

4. Install the `cm-sdu` RPM.

Enter one of the following commands:

- On RHEL 8.X operating systems, enter the following command:

  ```
  # cm node dnf -n admin install cm-sdu
  ```

- On SLES 15 SPX operating systems, enter the following command:

  ```
  # cm node zypper -n admin install cm-sdu
  ```

5. Complete the following steps to enable the `cray-sdu-rda` service:

   a. Enter the following command to display the name of the SDU/RDA image:

   ```
   # grep SDU_IMAGE /etc/sysconfig/cray-sdu-rda
   ```

   The following example output shows the name of the SDU/RDA image as `cray-sdu-rda`:

   ```
   SDU_IMAGE="registry.local/cray-sdu-rda/cray-sdu-rda:1.X.X"
   ```

   b. Enter the following command, and verify that the output from the `podman` command matches the name of the SDU/RDA image:

   ```
   # podman image ls
   ```

   The following example output shows the same name and version tag as in the `grep` command output:

   ```
   REPOSITORY                               TAG   IMAGE        ID CREATED    SIZE
   docker.io/library/cray-sdu-rda           1.X.X a2d7315bfebe 3 months ago 815 MB
   registry.local/cray-sdu-rda/cray-sdu-rda 1.X.X a2d7315bfebe 3 months ago 815 MB
   ```

6. Enter the following commands to enable and start the `cray-sdu-rda` service:

   ```
   # systemctl enable cray-sdu-rda
   # systemctl start cray-sdu-rda
   ```

7. Enter the following commands to verify that the `cray-sdu-rda` service is running:

   ```
   # sdu is_service_ready
   # sdu wait_for_service
   ```

   For example:

   ```
   # sdu is_service_ready
   running
   # sdu wait_for_service
   running
   ```

8. Enter the following command to display the admin node product number and the cluster serial number:

   ```
   # dmidecode -t system | grep -i -e product -e serial
   Product Name: ProLiant DLXXX GenXX
   Serial Number: ABCD12345678
   ```

9. Enter the `sdu bash` command in the following format to enable RDA inside the container:

   ```
   sdu bash cray-rda-setup enable [-p product_name] -s serial_number
   ```

   The variables are as follows:

Variable	Specification
product_name	The admin node product name as displayed in the dmidecode command output. Enter the product name as a single alphanumeric string without any spaces. If you do not specify the product name, the system prompts you for the product name. Be aware that the prompt, however, refers to the product name as the PRODUCT NUMBER. **NOTE:** If the dmidecode command returns a Product Name that includes space characters, omit the space characters when you specify the *product_name*.
serial_number	Required. The cluster serial number displayed in the dmidecode command output.

For example:

```
# sdu bash cray-rda-setup enable -s ABCD12345678
Enter product number (required) []: ProLiant DL380 Gen10
PRODUCT_NUMBER must be alphanumeric
Enter product number (required) []: ProLiantDL380Gen10

Creating new /etc/rda/dmi.dat and removing old client key and certificate

=== RDA Setup Utility ===
Apply file ownership and permissions... rda-setup-files: Setup RDA file owner and permissions

Done
.
.
.
Done!
```

Using the GUI to administer the cluster

The cluster manager GUI enables you to complete cluster management tasks on one or more nodes. The tasks you can complete depend on your privileges and the number of selected nodes, as follows:.

- A root user can view all the GUI displays and can complete all GUI-enabled tasks.
- A nonroot user can view the cluster definition and the monitoring information. A nonroot user cannot use the GUI to complete any tasks.
- If the GUI is configured for anonymous access, an anonymous user can view the cluster definition and the monitoring information. If the GUI is not configured for anonymous access, an anonymous user cannot interact with the GUI. For information about anonymous access, see the following:

  Starting the GUI without authentication

Starting the GUI

The cluster manager GUI is a Java application. You can download a copy of the GUI client to a computer that is connected to the admin node over your site network. For example, you can download the client to a laptop or desktop computer.

To close an unwanted dialog window, press the `ESC` key.

Procedure

1. On the client computer, verify that Java 8.X is installed.
2. Open a web browser on your client computer.
3. In the browser address bar, enter the address of the cluster admin node.

   Use the following format:

   `https://`*`admin_node_addr`*

   For *admin_node_addr*, specify the IP address or the fully qualified domain name (FQDN) of the cluster admin node.

   If the URL you enter in the address bar begins with `http` rather than `https`, the cluster manager redirects to `https`.
4. On the cluster manager home page, click **Launch HPE Performance Cluster Manager GUI**.
5. In the **Login** pop-up window, enter the following:

   - In the **Login** field, enter the cluster user name. By default, this is `root`.
   - In the **Password** field, enter the cluster password. By default, this is `cmdefault`.

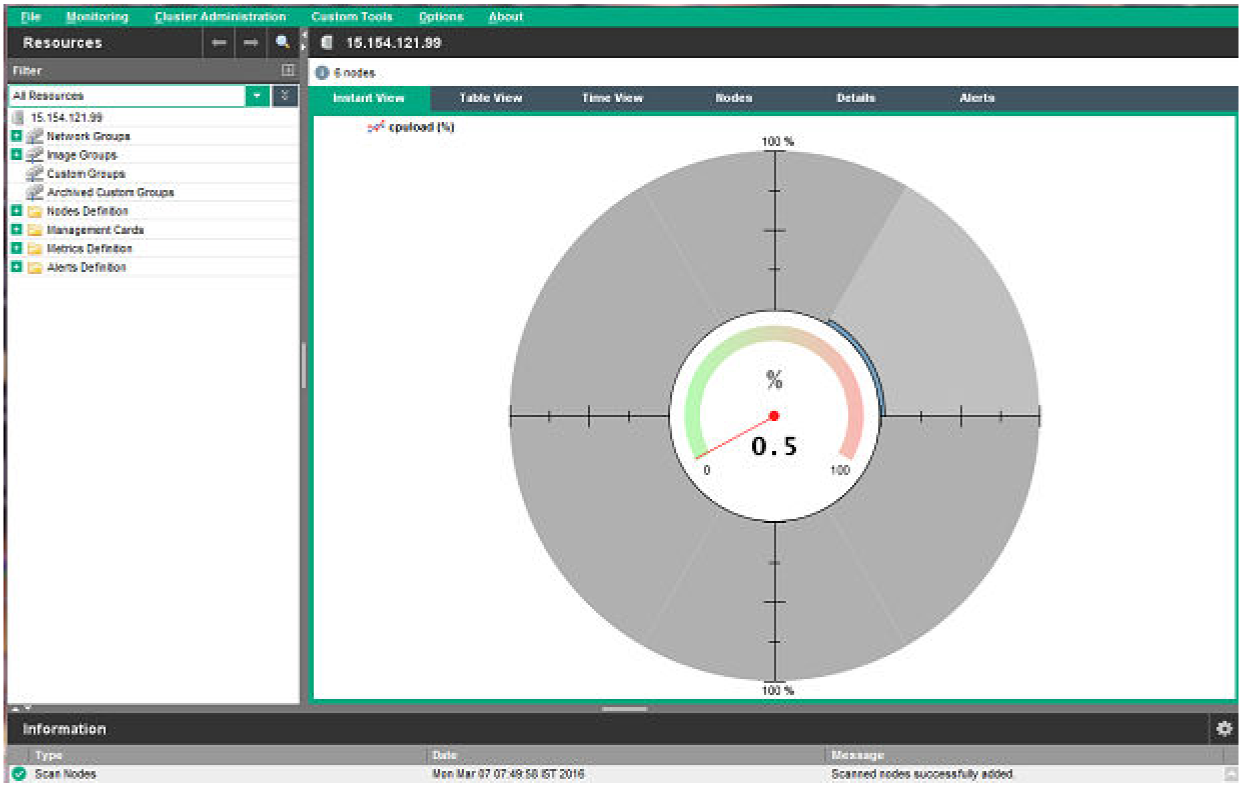
   NOTE: You can create one or more user accounts for GUI access. For information, see the following:

   Creating additional user accounts
6. Follow the onscreen prompts to launch the GUI.

   Depending on your configuration, the cluster manager might prompt you to enter more information. For example:

- The cluster manager might prompt you to enter the IP address of the admin node.
- Your client computer might have more than one network interface. In this case, the cluster manager might prompt you to enter the correct network interface to use for communication with the admin node.

7. Observe the main window and notice the functional areas.



Type	Date	Message
Scan Nodes	Mon Mar 07 07:49:58 IST 2016	Scanned nodes successfully added.

Figure 1: Main window

The main window contains the following functional areas:

- The top bar. This bar allows you to click or select GUI actions.
- The left pane. This pane lists resources such as **Network Groups**, **Image Groups**, and **Nodes Definitions**.

  Click the **+** button to expand a resource.

  The left pane contains a filter that allows you to select resources for display.
- The right pane, which displays the global cluster view.
- The bottom pane, which displays log information.

8. (Optional) Click **Options** > **Properties** to adjust the GUI to suit your viewing preferences.

There are several settings you can adjust to change font sizes, colors, and other aspects of the visual representation.

If you adjust any settings, exit the GUI and restart the GUI. The cluster manager records your preferences and reloads them the next time you launch the GUI.

9. (Optional) Assume the administrator role.

Administrator privileges enable you to perform cluster configuration tasks. You can perform cluster configuration tasks on only one instance of the GUI at a time. Complete the following steps:

- Click **Options** > **Enter Admin mode**.
- On the popup window that appears, provide the cluster administrator login credentials and click **OK**.

To exit administrator mode, click **Options** > **Leave Admin mode**.

The GUI is a Java application. You can use Java Web Start to download the GUI from the web server that runs on the admin node. Then, manually copy the GUI Java file onto the client.

The following file shows some of the ports that the cluster manager uses:

```
/opt/clmgr/etc/cmuserver.conf
```

To change any of the ports in this file, search in the `cmuserver.conf` file for the variables start with the letters `CMU_HTTP` and modify them for your site.

Starting the GUI without authentication

By default, the cluster manager GUI prompts user to authenticate before it presents cluster manager information. If you want the GUI to present information without prompting the user to enter a username and password, complete the procedure in this topic. Without authentication, the GUI presents information in read-only form.

Procedure

1. Log into the admin node as the root user.
2. Open the following file:

   ```
   /opt/clmgr/etc/cmuserver.conf
   ```
3. Search for `CMU_JAVA_SERVER_ARGS` within the `cmuserver.conf` file.
4. Add the following string to the existing list of `CMU_JAVA_SERVER_ARGS` parameters:

   ```
   -Dcmu.gui.anonymousAccess=true
   ```
5. Save and close the `cmuserver.conf` file.
6. Restart the `cmdb` service:

   ```
   # systemctl restart cmdb.service
   ```

Creating additional user accounts

You can create user accounts for specific reasons. For example, you can create an account for users to use when they log into the graphical user interface (GUI).

Procedure

1. Log into the admin node as the root user.
2. Create a new Linux user.

   For example:

   ```
   # useradd username
   ```

   For *username*, specify the name for the account.

3. Open the following file:

   ```
   /opt/clmgr/etc/admins
   ```

4. Use the instructions in the `admins` file to configure the user account.

   For example, to configure an account for users to use when they want to log into the GUI, grant the `GUI` permission.

5. Save and close the `admins` file.

Administrator menu

The GUI has the following operator modes:

- Normal mode.

  In normal mode, the GUI allows you to monitor node status and visualize static data. You cannot perform any other action on the cluster nodes in normal mode. This helps prevent unauthorized users from performing actions that can be harmful to a node.

- Administrator mode.

  In administrator mode, you can perform all the actions available with normal mode plus additional administrative actions.

  With one or more nodes selected in the left panel, right-click to access a contextual menu. This menu allows you to perform actions on selected nodes.

  The contextual menu is available in network group, image group, and custom group views. This menu is also accessible by right-clicking in the overview frame.

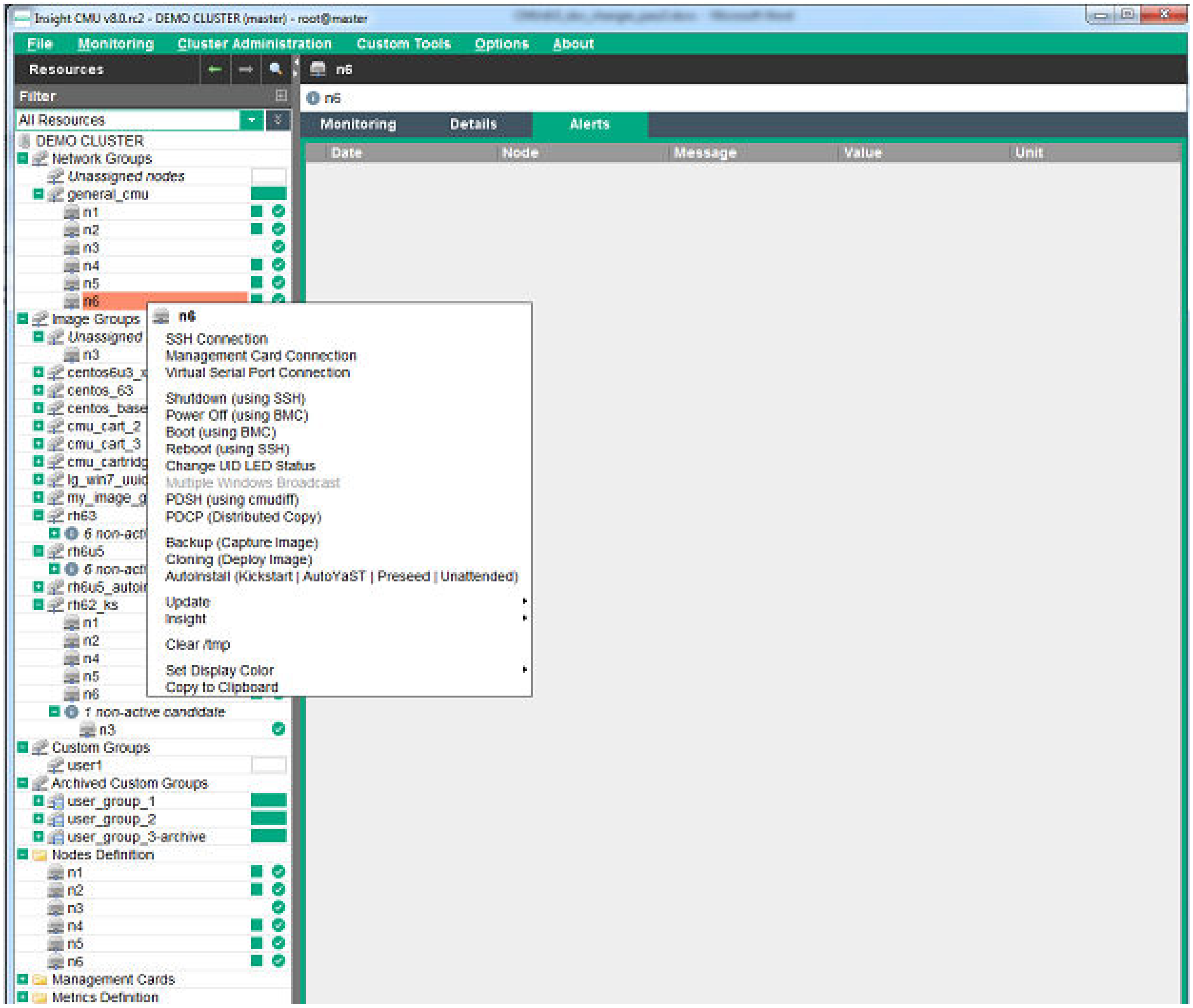


Figure 2: Contextual menu for administrator mode

For information about entering administrator mode, see the following:

Starting the GUI

Adding nodes

The **Add Node** menu item adds one node to the cluster.

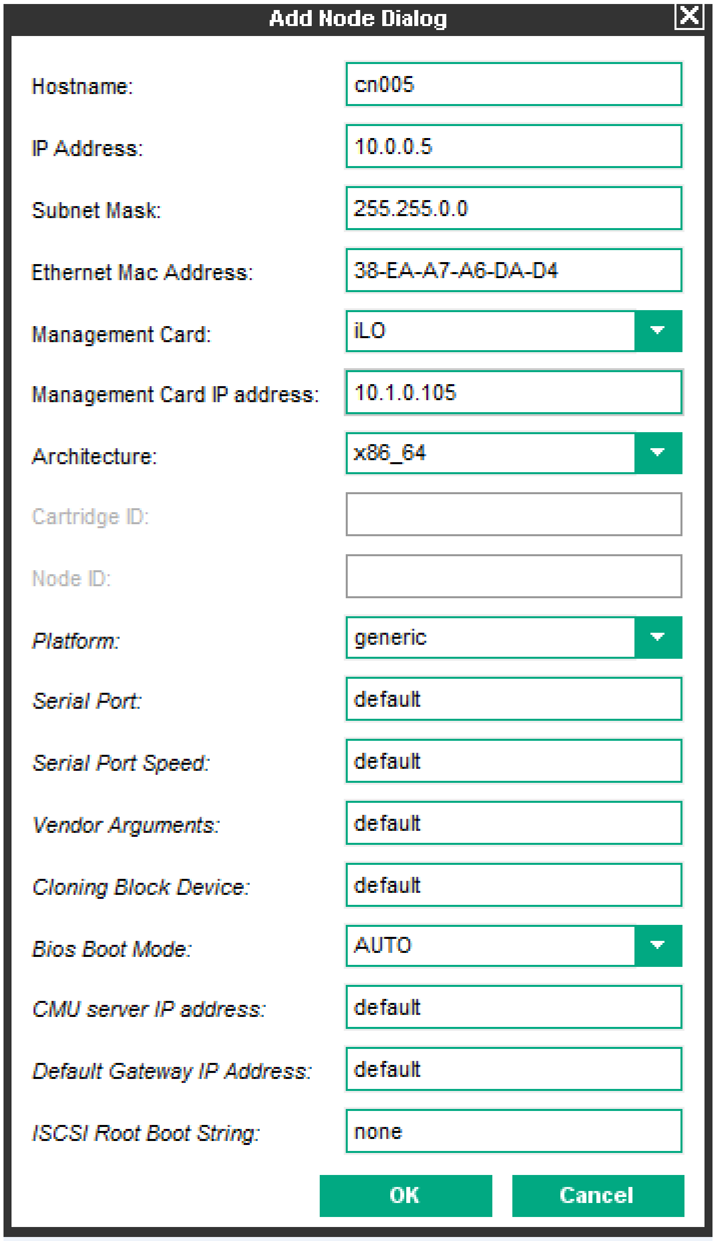
Prerequisites

Know the MAC address of the node you want to add to the cluster.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Node Management**.
3. In the right pane, click **Add Node**.

   The following dialog box displays.


Add Node Dialog

Hostname:	cn005
IP Address:	10.0.0.5
Subnet Mask:	255.255.0.0
Ethernet Mac Address:	38-EA-A7-A6-DA-D4
Management Card:	iLO
Management Card IP address:	10.1.0.105
Architecture:	x86_64
Cartridge ID:	
Node ID:	
Platform:	generic
Serial Port:	default
Serial Port Speed:	default
Vendor Arguments:	default
Cloning Block Device:	default
Bios Boot Mode:	AUTO
CMU server IP address:	default
Default Gateway IP Address:	default
ISCSI Root Boot String:	none

Figure 3: Add node dialog

4. Complete the fields in the dialog box and click **OK**.

   A dialog box displays the successful addition of a node completion.

5. On the **Add another node** popup, click **Yes** or **No** in the dialog box that asks if you want to add another node. Proceed as follows:

   Proceed as follows:

   - If you are finished adding nodes, click **No** and proceed to the next step in this procedure.
   - If you want to add another node, click **Yes** and repeat the preceding steps in this procedure.

6. On the **Update Configs** popup that asks **Do you want to run update-configs now?**, click **Yes** to update the configuration.

7. After the **Update Configs** terminal window appears, complete the following steps:

- Click in the terminal window.
- Press **Enter** to close the window.

8. Add the new node or nodes to a network group.

   Proceed to the following:

   Adding network groups from the GUI

Modifying nodes

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Node Management**.
3. In the right pane, right-click the node you want to modify and select **Manage** > **Modify Node**.

   The **Modify Node Dialog** window displays.
4. Update information for the node in the **Modify Node Dialog** box, and click **OK**.

   You cannot change the name of the node in the **Modify Node Dialog** dialog box.
5. On the **Update Configs** popup that asks **Do you want to run update-configs now?**, click **Yes** to update the configuration.
6. After the **Update Configs** terminal window appears, complete the following steps:

   - Click in the terminal window.
   - Press **Enter** to close the window.

Modifying software on a compute node

The procedure in this topic explains how to install, remove, or update software packages on one or more compute nodes. Use this task to update RPMs on nodes if there are newer versions of the RPMs available.

Procedure

1. Log into the admin node as the root user.
2. Use `ls` or another command to make sure that the RPMs you want to change reside in the following directory:

   `/opt/clmgr/image/rpmlists/generated/`
3. (Conditional) Add a repository.

   Complete this step if the repository you want to add or update is not present.

   Use the `cm repo add` and the `cm repo select` commands.
4. Enter the following command to verify that the repository is selected:

   ```
   # cm repo show
   ```

   The command displays selected repositories with an asterisk (`*`) in column 1.

5. (Conditional) Use the `cm repo select` command to select a repository.

   Complete this step if the repository or repository group is not selected.

6. Click **Options** > **Enter Admin mode** and enter the cluster credentials.

7. Click **Cluster Administration** > **Node Management**.

8. In the left pane, select the node or nodes that you want to modify.

   If necessary, click **+** to expand the node group that includes the node.

9. Right click the node(s), and select one of the following:

   - **Node RPMs** > **Refresh RPMs**.

     For example, select **Refresh RPMs** if you had deleted an RPM on a compute node.

   - **Node RPMs** > **Update RPMs**

     For example, select **Update RPMs** to upgrade node RPMs to a higher version after a patch or new release.

10. Specify an RPM or RPM group in one of the following ways.

    - In the **RPM list** field, enter the name of the RPM. Alternatively, click the folder icon to browse the list of available RPMs, and select the RPM.
    - Click the **Select repositories** checkbox. The preselected repositories appear in orange after you click the **Select repositories** checkbox. If the repository that you want to update is not highlighted in orange, enter the `cm repo select` command in a command window to select that repository. When you click **OK**, the cluster manager updates all repositories that appear in orange in the window.
    - Click the **Select repository groups** checkbox. The preselected repository groups appear in orange after you click the **Select repository groups** checkbox. If the repository group that you want to update is not highlighted in orange, enter the `cm repo select` command in a command window to select that repository group. When you click **OK**, the cluster manager updates all repository groups that appear in orange in the window.

11. Click **OK**.

Importing nodes

The procedure that follows explains the following:

- How to create a text file with node information.
- How to direct the cluster manager to use the information in that file to import those nodes. All imported nodes belong to the default image group.

Prerequisites

Make sure that the file is formatted correctly. Incorrect formatting can break the operation.

Procedure

1. Create a text file that describes the node information for the node that you want to import into the cluster database.

   Within the file, enter information for an individual node all on one line. The fields in the line are as follows:

Node name
IP address
Subnet mask
MAC address
Image name

Management card IP address
Management card type
Architecture
Cartridge ID
Node ID
Platform name
Serial port
Serial port speed
Vendor arguments
Cloning bock device
BIOS mode
Management server IP address
Default gateway
ISCSI root

Management card MAC address

For example:

```
n0 10.117.31.2 255.255.255.0 c8-cb-b8-cb-d4-c6 rhel8.X 10.117.30.2 IPMI
x86_64 -1 -1 generic ttyS1 115200 "default" default auto default default
"none" c8-cb-b8-cb-d4-c9
n1 10.117.31.3 255.255.255.0 c8-cb-b8-cb-d5-96 rhel8.X 10.117.30.3 IPMI
x86_64 -1 -1 generic ttyS1 115200 "default" default auto default default
"none" c8-cb-b8-cb-d5-99
n2 10.117.31.4 255.255.255.0 c8-cb-b8-c6-c8-b8 rhel8.X 10.117.30.4 IPMI
x86_64 -1 -1 generic ttyS1 115200 "default" default auto default default
"none" c8-cb-b8-c6-c8-bb
```

2. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
3. Click **Cluster Administration** > **Node Management**.
4. In the right pane, click **Import Compute Nodes**.
5. Browse to the text file you created.
6. Click **Open** to add the nodes from the file to the cluster.
7. On the **Update Configs** popup that asks **Do you want to run update-configs now?**, click **Yes** to update the configuration.
8. After the **Update Configs** terminal window appears, complete the following steps:

    - Click in the terminal window.
    - Press **Enter** to close the window.

Deleting nodes

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the right pane, right-click the node(s) you want to delete and select **Manage** > **Delete Node(s)**.

   To select more than one node for deletion, do one of the following:

   - Press Shift and click the left mouse button.

     or

   - Press Ctrl and click the left mouse button.

   After you delete a node, it cannot be recovered. You can, however, add the node back in.
3. On the **Delete Node(s)** popup, click **OK** to delete the node.
4. On the **Update Configs** popup that asks **Do you want to run update-configs now?**, click **Yes** to update the configuration.
5. After the **Update Configs** terminal window appears, complete the following steps:

   - Click in the terminal window.
   - Press **Enter** to close the window.

Provisioning a compute node or an ICE leader node by using the GUI

Procedure

1. Create an image using the command-line based method in the following:

   Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

   Observe the following rules:

   - The image must be compatible with the node hardware.
   - The nodes are ready to be powered on by the node controller.

   NOTE: Do not complete this procedure for scalable unit (SU) leader nodes. For information, see the following:

   Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters
2. In the left pane, select the node, nodes, or node group that you want to host the image.
3. Right-click the nodes, and select **Provision Image (Deploy)**.
4. In the **Provision Image** popup window, complete the following fields and click **OK**:

   - In the **Image group** field, select the new image group.
   - In the **Kernel** field, select the kernel that you want to deploy on the selected nodes.

- In the **Rootfs type** field, select either **disk** or **tmpfs**.
- To wipe the disk and install the new image, complete the following:
    - In the **Force disk** field, specify a comma-separated list of disk devices.
    - Check the **Wipe disk** box. If checked, the cluster manager wipes and then repartitions the disk.
- In the **Transport** field, select a transport method.

  For more information about various transport methods, see the installation guide for your platform. For links to the installation guides, see the following:

  Cluster manager documentation
- Check or clear the **Ignore power errors** box.

  When **Ignore power errors** is clear, the cluster manager stops the provisioning action if power errors occur during any step in the process, such as halting a node or checking node status. If this box is checked, the cluster manager continues the provisioning action if errors occur.
- In the **Wait for halt (seconds)** field, accept the default of 15 seconds or specify a different value.
- Check or clear the **Force PXE** box.

  If checked, the cluster manager PXE boots the node. Check this box if the node has been configured to boot to disk before it PXE boots.

5. Review the content of the **Image Provisionment Confirmation** popup window and click **OK**.
6. Examine the status window and check for any nodes that failed to deploy.

   When provisioning is complete, the final status displays. The compute nodes that provisioned successfully display in the chosen image group. The compute nodes that failed remain in the default image group.

   If the node name has a suffix of `[non-active]` in the image group containing the image, then the node failed to provision. If the node name has a suffix of `[active]`, then the node is provisioned correctly.

Exporting node information to a flat text file

You can export node information to a text file. When you store the text file on a system outside the cluster, you have a backup copy of the cluster node information recorded in a simple report.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Node Management**.
3. Click **Export Compute Nodes**.
4. In the **Export** dialog box, complete the following steps:
    - In the **File Name** field, enter a name for the exported file.
    - Click **Export**.
5. Save the file.

Retrieving node information from the GUI

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Node Management**
3. In the left pane, locate the node for which you want to retrieve information.

   If necessary, click **+** to expand the node group that includes the node.
4. Select a node.
5. In the right pane, click **Details**.

Managing custom groups of nodes and network groups of nodes from the GUI

The cluster manager lets you put nodes into groups and lets you manage nodes as groups. You can create the following types of groups:

- Custom groups. You can create a custom group for any group of nodes. For example, you could create node groups for the following purposes:
  - A group of nodes that is dedicated to a particular job or project.
  - A group of nodes that have GPUs attached to them.
  - A group of nodes that you want to take down for servicing.

  The following topics explain how to add and delete custom groups:

  - **Adding custom groups from the GUI**
  - **Deleting custom groups from the GUI**

- Network groups. A cluster can have multiple network groups. These groups typically include nodes grouped according to the cluster network topology.

  If you create network groups, monitoring is more efficient. The cluster manager monitoring tools can monitor nodes that send data to each other.

  Typically, a network group consists of one of the following:

  - Compute nodes attached to a common switch.

    The cluster manager does not configure these groups by default. If you want to run monitoring, Hewlett Packard Enterprise recommends that you configure network groups for compute nodes.
  - ICE compute nodes attached to a common leader node.

    The cluster manager configures these groups by default. These groups also include the associated chassis controllers and InfiniBand switches. These groups are called **reserved network groups**. Each leader node and the ICE compute nodes attached to that leader node constitute a reserved network group. If you enter the following command, you can see these reserved network groups:

    ```
    # cm group network show
    admin               # A reserved network group
    rack1               # A reserved network group
    ```

```
rack2                # A reserved network group
mynetworkgroup       # A user-defined network group
```

You cannot include nodes from a reserved network group in your own, user-defined network group. You cannot delete a default, reserved network group. The admin node is in its own reserved network group called `admin`.

The following topics explain how to add and delete network groups:

- **Adding network groups from the GUI**
- **Deleting network groups from the GUI**

To see the names of all the node groups defined in the cluster, including the default groups, enter the following command:

```
# ls /etc/dsh/group
```

You can delete the node groups that you create. When you delete a custom group, you can archive the custom group. When you archive a custom group, the cluster manager saves all the node information about the group so you can visualize that data later. Network groups tend to be useful for long time periods, so they cannot be archived.

Cluster manager support for most node operations is identical. Certain operations, such as the following, are performed the same way for both custom groups and network groups:

- Adding groups
- Deleting groups
- Adding nodes to groups
- Deleting nodes from groups

You can use either the cluster manager GUI or the `cm group` command for most node group operations. However, to specify the simultaneous creation of `pdsh` groups, use the `cm group` command.

If you plan to create `pdsh` groups for your node groups, use the `cm group` command to manage the node groups. The command ensures correct additions and deletions from both the node group and the `pdsh` group. Additions to and deletions from node groups from the GUI do not propagate to the `pdsh` groups.

The following additional topics contain information about managing node groups:

- **Archiving custom groups from the GUI**
- **Visualizing history data from the GUI**
- **Limitations for displaying archived custom groups from the GUI**
- **Creating custom groups from the CLI**
- **Creating network groups from the CLI**
- **Creating `pdsh` group files**

Viewing node groups from the GUI

You can view the nodes grouped in the following ways:

- Network groups. Typically, these are nodes under a common switch.
- Custom groups. These groups can consist of any group of nodes with similar characteristics.

- System groups. These are groups of nodes based on roles. For example, groups of compute nodes, groups of ICE compute nodes, switches, and so on.
- Image groups. These are nodes that all host the same system image.

various ways, including system groups, image groups, network groups, and custom groups. All the groups appear in the left pane. The following are some ways to view information about system groups:

- To expand a resource, click the **+** button.
- To hide default system groups that are empty, complete the following steps:
  1. In the left pane, find the **Filter** row, and click the **X** to the right.
  2. Clear **Hide empty system groups**. This action hides empty system groups in the left pane. Empty system groups continue to appear in the right pane.

The GUI includes nodes that are deployed successfully in an active image group. The GUI lists inactive nodes as nonactive candidates.

If either of the following are true, the nodes appear in a default network group that contains the unclassified nodes:

- No network group is defined for the node.

  Or
- The node is not included in a network group.

Adding custom groups from the GUI

In addition to the standard cluster manager node groups, you can create a custom group for a site-specific purpose. Nodes can belong to more than one standard group, and nodes can belong to more than one custom group. Custom groups are optional.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Custom Group Management**
3. In the right pane, in the **Custom Groups** display, click **Create**.

   From the **Custom Groups** display, you can add, delete, or rename a custom group.
4. In the **New Custom Group** popup window, in the **Group Name** field, enter a name for the custom group and click **OK**.

   Choose a group name that conforms to the following rules:

   - The first character must be an alphanumeric character.
   - The remaining characters can be one of the following:
     - Alphanumeric characters
     - Hyphen (–)

- Period (.)
- Underscore character (_)

Notice that the new custom group appears in the left pane under the **Custom Groups** heading.

5. Select nodes from the **Nodes in Cluster** column and click the **=>** button to add them to the **Nodes in Custom Group** column.

   After the group is created, you can add or delete nodes. Select any number of nodes from the **Nodes in Cluster** section and use the arrows to move the nodes to the **Nodes in Custom Group** section.

Deleting custom groups from the GUI

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Custom Group Management**
3. In the right pane, in the **Custom Groups** display, do the following:
   - Use the **Select a Custom Group** pull-down menu to select the group you want to delete.
   - Click **Delete**.
4. On the **Archive Custom Group in History Engine** popup window, select **Yes** to archive the group or select **No** to simply delete the group.
5. (Conditional) In the **Archiving Custom Group** popup window, supply the name you want to assign to the archive version of the custom group and click **OK**.

   Complete this step if you want to archive the custom group.

Adding network groups from the GUI

For the monitoring functions to work, each compute node must be included in a network group. The cluster manager automatically includes ICE compute nodes in network groups based on the rack number. Ensure that compute nodes are included in a network group.

The cluster monitoring function requires each node to be included in a network group.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Network Group Management**
3. In the right pane, click **Create**.
4. In the **New Network Group** window, specify a name for the network group you want to create and click **OK**.
5. Complete the following actions to move nodes into network groups:
   - Select any number of nodes from the **Nodes not in any Network Group** section on the left.
   - Use the arrows to move the nodes to the **Nodes in Network Group** section on the right.

You can include up to 519 nodes in a single network group. By default, the cluster manager checks all network groups when you enable, start, or restart native monitoring. If necessary, the cluster manager reconfigures the network groups as follows:

- If a network group contains 520 nodes or more, the cluster manager splits the group. The cluster manager keeps 519 nodes in the original group with the original group name. It puts the excess nodes into an additional group. The additional group name is the original group name with a suffix of `_1`. For example, if the original network group was named `mynetworkgroup`, the cluster manager puts the additional nodes into another group called `mynetworkgroup_1`.
- If the cluster manager finds unassigned nodes in the `MonitNe` default group, it assigns those nodes to one or more groups named `mongroup`*X*, where *X* is an integer number starting with 1. The cluster manager groups nodes together into a network group if the nodes are under the same leader node.

For optimal, lightweight performance, Hewlett Packard Enterprise recommends that each network group correspond to an Ethernet switch. A switch in the cluster must physically represent a network group, and the associated nodes must be connected to that switch.

Deleting network groups from the GUI

You can delete network groups. After you delete a network group, you can put the nodes from the group into another network group.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Network Group Management**.
3. In the right pane, use the **Select a Network Group** pull-down menu to select a network group that you want to delete.
4. Click **Delete**.

   After you delete a network group:

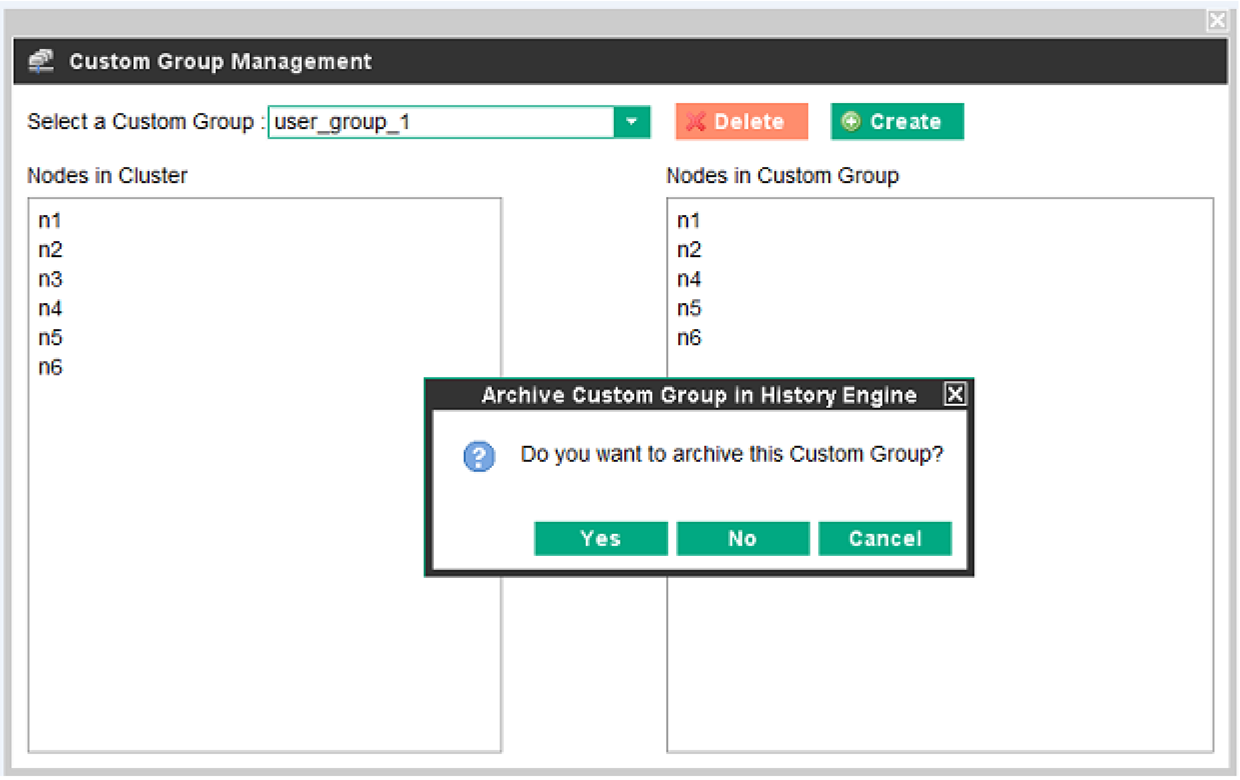
   - You cannot recover that network group.
   - You can put the nodes from the network group into another network group.
   - You can create a new network group with the same name as the network group you deleted.

Archiving custom groups from the GUI

Procedure

1. Access the **Custom Group Management** window.
2. In the **Select a Custom Group** field, use the drop-down list to select the custom group to delete.
3. Click **Delete**.

   A window displays asking if you would like to archive the selected custom group.



Select a Custom Group : user_group_1

Do you want to archive this Custom Group?

Figure 4: Archiving deleted custom groups

4. Click **Yes** to archive the group.

   After the custom group is deleted, it displays in the **Archived Custom Groups** list in the left-frame tree.

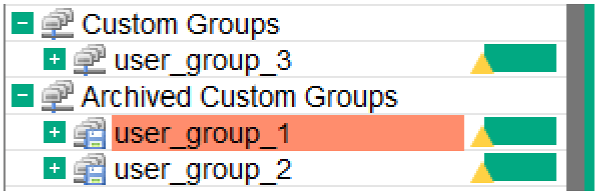


Figure 5: Archived custom groups

NOTE: As an alternative, use the following command to archive a custom group:

```
cm group custom del group_name
```

For more information, see the help output.

Visualizing history data from the GUI

When selecting an archived custom group in the left pane tree view, a static **Time View** picture appears in the right pane. This picture shows the activity view of the custom group during its existence. All options available with **Time View** are also available when visualizing archived custom groups.

Limitations for displaying archived custom groups from the GUI

To display an archived custom group, the following conditions must be satisfied.

- Time must not exceed 24 hours.
- The number of nodes must not exceed 4096.
- The number of metrics must not exceed 100.
- The product of the three parameters above must not exceed 409600.

The table below displays examples of valid combinations of these three parameters.

Table 1: Valid archived custom group parameters

Nodes	Metrics	Hours	Nodes Metrics Hours
4096	10	10	409600
4096	5	20	409600
4096	100	1	409600
256	100	12	307200
2048	8	24	393216
1024	16	24	393216

IMPORTANT: If the preceding criteria are not met, the archived custom group does not display. Instead, a warning message displays.

Uploading files to the admin node

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Right-click the cluster name, and select **Upload file(s) to /tmp**.
3. In the **Upload file(s)** popup, specify file or files you want to upload.
4. Click **Upload file(s)**.

   The cluster manager uploads the selected files to the `/tmp` folder on the admin node, and a progress window displays.

   When the upload is complete, the progress window disappears and the following message displays in the bottom information panel:

   ```
   Uploading: - Operation completed
   ```

Creating a secure shell (SSH) connection to a node from the GUI

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, locate the node.

If necessary, click **+** to expand the node group that includes the node.

You cannot open an SSH connection if you have more than one node selected.

3. Right-click the node and select **SSH Connection**.

Changing the default gateway IP address of a node from the GUI

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Node Management**.
3. In the left pane, select a node.
4. Right-click the node and select **Manage** > **Modify Node**.
5. In the **Modify Node Dialog** box, in the **Default Gateway IP Address** field, enter one of the following values and click **OK**:

   - The keyword `default`.
   - The keyword `cmumgt`.
   - The actual, numeric IP address of the gateway.

Connecting to a node controller from the GUI

Prerequisites
The node has a node controller, and the node controller is configured properly. The node controller can be an iLO device or a baseboard management controller (BMC).

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, locate the node.

   If necessary, click **+** to expand the node group that includes the node.

   You cannot connect to a node controller if you have more than one node selected.
3. Right-click the node and select one of the following:

   - **Management Card Connection**.

     This selection opens a telnet or SSH connection to the node controller.
   - **Virtual Serial Port Connection**

     This selection runs the commands necessary to open a virtual serial port connection to the node controller.

   The GUI does not present the preceding menu selections if the node does not have a node controller.
4. When prompted, provide the node controller credentials, and click **OK**.

Shutting down nodes

Prerequisites

The node or nodes that you want to shut down must use a node controller. To determine if a node includes a node controller, select the node and examine the **Details** tab in the right pane.

If the node does not use a node controller, perform a reboot rather than a shutdown. For nodes without node controllers, a shutdown action can cause a hang, which can necessitate a manual shutdown.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click the nodes, and select **Shutdown (using SSH)**.
4. On the **Shutdown** popup window, do the following:
    - (Optional) In the **Message** field, specify a message for system users.
    - (Optional) Use the **Delay (in mins)** pull-down menu to specify an amount of time to wait before shutting down the node or nodes.
    - Click **OK**.

Powering off nodes

Prerequisites

The node or nodes that you want to power off must use a node controller. To determine if a node includes a node controller, select the node and examine the **Details** tab in the right pane.

The node controllers in the nodes you want to power off must all have the same passwords.

Before you power off a node, perform a node shutdown. For information about how to shut down a node, see the following:

Shutting down nodes

NOTE: Failure to complete the shutdown procedure before you power down a node can result in damage to the file system. Make sure to shut down the node before you power off the node.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click the nodes, and select **Power Off (using BMC)**.
4. On the **Power Off** popup window, click **OK**.

Booting or rebooting nodes

Prerequisites

The node or nodes that you want to boot or reboot must all use a node controller. To determine if a node includes a node controller, select the node and examine the **Details** tab in the right pane. The node controllers in the nodes you want to power off must all have the same passwords.

Before you boot or reboot a node, perform a node shutdown. For information about how to perform a shutdown, see the following:

Shutting down nodes

NOTE:

Perform a shutdown before you boot or reboot. After you click **Boot/Reboot (using BMC)**, the cluster manager attempts to shut down the nodes. If the shutdown fails, the node controller resets. This can damage the file system.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click and select **Boot/Reboot (using BMC)**.
4. In the **Boot** popup window, do the following:
    - Select one of the following boot modes:
        - Default
        - NetBoot OS
    - Click **OK**.

Rebooting nodes

The following procedure reboots a node using `ssh`.

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click and select **Reboot (using SSH)**.
4. On the **Reboot** popup window, do the following:
    - (Optional) In the **Message** field, specify a message for system users.
    - (Optional) Use the **Delay (in mins)** pull-down menu to specify an amount of time to wait before rebooting the node or nodes.
    - Click **OK**.

Changing the status of the unit identification (UID) light emitting diode (LED)

Prerequisites

The node must have an iLO node controller, and the node must be equipped with a status LED. To determine if a node includes an iLO node controller, select the node and examine the **Details** tab in the right pane.

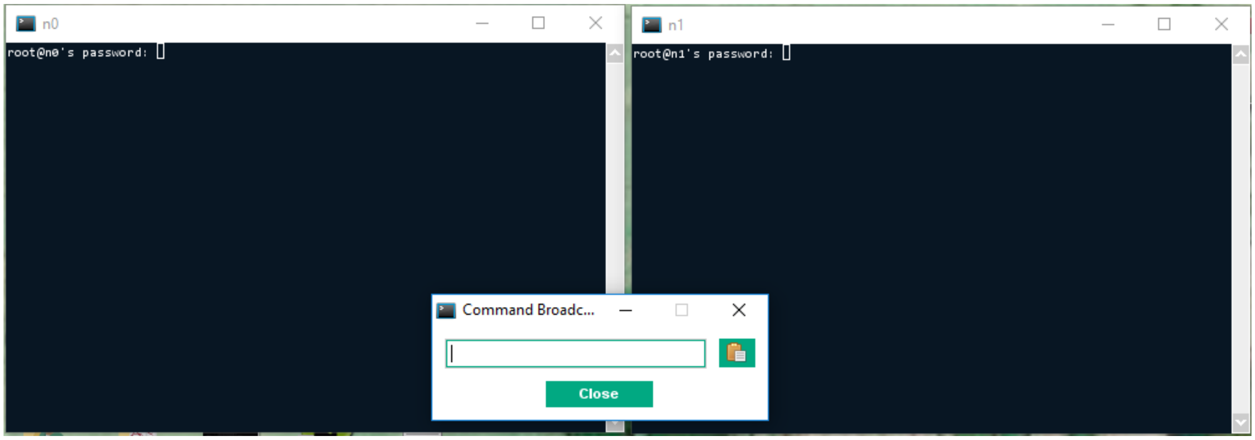
Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click and select **Change UID LED Status**.
4. On the **Change Locator LED** popup window, do the following:
   - Select **ON** or **OFF**.

     If the switch is on, this procedure turns on the LED.
   - Click **OK**.

Running a command in multiple windows on multiple nodes

The procedure in this topic shows you how to configure terminal windows for two or more nodes and run commands on the nodes concurrently. This procedure yields a display similar to the following on your monitor:



```
root@n0's password:
```

```
root@n1's password:
```

Figure 6: Running commands on multiple windows

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select two or more nodes.

   Hewlett Packard Enterprise recommends that you select no more than 64 nodes.
3. Right-click and select **SSH Connection**.
4. On the **Broadcast** popup window, do the following:

- In the **Connection type** field, select one of the following:
    - **SSH**. Enables a secure shell connection to each node through the network when the network is up.
    - **Management Card**. For nodes that have node controllers, this connection type enables a connection through the node controller.
    - **Virtual Serial Port**. For nodes that have node controllers, this connection type enables a connection to the virtual serial port through the node controller.
- In the **Terminal size** field, accept the default of 80 by 24 rows, or specify a different size. Specifying smaller values yields smaller terminal windows. Specifying larger values yields larger terminal windows.
- Choose one of the following window placement schemes:
    - **Let window manager place windows**.
    - **Automatic grid placement**.
    - **Manual grid placement (in pixels)**.

      Also specify the number of pixels you want the cluster manager to maintain between windows. In the **X shifting** (horizontal) and **Y shifting** (vertical) fields, specify a number.

  By default, the cluster manager arranges the windows so that all windows appear on your screen and no windows appear outside of your screen.
- Click **OK**.

5. Run commands in one of the following ways:
    - To run a command on all the nodes, enter the command in the **Command Broadcast** window.
    - To run a command that is on your clipboard, click the clipboard icon on the right in the **Command Broadcast** window.
    - To run a command on only one node, enter the command directly into the window for that node.

Running a command in one window on multiple nodes with PDSH

The cluster manager can facilitate a parallel distributed shell (PDSH) connection to specific nodes so you can run commands on multiple nodes with one `pdsh` command. You can use either `cm_diff` or `dshbak` to format the `pdsh` output in a way that omits repetitive, identical results from multiple nodes.

For examples, see the following:

- **Example 1. PDSH (using HPCM)**
- **Example 2. PDSH (using HPCM)**

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select two or more nodes.

3. Right-click and select **PDSH (using HPCM diff)**.
4. (Optional) Change the display filter when the terminal window appears.

   The following are the possible display mechanisms:

   - `cm_diff` in interactive mode. Default.
   - `cm_diff` in noninteractive mode.
   - `dshbak`.

   To display the full list of `cm_diff` options, enter the following command:

   ```
   # cm_diff -h
   ```

Example 1. PDSH (using HPCM)

The `cmudiff` output consists of two fields. For example:

```
cmu_pdsh> date

qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq
Responses: 4 { node[41-44] }
Reference: node41 - 1 line
Ignored:
[ ]      <none>
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqwqqqqqqqqqqqqqqqqqq
                                                             x
   m+  Mon Aug 27 11:19:47 EDT 20XX                          x (all different)
                                                             x
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqvqqqqqqqqqqqqqqqqqq
exit code   0 :   4 { node[41-44] }
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq
```

The header displays the following information.

- The number of responses

  In the example output, the number of responses is 4. This means that 4 compute nodes responded.

- The reference node

  This is the node chosen by `cmudiff` as a reference. The output highlights differences in the output from this reference node.

- The list of ignored lines

The output displays below the header. In the example output, the output is only 1 line. The `m` on the left indicates that the output from some compute node differs from the reference node. Some details about the output processing results display on the right.

Characters that differ from the reference node appear in a different color. In the example, the time drift in the `seconds` field differs.

When you click the line that starts with `m+`, you can see the individual differences. The following example shows the differences:

```
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq
Responses: 4 { node[41-44] }
Reference: node41 - 1 line
Ignored:
```

```
[ ]      <none>
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqwqqqqqqqqqqqqqqqqqq
                                                             x
   m-  Mon Aug 27 11:19:47 EDT 20XX                          x (all different)
25% x  Mon Aug 27 11:19:47 EDT 20XX                          x x   1: node41
25% x  Mon Aug 27 11:19:37 EDT 20XX                          x x   1: node42
25% x  Mon Aug 27 11:19:51 EDT 20XX                          x x   1: node43
25% x  Mon Aug 27 11:19:11 EDT 20XX                          x x   1: node44
                                                             x
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqvqqqqqqqqqqqqqqqqqq
exit code   0 :   4 { node[41-44] }
qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq
```

Depending on the output length, you can pipe the `cmudiff` output to the `less` editor to enable scrolling through the output with arrows. Enter `q` to terminate output editing.

Example 2. PDSH (using HPCM)

This example uses `cm_diff` to detect BIOS firmware version differences. After you start the PDSH session, enter the following command:

```
cmu_pdsh> dmidecode
```

The cluster manager returns several lines of information, incuding the following:

```
responses: 5, no data: 0
reference: o185i032
ignored: <none>
output: 824 lines
[[ use directional arrows to navigate, press'q' to return ]]
----------------------------------------------------------------------
     | # dmidecode 2.9                                                |
     | SMBIOS 2.6 present.                                            |
    m| 62 structures occupying 3513 bytes.                            | (3 populations, not displayed)
    m| Table at 0x000FCBD0.                                           | (2 populations, not displayed)
     |                                                                |
     | Handle 0x0000, DMI type 0, 24 bytes                            |
     | BIOS Information                                               |
     |       Vendor: HPE                                              |
     |       Version: U32                                             |
    m|       Release Date: 11/13/20XX                                 | (2 populations, not displayed)
     |       Address: 0xE0000                                         |
     |       Runtime Size: 64 KB                                      |
     |       ROM Size: 64 MB                                          |
     |       Characteristics:                                         |
     |             ISA is supported                                   |
     |             PCI is supported                                   |
```

The window displays a small portion of the output. Use the arrows to scroll up and down.

A difference is found in the BIOS release date. The following comment, which appears on the right, suggests that two groups of nodes are present with two different BIOS release dates:

```
2 populations, not displayed
```

The output indicates that one of the two populations might be a single node without a firmware upgrade.

Use the `-d` option to narrow the search to the failing nodes and display node populations.

```
cmu_pdsh> cm_diff -d
cm_diff filter is <ON>, with parameters -d
cmu_pdsh>
cmu_pdsh> dmidecode
```

```
responses: 5, no data: 0
reference: o185i032
ignored: <none>
output: 824 lines
[[ use directional arrows to navigate, press'q' to return ]]
```

```
---------------------------------------------------------------------
     | # dmidecode 2.9                                                | 
     | SMBIOS 2.6 present.                                            | 
    m| 62 structures occupying 3513 bytes.                            | (3 populations) o185i[039,043] are 94% similar, 0185i[04
    m| Table at 0x000FCBD0.                                           | (2 populations) o185i[040,042] are 94% similar
     |                                                                | 
     | Handle 0x0000, DMI type 0, 24 bytes                            | 
     | BIOS Information                                               | 
     |       Vendor: HPE                                              | 
     |       Version: U32                                             | 
    m|       Release Date: 11/13/20XX                                 | (2 populations) o185i[040,042] are 83% similar
     |       Address: 0xE0000                                         | 
     |       Runtime Size: 64 KB                                      | 
     |       ROM Size: 64 MB                                          | 
     |       Characteristics:                                         | 
     |             ISA is supported                                   | 
     |             PCI is supported                                   | 
```

The comment now says `(2 populations) o185i[040,042] are 83% similar`. This comment suggest that those two compute nodes have a different BIOS release date than all other nodes.

NOTE: An unresponsive node causes the answer from other nodes to be delayed until a timeout occurs from the unresponsive node. You can reduce this delay by setting the `ConnectTimeout` value in the in `.ssh/config variable`.

For example:

```
# vi /root/.ssh/config
Host *
StrictHostKeyChecking no
ConnectTimeout 1
```

To review detailed BIOS settings on multiple nodes, use the following command:

```
cm node bios show --cmdiff -n node,node
```

For *node*,*node*, specify two or more node hostnames.

Copying a file from the admin node to another node or nodes

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. In the left pane, select one or more nodes.
3. Right-click and select **PDCP (Distributed Copy)**.
4. On the **Parallel distributed copy** popup window, do the following:

    - In the **Source:** field, specify the full path to the source file on the admin node. Alternatively, click the folder icon at the right and browse to the source location.
    - In the **Destination:** field, specify the full path to the target destination on the node. Alternatively, click the folder icon at the right and browse to the target location.
    - Check or clear the following boxes:
        - **recursive**. Check this box if you want this action to include all underlying directories and files.
        - **preserve permissions**. Check this box if you want to retain the file permissions after the copy operation.
    - Click **OK**.

Adding a network

Procedure

1. Click **Options** > **Enter Admin Mode** and enter the cluster credentials.
2. Use one of the following methods to display the **Add Network** dialog box:

   Method 1:

   - Click **Cluster Administration** > **Network Management**.
   - In the right pane, click **Add Network**.

   Method 2:

   - In the left pane, expand **Networks**.
   - Select one of the networks.
   - Right-click **Manage** > **Add Network**.

3. Complete the fields in the **Add Network** dialog box and click **OK**.

   The fields in the dialog box are as follows:

   - **Network Name**

     Enter a name for the network.
   - **Base IP**

     Enter a valid IP address.
   - **Broadcast IP**

     Enter a valid IP address.
   - **Gateway IP**

     Enter a valid IP address.
   - **Type**

     Enter one of the following keywords:

     - `cooling`
     - `data`
     - `ha`
     - `ib`
     - `lead-bmc`
     - `lead-mgmt`
     - `mgmt`
     - `mgmt-bmc`

- **Vlan**

  Enter the number that corresponds to the VLAN you want to add.

- **NetMask**

  Enter a valid IP address.

- **Management Server IP**

  Enter a valid IP address.

- **Rack**

  Enter a valid IP address.

  This field is optional.

Deleting a network

Procedure

1. Click **Options** > **Enter Admin Mode** and enter the cluster credentials.
2. Use one of the following methods to select the network or networks that you want to delete:

   Method 1:

   - Click **Cluster Administration** > **Network Management**.
   - In the right pane, select one or more networks to delete.
   - Click **Delete Network**.

   Method 2:

   - In the left pane, expand **Networks**.
   - Select one or more networks to delete.
   - Right-click **Manage** > **Delete Network(s)**.

3. On the **Delete Network(s)** popup, click **OK** to delete the network.

Modifying a network

Procedure

1. Click **Options** > **Enter Admin Mode** and enter the cluster credentials.
2. Use one of the following methods to display the **Modify Network** dialog box:

   Method 1:

- Click **Cluster Administration** > **Network Management**.
- In the right pane, select the network that you want to modify.
- Click **Modify Network**.

Method 2:

- In the left pane, expand **Networks**.
- Select the network that you want to modify.
- Right-click **Manage** > **Modify Network**.

3. Update the fields in the **Modify Network** dialog box as needed and click **OK**.
4. (Conditional) Add the nodes on the network back into the cluster.

   Complete this step if nodes had resided on the network you modified.

   Use the `cm node nic set` command to add the nodes back into the cluster. For syntax, enter the following command:

   ```
   # cm node nic set -h
   ```

Saving user settings for the GUI

The cluster manager saves user preferences to the following file on your client:

`cmu_gui_local_settings`

You can restore preferences from this file.

Customizing the cluster manager GUI

You can add site-specific menu options to the cluster manager GUI.

Procedure

1. Open the following file:

   `/opt/clmgr/etc/cmu_custom_menu`

2. Add custom menu options for your site to the `cmu_custom_menu` file.

   The `cmu_custom_menu` file contains comments that explain how to edit the file. The file includes instructions that explain how to add commands and GUI options. The file provides commented, ready-to-use examples.

   As explained in the file, you can use the custom menu keyword `CMU_SUDO` in the `/opt/clmgr/etc/cmu_custom_menu` file to apply `sudo` support to a command.

3. Save and close the `cmu_custom_menu` file.
4. Use the `cmu_custom_run` command, and review the output, to confirm the new command that you created.

   Example 1. The following command displays the help text for the `cmu_custom_run` command:

   ```
   # cmu_custom_run
   ```

Example 2. The following command displays the list of custom commands. Confirm that the list includes the new command you created in this procedure.

```
# ./cmu_custom_run -l
Title                                                             Command
-----------------------------------------------------------------|-------
Clear /tmp                                                        env
WCOLL=CMU_TEMP_NODE_FILE /opt/clmgr/bin/pdsh -S 'rm -rf /tmp/*'
Uptime Martha /tmp                                                env
WCOLL=CMU_TEMP_NODE_FILE /opt/clmgr/bin/pdsh -S 'uptime'
HPE iLO4+ Agentless Management Service|Get/Refresh SNMP Data      /opt/
clmgr/bin/cmu_get_ams_data -f CMU_TEMP_NODE_FILE
HPE iLO4+ Agentless Management Service|Configure ILO              /opt/
clmgr/bin/cmu_config_ams -f CMU_TEMP_NODE_FILE
HPE iLO4+ Agentless Management Service|Test ILO Config            /opt/
clmgr/bin/cmu_config_ams -t -f CMU_TEMP_NODE_FILE
```

5. On the admin node, create a text file that lists the names of the nodes upon which you want the new command to run.

6. Enter the `cmu_custom_run` command, in the following format, to run the new command:

   ```
   cmu_custom_run -t 'command_title' -f file
   ```

   The variables are as follows:

Variable	Specification
command_title	The name for the new command.
file	The full path and name of the file that lists the nodes upon which you want this command to run.

   For example, the following command runs on the nodes in `/tmp/nodelist`:

   ```
   # cmu_custom_run -t 'HPE iLO4+ Agentless Management Service|Test ILO Config' -f /tmp/nodelist
   executing "/opt/clmgr/bin/cmu_config_ams -t -f /tmp/nodelist"...
   ----------------
   100.117.20.168
   ----------------
   AMS is configured
   ```

7. (Conditional) Edit the `/etc/sudoers` file.

   Complete this step if the customized command requires root permission and you want to control access by nonroot users.

Using the CLI to administer the cluster

Creating custom groups from the CLI

The `cm group custom` command creates a custom group that includes the nodes you specify on the command line. When you create a custom group, you can run commands simultaneously on all or some of the nodes in the group.

NOTE: As an alternative to `cm group custom` command, you can also create a custom node group in one of the following ways:

- At installation time in the cluster definition file. Use the `custom_groups` configuration attribute in the cluster definition file.

  For more information, see the installation guide for your platform. For links to the installation guides, see the following:

  Cluster manager documentation

- After cluster deployment in the GUI. Like the `cm group` command, the GUI enables you to create, delete, or archive custom node groups.

  For more information, see the following:

  Managing custom groups of nodes and network groups of nodes from the GUI

The `cm group` command has several parameters. These parameters allow you to create, delete, and manipulate custom groups. For a full parameter list, enter the following command:

```
# cm group -h
```

Procedure

1. Log into the admin node as the root user.
2. Enter the `cm group custom` command in the following format to create a custom group:

   ```
   cm group custom add -c group_name -n node,node,...
   ```

   The variables are as follows:

Variable	Specification
group_name	A group name that conforms to the following rules: • The first character must be an alphanumeric character. • The remaining characters can be one of the following: ◦ Alphanumeric characters ◦ Underscore character (`_`) ◦ A period (`.`) ◦ Hyphen (`-`)
node,node	Two or more node hostnames.

Example: Creating custom groups for a 100-node cluster

Assume you have a 100-node cluster. The following nodes have specific features:

- Nodes 1 through 20 have graphics processing units (GPUs)
- Nodes 50 through 70 have additional memory.

There are two methods for creating custom groups.

Method 1 creates the node groups at installation time in the cluster definition file. For example, the following lines from a cluster definition file include all nodes in a custom group called `comp`, all nodes with GPUs in a custom group called `gpu-nodes`, and all nodes with additional memory in a custom group called `big-mem-nodes`:

```
[templates]
# compute node templates
name=compute, mgmt_net_name=head, mgmt_bmc_net_name=head-bmc,
mgmt_net_interfaces="eno1",
mgmt_net_bonding_master=bond0, mgmt_net_bonding_mode=active-backup,
redundant_mgmt_network=no, switch_mgmt_network=yes, transport=udpcast,
tpm_boot=no,
dhcp_bootfile=grub2, disk_bootloader=no, predictable_net_names=yes,
console_device=ttyS0,
conserver_ondemand=no, conserver_logging=yes, rootfs=disk, card_type=IPMI,
baud_rate=115200, bmc_username=admin, bmc_password=admin

[discover]
hostname1=n1, internal_name=service1, mgmt_bmc_net_macs="20:67:7c:e4:d0:ee",
mgmt_net_macs="20:67:7c:d7:30:38", template_name=compute,
custom_groups="comp,gpu-nodes"
...
hostname1=n20, internal_name=service20,
mgmt_bmc_net_macs="f4:03:43:49:ca:c8", mgmt_net_macs="f4:03:43:49:ca:d8",
template_name=compute, custom_groups="comp,gpu-nodes"
hostname1=n21, internal_name=service21,
mgmt_bmc_net_macs="ec:eb:b8:94:38:8a", mgmt_net_macs="40:67:7c:d7:30:38",
template_name=compute, custom_groups="comp"
```

```
...
hostname1=n49, internal_name=service49,
mgmt_bmc_net_macs="98:f2:b3:21:23:f4", mgmt_net_macs="98:f2:b3:21:23:f4",
template_name=compute, custom_groups="comp"
hostname1=n50, internal_name=service50,
mgmt_bmc_net_macs="f4:03:43:49:1b:24", mgmt_net_macs="f4:03:43:49:1b:14",
template_name=compute, custom_groups="comp,big-mem-nodes"
...
hostname1=n70, internal_name=service70,
mgmt_bmc_net_macs="d0:67:26:d7:5a:b8", mgmt_net_macs="d0:67:26:d7:5a:ba",
template_name=compute, custom_groups="comp,big-mem-nodes"
hostname1=n71, internal_name=service71,
mgmt_bmc_net_macs="22:67:7c:db:fd:8e", mgmt_net_macs="22:67:7c:d6:2d:5c",
template_name=compute, custom_groups="comp"
...
hostname1=n100, internal_name=service100,
mgmt_bmc_net_macs="f4:03:43:49:2b:b8", mgmt_net_macs="f4:03:43:49:2b:b9",
template_name=compute, custom_groups="comp"
```

Method 2 creates the node groups by entering the following commands after the cluster is installed:

- `# cm group custom add -c comp -n node[1-100]`
- `# cm group custom add -c gpu-nodes -n node[1-20]`
- `# cm group custom add -c big-mem-nodes -n node[50-70]`

The following commands show how to refer to the custom groups:

- The following command shows all the nodes in the custom group `gpu-nodes`:

  ```
  # cm node show -n @gpu-nodes
  ```

- The following command shows all the nodes in the custom group `gpu-nodes` and the custom group `big-mem-nodes`:

  ```
  # cm node show -n @gpu-nodes,@big-mem-nodes
  ```

- The following command shows image information for all the `gpu-nodes`:

  ```
  # cm node show -I -n @gpu-nodes
  NODE    IMAGE.NAME  KERNEL                 IMAGEPENDING CLONINGDATE                  CLONINGBLOCKDEVICE IMAGETRANSPORT
  node43  rhel8.3     4.18.0-240.el8.x86_64  True         202X-04-27T01:32:18.653+0000 default            udpcast
  node44  rhel8.3     4.18.0-240.el8.x86_64  True         202X-04-27T01:32:18.653+0000 default            udpcast
  ```

- The following command shows all compute nodes in the cluster that do not have graphics processing units (GPUs):

  ```
  # cm node show -Z -n @comp --exclude  @gpu-nodes
  node[21-100]     # the -Z in the command line generates compressed output
  ```

- The following command shows all the compute nodes in the cluster that do not have GPUs or big memory:

  ```
  # cm node show -Z -n @comp --exclude @gpu-nodes,@big-mem-nodes
  node[21-49,71-100]
  ```

- The custom group named `gpu-nodes` contains 20 nodes. These nodes are named `node1` through `node20`. The following command uses the `-Z` parameter, which compresses output. The example that follows shows all the nodes in the custom group `gpu-nodes` that have node names that start with `node1`:

  ```
  # cm node show -Z -n @gpu-nodes:'node1*'
  node[1,11-19]
  ```

Creating network groups from the CLI

The `cm group network` command creates a custom network group that includes the nodes you specify on the command line. When you create a custom network group, you can run commands simultaneously on all or some of the nodes in the network group.

NOTE: As an alternative to `cm group network` command, you can create a network group in one of the following ways:

- At installation time in the cluster definition file.

  For more information, see the installation guide for your platform. For links to the installation guides, see the following:

  Cluster manager documentation

- After cluster deployment in the GUI. Like the `cm group` command, the GUI enables you to create, delete, or archive custom node groups.

  For more information, see the following:

  Managing custom groups of nodes and network groups of nodes from the GUI

The `cm group` command has several parameters. These parameters allow you to create, delete, and manipulate network groups. For a full parameter list, enter the following command:

```
# cm group -h
```

Procedure

1. Log into the admin node as the root user.
2. Enter the `cm group network` command in the following format to create a network group:

   ```
   cm group network add -c group_name -n node,node,...
   ```

   The variables are as follows:

Variable	Specification
group_name	A group name that conforms to the following rules: - The first character must be an alphanumeric character. - The remaining characters can be one of the following: ◦ Alphanumeric characters ◦ Underscore character (_) ◦ Period (.) ◦ Hyphen (-)
node,node	Two or more node hostnames.

The following are example commands for network groups:

- The following command adds node `n3` to network group `big-mem-nodes`:

  ```
  # cm group network add -n n3 big-mem-nodes
  ```

- The following command deletes a network group called `big-mem-nodes`:

  ```
  # cm group network del big-mem-nodes
  ```

Changing global cluster configuration settings

This topic explains how to enable optional features on a cluster-wide basis. The features you can enable depend on your hardware platform and your site requirements. When you use the cluster configuration tool, you use the menus to set systemwide, global values. The values you set apply to all nodes that you configure into the cluster after you set the value, and the effects are as follows:

- During a bare-metal installation, the cluster configuration tool sets global values before you configure any nodes into the cluster. All the nodes you configure receive the global values you set in the cluster configuration tool.
- You can add nodes or change global values on a production system. In these cases, you might need to use commands to reset values on older nodes that you had configured previously.

The `configure-cluster` and several `cm` commands work in concert with the cluster definition file. This file defines the following:

- The roles of the various cluster nodes
- Global system attributes
- Data networks and their respective switches
- Management networks and their respective switches

The cluster definition file is a configuration file. The file provides a convenient and efficient method of specifying large-scale changes.

For an overview and examples of the cluster definition file, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

To preserve custom configuration changes across `cm node update config` calls, see the following:

Preserving custom configuration changes

The following topics explain how to use `configure-cluster` or related commands to change global cluster configuration settings:

- **Changing the site domain name service (DNS) server information**
- **Enabling or disabling a backup domain name service (DNS) server**
- **Configuring the `blademond` rescan interval**

Changing the site domain name service (DNS) server information

The following procedure explains how to change or update your site DNS server information in the cluster configuration database.

Procedure

1. From a graphics screen, or through an `ssh` connection, log into the admin node as the root user.
2. Enter the following command to start the cluster configuration tool:

   ```
   # /opt/sgi/sbin/configure-cluster
   ```

3. On the cluster configuration tool main menu, select **D Configure House DNS Resolvers**, and select **OK**.
4. On the **Enter up to three DNS resolvers IPs** screen, enter the IP addresses you want to configure.
5. Select **OK**.

Enabling or disabling a backup domain name service (DNS) server

Typically, the DNS on the admin node provides name services for the cluster. When you configure a backup DNS, however, the compute nodes can use a different, dedicated compute node as a secondary DNS server if the admin node is not available. This feature is optional.

The following examples show how to use commands to enable or disable a backup DNS.

- Example 1. To retrieve current DNS backup information, enter the following:

  ```
  # /opt/sgi/sbin/backup-dns-setup --show-backup
  n0
  ```

- Example 2. To enable a backup DNS on `n0`, enter the following:

  ```
  # /opt/sgi/sbin/backup-dns-setup --set-backup n0
   Shutting down name server BIND waiting for named to shut down (29s)
  done
  Configuration manager initiating node configuration.
  .
  ```

```
.
.
```

- Example 3. To disable the backup DNS, enter the following:

  ```
  # /opt/sgi/sbin/backup-dns-setup --delete-backup
  Shutting down name server BIND waiting for named to shut down (28s) done
  Configuration manager initiating node configuration.
  .
  .
  .
  ```

To use the cluster configuration tool to enable or disable the backup DNS, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Adding, changing, or removing options from the domain name service (DNS) `named` service

The following procedure explains how to change the DNS `named` service settings on the admin node.

Procedure

1. Log into the admin node as the root user.
2. Open the following file in a text editor:

   `/opt/clmgr/etc/named-options.conf`
3. Edit the `named-options.conf` file in a way that adds, changes, or removes options from the `named` service.
4. Save and close the file.
5. Enter the following command to update the configuration, including the `/etc/named.conf` file:

   ```
   # cm node update config --sync dns --node admin
   ```
6. (Conditional) Update the configuration on the scalable unit (SU) leader nodes.

   Complete this step if the cluster has SU leader nodes.

   Enter the following command:

   ```
   # cm node update config --sync dns -t role su-leader
   ```

Enabling extension mechanisms for domain name service (EDNS)

EDNS can cause excessive logging activity when not working properly. The cluster manager limits EDNS logging. This section describes how to delete this code and allow EDNS to work unrestricted and log messages.

To enable EDNS on your cluster, perform the following steps:

Procedure

1. Open the `/opt/clmgr/lib/Tempo/Named.pm` file with your favorite editing tool.
2. (Optional) Remove the limit on the `edns_udp_size` parameter.

Comment out or remove the following line:

```
$limit_edns_udp_size = "edns-udp-size 512;";"
```

3. Remove the following lines so that EDNS logging is no longer disabled:

```
logging {
category lame-servers {null; };
category edns-disabled { null; };  };
```

4. Save and close the file.

Configuring the `blademond` rescan interval

When enabled, the system checks every two minutes for changes to the number of ICE compute nodes in the system. If you remove or add an ICE compute node, the system automatically does the following:

- Detects the change
- Updates the system
- Integrates the change on the rack

By default, the interval between checks is set to `120`, which is two minutes.

Use the following procedure to configure the `blademond` rescan interval from the cluster configuration tool.

Procedure

1. From a graphics screen, or through an `ssh` connection, log into the admin node as the root user.
2. Enter the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

3. On the **Main Menu** screen, select **C Configure blademond rescan interval (optional)**, and select **OK**.
4. On the pop-up window that appears, accept the default of `120`, which is two minutes, and select **OK**.

   Alternatively, enter a different value and select **OK**.

Changing the interfaces used for the house network or the management networks

Procedure

1. Enter the following command to start the cluster configuration tool:

```
# configure-cluster
```

2. On the cluster configuration tool main menu, select **U - Update Admin Interfaces**.
3. On the **House Network Interface Selection** screen, do the following:

   - Use the spacebar and arrow keys to select the network interface card (NIC) you want to use for the cluster house network.

     Verify that the NIC you select has the IP address that you want people to use when they log into the cluster admin node from an outside public network.
   - Click **OK**.

4. On the **Management Network Interfaces Selection** screen, do the following:

   - Use the spacebar and arrow keys to select one or two NICs for the management network.
   - Click **OK**.

5. On the screen that asks **Do you want to use a separate, dedicated NIC to handle BMC traffic on the Management Network?**, click **Yes** or **No**.

   If you clicked **No**, proceed to the next step in this procedure. When you click **No**, the cluster manager uses the NICs you selected in the previous step for node controller traffic.

   If you clicked **Yes**, the installer presents you with the **Management BMC Network Interfaces Selection** screen. Select one of the NICs on that screen for the separate node controller network, and click **OK**.

6. On the screen that asks **Choose Admin bonding mode used for the management network**, do the following:

   a. Click **active-backup** or **802.3ad (LACP)**.

   These modes are as follows:

Mode	Effect
active-backup	Only one link in a bonded interface is active at a time. This mode requires no matching configuration on the management switch. Default.
802.3ad (LACP)	All links in a bonded interface are active at the same time. This mode requires that the Ethernet switch connected has matching LACP configuration on all links in the bonded interface. Hewlett Packard Enterprise recommends using this bonding mode when more than one interface connects to a management network on the admin node.

   b. On the **Main Menu** screen, click **OK**.

7. Examine the contents of the following file to view the interfaces you specified:

```
/etc/opt/sgi/configure-cluster-ethernets
```

Managing slots

The following topics explain how to manage slots:

- **Displaying the slot configuration**
- **Booting from a different slot on clusters without scalable unit (SU) leader nodes**
- **Booting from a different slot on clusters with scalable unit (SU) leader nodes**
- **Copying a slot**
- **Customizing slot labels**
- **Modifying boot options**

Displaying the slot configuration

This topic explains how to complete the following tasks:

- Identify the slot that is in use
- Display information about slots that are configured but not in use

Procedure

1. Log in to the admin node as the root user.
2. Verify the current boot slot:

   ```
   # cadmin --show-current-root
   admin node currently booted on slot: 1
   ```

3. Show information about the slots available to be booted:

   ```
   # cadmin --show-root-labels
   CD slot 1: CM 1.0 / sles15spX: Production
      slot 2: CM 1.0 / sles15spX: Backup for slot 1
   CD slot 3: CM 1.0 / sles15spX: Chris's slot
      slot 4: CM 1.0 / rhel8.X: Do not destroy until June 30 20XX
      slot 5: CM 1.0 / rhel8.X: (none)
   ```

Booting from a different slot on clusters without scalable unit (SU) leader nodes

If you configured more than one slot, you can boot from the boot partition of any slot. The following procedure explains how to change the system to boot from a different slot.

Procedure

1. Log in as the root user to the admin node.
2. Change the default slot.

   You can specify the new slot now, or you can specify the new slot during the reboot. This step explains how to change the boot slot now. Enter the `cadmin` command in the following format:

   ```
   cadmin --set-default-root --slot num
   ```

   For *num*, specify the new boot slot number. *num* can be an integer from 1 to 10, inclusive.

   For example, to specify a boot from slot 2, enter the following:

   ```
   admin:~ # cadmin --set-default-root --slot 2
   ```

   For information about the operating systems installed in each slot, see the following:

   Displaying the slot configuration

3. Enter the following command to shut down the entire system:

   ```
   # cm power shutdown -t system
   ```

4. Enter the following command to reboot the admin node:

   ```
   # reboot
   ```

5. Connect to the system console to monitor the reboot.

   Optionally, select a nondefault slot from which you want to boot.

During the reboot, the system displays a screen that shows all the available slots and highlights the current boot slot. To select a different boot slot, use the arrow keys to select a new slot and press **Enter**.

If you do not select a new slot, the system boots from the highlighted slot after approximately 10 seconds.

6. Log in as the root user again.

7. Enter the following command to reboot all the rack leaders and compute nodes:

   ```
   # cm power on -t system
   ```

   If the IP addresses are configured differently within different slots, the `cm power` command might not be able to communicate with the node controllers immediately after you reboot the admin node. If you have trouble connecting to the rack leaders and compute node node controllers after you change slots, wait a few minutes. Issue the `cm power` command again. The wait enables the nodes to obtain new IP addresses.

Booting from a different slot on clusters with scalable unit (SU) leader nodes

Procedure

1. Power off all the compute nodes in the cabinets:

   ```
   # cm power off -t node r1c*t*n*
   ```

2. Check the status, and ensure that the nodes are powered off:

   ```
   # cm power status -t node r1c*t*n*
   ```

3. Set the admin node and scalable unit (SU) leader nodes to the slot you want to boot.

   For example, assume that you want to boot slot 3. Enter the following commands:

   ```
   # cadmin --set-default-root --slot 3
   # su-leader-setup --set-default-slot 3
   ```

4. Reboot the SU leader nodes:

   ```
   # cm power reboot -t leader leader_hostnames
   ```

   For *leader_hostnames*, specify the hostnames of the SU leader nodes. Use a comma to separate multiple node hostnames. You can use wild card characters in the specification.

5. Console into one of the SU leader nodes and monitor the nodes as they come back online:

   ```
   # console leader1
   ```

   Provide the username and password credentials as prompted.

6. Verify that the SU leader nodes are booted with specified slot:

   ```
   # df -h | grep "/$"
   # ctdb status
   ```

7. Reboot the admin node:

   ```
   # reboot
   ```

8. After the admin node reboots, make sure that shared storage is mounted on the admin node:

   ```
   # df -h
   # enable-su-leader --skip-sync
   ```

9. Power-on all the compute nodes:

   ```
   # cm power on -t node r*c*t*n*
   ```

Monitor the console. If the compute nodes fail to boot and fail to mount the shared storage, enter the following command:

```
# clush -g su-leader systemctl restart glusterd
```

Copying a slot

You can copy, or clone, the installation in one slot to a different slot at any time. Before you modify slot images or reconfigure a slot, copy it first. Hewlett Packard Enterprise recommends copying because the copy provides a backup. If you want to revert to the original configuration, you can use the copy.

The `cm node slot copy` command does the following:

- Synchronizes the data.
- Configures the `grub` and `fstab` entries to make the copied slot bootable.

Procedure

1. Log into the admin node as the root user.
2. Enter the `cm node slot copy` command in the following format:

   ```
   cm node slot copy src_nodes -s src_slot -d dest_slot [--confirm]
   ```

   The variables are as follows:

Variable	Specification

src_nodes	The nodes you want to copy. Specify these nodes in one of the following ways:
	• `-n` *`node,node,...`*
	Specify one or more node hostnames. If specifying individual hostnames, specify them in a comma-separated list. You can use wildcard characters.
	For information about how to specify hostnames, see the following:
	Using the `cm` commands
	For example:
	`@gpu-nodes`
	`n0`
	`node?`
	`node[13]`
	`node[10-14]`
	`node[001-022]`
	`node[2-6,20-26,36]`
	`node52*`
	`admin`
	• `-f` *`file`*
	For *file*, specify the full path to a file that contains a list of node hostnames.
	• `-t {custom,image,network,system}` *`group_names`*
	Copies only the nodes in the `custom`, `image`, `network`, or `system` group. For *group_names*, specify one group name or a comma-separated list of group names.
	Common system group names include `compute`, `ice_compute`, and `leader`.
	If you specify `-t system ALL` , the command runs on all cluster nodes including the admin node.

src_slot	The slot number that contains the configuration that you want to copy. If the *src_slot* slot is the mounted, or active, slot, the command shuts down the cluster database on the admin node before it starts the copy operation. The command ensures that the cluster database does not change during the copy operation and that there is no data loss.
dest_slot	The slot number to receive the copy of the configuration. This slot is the destination slot. **NOTE:** The copy process completely destroys all data in *dest_slot*.
`--confirm`	Optional. When specified, the command displays the wildcard-expanded node names and then prompts to continue.

For more information, including additional examples, enter one of the following commands:

- # **man cm-node-slot-copy**

  Or

- # **cm node slot copy -h**

For example, assume that you want to create a command that completes the following tasks:

- Copies the software for the admin node, the leader nodes, and the compute nodes from slot 1 to slot 2
- Overwrites the contents of slot 2.

The command is as follows:

```
# cm node slot copy -n '*' -s 1 -d 2
```

The example command does not copy ineligible nodes. Ineligible nodes include all scalable unit (SU) leader nodes and all nodes without disk root file systems. The command removes these nodes from the copy.

If the command is run on a cluster with ICE compute nodes, be aware that the ICE compute nodes do not participate in the copy process because they are diskless.

3. Enter the following commands on the source slot to restart the cluster manager:

```
# systemctl start cmdb.service
# systemctl start clmgr-power.service
# systemctl start config_manager.service
```

Customizing slot labels

After an installation, the slot label is `(none)`. You can use the `cadmin` command to label the slots on a multiple-boot cluster.

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to retrieve the current labels:

```
admin:~ # cadmin --show-root-labels
   slot 1: HPCM 1.X / sles15spX: (none)
   slot 2: HPCM 1.X / sles15spX: Backup for slot 1
   slot 3: HPCM 1.X / sles15spX: (none)
   slot 4: HPCM 1.X / rhel8.X: (none)
   slot 5: HPCM 1.X / rhel8.X: patch 11395
```

3. Enter the following command to specify the slot and the label:

```
cadmin --set-root-label --slot num --label "mylabel"
```

The variables are as follows:

Variable	Specification
num	Use an integer from 1 to 10, inclusive, to specify the slot you want to label.
mylabel	Enter the name you want to apply to the slot.

For example:

```
# cadmin --set-root-label --slot 1 --label "Installed 08/15/20XX"
# cadmin --show-root-labels
   slot 1: HPCM 1.X / sles15spX: Installed 08/15/20XX
   slot 2: HPCM 1.X / sles15spX: Backup for slot 1
   slot 3: HPCM 1.X / sles15spX: (none)
   slot 4: HPCM 1.X / rhel8.X: (none)
   slot 5: HPCM 1.X / rhel8.X: patch 11395
```

Modifying boot options

You can use the `cm image` command to set extra kernel boot parameters for the following on a per-image basis:

- ICE compute nodes
- Compute nodes
- Leader nodes

For example, assume that you want to add the `cgroup_disable=memory` kernel boot parameter to the `ice-sles15spX` image, which is installed on all compute nodes.

Enter the following command:

```
% cm image set -i ice-sles15spX --kernel-extra-params cgroup_disable=memory
```

To change the boot parameters, issue additional `cm image` commands.

To display kernel parameters, enter the following command:

```
# cm image show --settings [-i image]
```

The following `cm image` commands might be useful to you when you update boot parameters:

- `cm image unset -i image --kernel-extra-params`
- `cm image set -i image --nfsroot-extra-params`

  For example:

  ```
  # cm image set -i ice-sles15spX --nfsroot-extra-params
  ```

For information about boot options, see the installation guide for your platform and your operating system documentation. For links to the installation guides, see the following:

Cluster manager documentation

Rebooting, halting, powering on, and powering off

The `cm power` command controls power on actions, power off actions, and other actions. The command can also display the power status of system components. The following command format shows the basic parameters:

```
cm power action -t target_type [target_hostname]
```

The following table summarizes the tasks you can accomplish with the `cm power` command:

- You can accomplish several tasks from the admin node. These tasks are as follows:

Action	Target
Power on, power off, return status. The *action* keywords are as follows: ◦ `on` ◦ `off` ◦ `status`	A computing component with one of the following *target_type* keywords: ◦ `system`. Do not specify a *target_hostname* when the target is `system`. ◦ `node` This term refers to scalable unit (SU) leader nodes, compute nodes, and ICE compute nodes. ◦ `leader`. This term refers to HPE SGI 8600 ICE leader nodes. ◦ `rack` ◦ `chassis` ◦ `switch` ◦ `tray` Valid on HPE Cray EX systems only. ◦ `slot` Valid on HPE Cray EX systems only. ◦ `controller` Valid on HPE Cray EX switches or node controllers only.
Power cycle, power reset, chassis power soft. The *action* keywords are as follows: ◦ `hard_reboot` ◦ `reset` ◦ `press`	The node controller of the *target_type* computing component.
Turn on an identification light, turn off an identification light. The *action* keywords are as follows: ◦ `uid_on` ◦ `uid_off`	The node controller of the *target_type* computing component.

- You can accomplish additional tasks with a two-step process. First, use the `ssh` command to log into a *target* computing component. Second, enter one of the following keywords for each `action` you want to complete:

- `halt`
- `shutdown`
- `reboot`

Example 1. To reboot ICE leader nodes `r1lead` and `r2lead`, use the `leader` target type as shown in the following:

```
admin # cm power reboot -t leader r1lead,r2lead
```

Example 2. To power off nodes `n0` and `n1`, specify the following:

```
admin # cm power halt -t node n0,n1
```

For more information, including a comprehensive parameter list, enter the following command:

```
# cm power -h
```

The following topics provide more information about the `cm power` command:

- **Power commands for clusters without leader nodes and for HPE SGI 8600 clusters**
- **Power commands for ICE compute nodes and compute nodes**
- **Power commands for ICE leader nodes**
- **Power commands for HPE Cray EX, HPE Apollo 9000, and HPE SGI 8600 compute chassis**
- **Power commands for HPE Cray EX, HPE Apollo 9000, and HPE SGI 8600 switches**
- **Power commands for HPE Cray EX slot, tray, and controller components**
- **Powering on a cluster with scalable unit (SU) leader nodes**
- **Powering off a cluster with scalable unit (SU) leader nodes**

NOTE: Follow your site power-on and power-off procedures to ensure the following:

- That your leader, storage, license, login, and other special-purpose nodes boot before the compute nodes boot.
- That your leader, storage, license, login, and other special-purpose nodes power-off after compute nodes power off.

Power commands for clusters without leader nodes and for HPE SGI 8600 clusters

To manage the power status of the entire cluster (excluding the admin node), specify the target type of `system` and the desired action on the `cm power` command.

NOTE: Do not use the power commands in this topic on clusters with scalable unit (SU) leader nodes. For clusters with SU leader nodes, see one of the following topics:

- **Powering on a cluster with scalable unit (SU) leader nodes**
- **Powering off a cluster with scalable unit (SU) leader nodes**

The following are example power management commands for entire clusters.

- The following command powers down the cluster:

  ```
  # cm power off -t system
  ```

The compute nodes and/or ICE compute nodes are powered down first. Then, the rack leaders are powered down.

- The following command powers up the cluster:

  ```
  # cm power on -t system
  leader node r1lead power ON
  600 sec wait for leader r1lead to boot
  direct node n0 power ON
  leader node r1lead is BOOTED
  leader node r1lead is BOOTED
  compute node r1i0n0 BOOTED
  compute node r1i0n3 BOOTED
  compute node r1i0n4 BOOTED
  ...
  compute node r1i2n2 BOOTED
  compute node r1i2n11 BOOTED
  compute node r1i2n4 BOOTED
  compute node r1i2n14 BOOTED
  compute node r1i2n15 BOOTED
  ```

  The rack leaders and compute nodes power on first. On HPE SGI 8600 clusters, the cluster manager then powers on the ICE compute nodes.

- The following command queries the power on/off status of an HPE Apollo 9000 cluster:

  ```
  # cm power status -t system
  leader1      BOOTED
  r1c1t1n1     BOOTED
  r1c1t1n2     BOOTED
  r1c1t1n3     BOOTED
  .
  .
  .
  r1c1t8n2     On
  r1c1t8n3     BOOTED
  .
  .
  .
  ```

Power commands for ICE compute nodes and compute nodes

To manage ICE compute nodes and compute nodes with the `cm power` command, specify the following:

- A target type of `node`
- An action
- A target list

The following are example power management commands for ICE compute nodes and compute nodes.

- The following command powers on `n0`, which is a compute node:

  ```
  # cm power on -t node n0
  ```

- The following command powers on all compute nodes:

  ```
  # cm power on -t node 'n*'
  ```

  To manage the boot order of a group of compute nodes, not ICE compute nodes, see the following:

Changing compute node configuration elements

- The following command queries and displays the status of all compute nodes:

  ```
  # cm power status -t node 'n*'
  ```

- The following command powers down one compute node:

  ```
  # cm power off -t node n0
  ```

- The following command reboots compute node `n0` with a three-minute timeout:

  ```
  # cm power reboot -t node n0 -w 180
  ```

- The following command powers on the ICE compute node at rack 1, chassis 3, slot 10:

  ```
  # cm power on -t node r1i3n10
  ```

  If the associated leader node is off, this action powers on components in the following order:

  - The leader node itself. The command powers on the leader note and waits for its successful boot. There is a 10-minute timeout.
  - The associated chassis, if needed.
  - The ICE compute node itself.

- The following command powers on a group of ICE compute nodes in a rack:

  ```
  # cm power on -t node 'r1i0n[2-5]' -w 300
  cmc node r1i0c power ON
  compute node r1i0n3 already BOOTED
  compute node r1i0n5 power ON
  compute node r1i0n4 power ON
  compute node r1i0n2 power ON
  compute node r1i0n5 is BOOTED
  compute node r1i0n2 is BOOTED
  300 second timeout exceeded waiting for boot of r1i0n4
  ```

  The command powers on and attempts to boot the ICE compute nodes in slots 2, 3, 4, and 5 in chassis 0 of rack 1. Note the 5-minute wait time for booting.

- The following command queries the status of all ICE compute nodes in a rack:

  ```
  # cm power status -t node 'r1i*n*'
  ```

- The following command queries the power status of compute node `r1c4t6n3` on an HPE Apollo 9000 cluster:

  ```
  cm power status -t node r1c4t6n3
  ```

- The following command powers off the specified ICE compute node.

  ```
  # cm power off -t node r1i3n10
  ```

  The associated rack leader and chassis are unaffected.

- The following command reboots the specified ICE compute node:

  ```
  # cm power reboot -t node r1i3n10
  ```

- The following commands turn on the ID LED of node `r1i0n0` and then turn off the ID LED:

  ```
  # cm power uid_on -t node r1i0n0
  # cm power uid_off -t node r1i0n0
  ```

Power commands for ICE leader nodes

Rack leader power management requires you to use the `cm power` command to specify a target type of `leader`, an action, and a target list.

The following are example power management commands for rack leaders.

- The following command powers on the leader node for rack 1:

  ```
  # cm power on -t leader r1lead
  ```

- The following command shuts down the specified leader node:

  ```
  # cm power shutdown -t leader r3lead
  leader node r3lead has been issued a shutdown -h now command
  leader node r3lead is DOWN
  ```

  The associated ICE compute nodes and chassis are unaffected.

- The following command queries and then displays the status of all leader nodes:

  ```
  # cm power status -t leader '*'
  r1lead       BOOTED
  r2lead       BOOTED
  r3lead       OFF
  ```

- The following command reboots a leader node with a three-minute timeout:

  ```
  # cm power reboot -t leader r3lead -w 180
  ```

- The following commands turn on the ID LED for leader node `r1lead` and then turn off the ID LED:

  ```
  # cm power uid_on -t leader r1lead
  # cm power uid_off -t leader r1lead
  ```

Power commands for HPE Cray EX, HPE Apollo 9000, and HPE SGI 8600 compute chassis

On the HPE Cray EX, HPE Apollo 9000, and HPE SGI 8600 clusters, power management requires you to use the `cm power` command with the target type of `chassis`, an action, and a target list.

Specify a chassis by its rack number and its chassis number. For example, on an HPE SGI 8600 cluster, `r1i1` specifies chassis 1 on rack 1.

On an HPE SGI 8600 cluster, powering on an ICE compute node powers on its associated leader and chassis, but the converse is not true. Likewise, powering on/off a chassis powers on/off its associated switches and ICE compute blades.

The following are example power management commands for chassis.

- On an HPE Cray EX cluster, the following command powers on all chassis:

  ```
  # cm power on -t chassis "*"
  ```

- On an HPE Apollo 9000 cluster, the following command powers on the chassis, switches, and compute nodes in chassis 0 in rack 1 :

  ```
  # cm power on -t chassis r1c1
  ```

- On an HPE SGI 8600 cluster, the following command powers off the chassis, switches, and ICE compute nodes 1 in chassis1, rack 3:

  ```
  # cm power off -t chassis r3i1
  ```

- On an HPE SGI 8600 cluster, the following command powers off all chassis, switches, and ICE compute nodes in rack3:

  ```
  # cm power off -t chassis 'r3i*'
  ```

  Notice the use of apostrophes (`'  '`) with the wildcard to ensure that a matching filename is not targeted.

For more information about wildcards and node identification, see one of the following:

Using the `cm` commands

Node identification

Power commands for HPE Cray EX, HPE Apollo 9000, and HPE SGI 8600 switches

Like a chassis, you can manage switches selectively. You can turn them on and off and query their power status. For switches, use the `cm power` command with the following:

- A target type of `switch`
- An action
- A target list

Specify a switch by its switch number and its associated rack and chassis. For example, on an HPE SGI 8600 cluster, `r1i0s0` specifies switch 0 associated with chassis 0 on rack 1.

The following are power management command examples for switches.

- The following command returns the power status for a switch on an HPE Cray EX cluster:

  ```
  # cm power status -t switch x9000c1r3b0
  x9000c1r3b0 On
  ```

- The following command powers on switch 0 associated with chassis 0 in rack 1:

  ```
  # cm power on -t switch r1i0s0
  ```

- The following command powers off switch 1 associated with chassis 1 in rack 3:

  ```
  # cm power off -t switch r3i1s1
  ```

- The following command returns the status of all switches on an HPE Apollo 9000 cluster:

  ```
  # cm power status -t switch '*'
  r1c0s0       ON
  r1c0s1       ON
  r1c1s0       ON
  r1c1s1       ON
  ```

```
r1c2s0        ON
r1c2s1        ON
r1c3s0        ON
r1c3s1        ON
```

Power commands for HPE Cray EX slot, tray, and controller components

The following examples show power commands run on an HPE Cray EX cluster:

- The following command returns the status of the nodes in cabinet 9000, chassis 1, and slot 0:

  ```
  # cm power status -t slot "x9000c1s*"
  x9000c1s0     On
  x9000c1s1     On
  x9000c1s2     On
  x9000c1s3     On
  ```

- The following command returns the status of the slot associated with `x9000c1s0`:

  ```
  # cm power status -t slot "x9000c1s0"
  x9000c1s0     On
  ```

- The following command resets controller `x9000c1s0b0`:

  ```
  # cm power reset -t controller  "x9000c1s0b0"
  ```

Powering on a cluster with scalable unit (SU) leader nodes

This topic explains powering on a cluster at a high level. Depending on your cluster, you might have site-specific practices.

Procedure

1. Log into the admin node as the root user.
2. Power-up all SU leader nodes.

   For example:

   ```
   # cm power on -t leader 'leader*'
   ```

   Wait for the leader nodes to boot.
3. Enter the following command to display the status of the cluster trivial database (CTDB) that runs on the SU leader nodes:

   ```
   # cm node run -t role su-leader ctdb status
   ```
4. After the leaders report that the CTDB status is normal, make sure that shared storage is mounted on the admin node:

   ```
   # enable-su-leader --skip-sync
   # df -h
   ```
5. Power on all cluster components, including compute nodes, login nodes, and device nodes.

   For example:

   ```
   # cm power on -t node 'r*c*t*n*'
   ```

Powering off a cluster with scalable unit (SU) leader nodes

This topic explains powering off a cluster at a high level. Depending on your cluster, you might have site-specific practices.

Procedure

1. Power off all compute nodes, login nodes, and device nodes.

   If your site has an approved order for powering off nodes, use that order. Keep the admin node and SU leader nodes powered on.

   For example:

   ```
   # cm power off -t node 'r*c*t*n*'
   ```

2. Power off the SU leader nodes.

   For example:

   ```
   # cm power off -t leader 'leader*'
   ```

   Wait for all SU leader nodes to power down.

3. Shut down the operating system on the admin node, and power off the admin node.

   If the admin node is a high availability (HA) admin node, log into the virtual machine and power down the virtual machine.

   If the admin node is a single admin node, go to the console and power it down.

   For example, you can issue an `init0` to shut down the operating system and then power down the admin node. Use your site practices for this step.

Power and energy management

Power management includes the following:

- Monitoring energy use
- Managing energy use at the system level and at the job level
- Managing the thermal health of the cluster

For power management information, see the following:

HPE Performance Cluster Manager Power Management Guide

`clush` command

The `clush` command issues an `ssh` call to a group of nodes and runs the specified command on all the nodes in the group.

NOTE: The `clush` command is not aware of node hostname aliases by default. To use node hostname aliases in `clush` commands, create alternative node hostnames in your DNS or in your `/etc/hosts` file.

The following are `clush` command examples:

- Example 1. From the admin node, to run the hostname command on all the ICE leader nodes, enter the following:

  ```
  # clush -g leader hostname
  ```

- Example 2. From the admin node, to run the hostname command on all the scalable unit (SU) leader nodes, enter the following:

  ```
  # clush -g su-leader hostname
  ```

- Example 3. From the admin node, to display the hostname of all ICE compute nodes in the cluster, enter the following:

  ```
  # clush -g ice-compute hostname
  ```

  The preceding command runs the hostname command on all the ICE compute nodes.

  If the preceding command does not work, verify that the routing information protocol (RIP) is enabled on the management switch. The RIP protocol is enabled by default, but it is possible that the protocol has been disabled. To run clush commands on all ICE compute nodes from the admin node, the RIP protocol must be enabled on the management switches.

  For example, to retrieve the status of the RIP protocol on mgmtsw1, enter the following command:

  ```
  # switchconfig rip -i -s mgmtsw1
  ```

  To set the RIP protocol on mgmtsw1, enter the following command:

  ```
  # switchconfig rip -e -v all -s mgmtsw1
  ```

- Example 4. From the admin node, to run the hostname command on just r1lead and r2lead, enter the following:

  ```
  # clush -w r1lead,r2lead hostname
  ```

- Example 5. From the admin node, to run the uptime command on the compute nodes in racks 1 and 2 on an HPE Apollo 9000 cluster, enter the following.

  ```
  # clush -w r[1-2]c[1-4]t[1-8]n[1-4] uptime
  ```

  The parameters are as follows:

  - The r[1-2] targets racks 1 and 2.
  - The c[1-4] targets all the chassis in the rack.
  - The t[1-8] targets all the trays.
  - The n[1-4] targets all the nodes in the tray. There are a maximum of four nodes in a tray.

pdsh and pdcp commands

The pdsh command is the parallel shell utility. The pdcp command is the parallel copy/fetch utility.

NOTE: On large clusters, consider using the clush command rather than the pdsh command or the pdcp command. The clush command is designed to scale better on large clusters.

For more information, see the following:

clush command

The cluster manager creates the following groups by default:

- compute. This group includes all the compute nodes in the cluster.
- leader. This group exists only on clusters with ICE leader nodes.
- ice-compute. This group exists on clusters with ICE leader nodes.
- su-leader. This group exists on clusters with scalable unit (SU) leader nodes.

The cluster manager populates some pdsh group files for the various node types. On the admin node, the cluster manager populates the leader and compute group files with the list of online nodes in each of those groups. On a leader node, the cluster manager populates the ice-compute group with the list of all the online ICE compute nodes in that rack. On a compute node, the cluster manager populates the compute group with the list of all the online compute nodes in the whole system.

To see the names of all the node groups defined in the cluster, including the default groups, enter the following command:

```
# ls /etc/dsh/group
```

The following topics explain how to use the pdsh and pdcp commands:

- **pdsh command examples**
- **Creating pdsh group files**

For more information, see the pdsh(1) and pdcp(1) manpages.

pdsh command examples

The following are pdsh command examples:

- Example 1. From the admin node, to run the hostname command on all the ICE leader nodes, enter the following:

  ```
  # pdsh -g leader hostname
  ```

- Example 2. From the admin node, to run the hostname command on all the scalable unit (SU) leader nodes, enter the following:

  ```
  # pdsh -g su-leader hostname
  ```

- Example 3. From the admin node, to display the hostname of all ICE compute nodes in the cluster, enter the following:

  ```
  # pdsh -g ice-compute hostname
  ```

  The preceding command runs the hostname command on all the ICE compute nodes.

  If the preceding command does not work, verify that the routing information protocol (RIP) is enabled on the management switch. The RIP protocol is enabled by default, but it is possible that the protocol has been disabled. To run pdsh commands on all ICE compute nodes from the admin node, the RIP protocol must be enabled on the management switches.

  For example, to retrieve the status of the RIP protocol on mgmtsw1, enter the following command:

  ```
  # switchconfig rip -i -s mgmtsw1
  ```

  To set the RIP protocol on mgmtsw1, enter the following command:

  ```
  # switchconfig rip -e -v all -s mgmtsw1
  ```

- Example 4. From the admin node, to run the hostname command on just r1lead and r2lead, enter the following:

  ```
  # pdsh -w r1lead,r2lead hostname
  ```

- Example 5. From the admin node, to run the uptime command on the compute nodes in racks 1 and 2 on an HPE Apollo 9000 cluster, enter the following.

  ```
  # pdsh -w r[1-2]c[1-4]t[1-8]n[1-4] uptime
  ```

  The parameters are as follows:

  - The r[1-2] targets racks 1 and 2.
  - The c[1-4] targets all the chassis in the rack.
  - The t[1-8] targets all the trays.
  - The n[1-4] targets all the nodes in the tray. There are a maximum of 4 nodes in a tray.

NOTE: The pdsh command is not aware of node hostname aliases by default. To use node hostname aliases in pdsh commands, create alternate node hostnames in your DNS or in your /etc/hosts file.

Creating pdsh group files

You can direct the cluster manager to create pdsh group files for custom groups or network groups at the time you create the node group. Use one of the following methods:

- Method 1 - Create the pdsh group at the time you create a specific network group or custom group

  If you specify --pdsh-group on the cm command line, the cluster manager creates a pdsh group for the node group at the time it creates the node group.

  For example, the following command creates custom group giraffe with nodes n0 and n1 and also creates a pdsh custom group with those same nodes:

  ```
  # cm group custom add -c giraffe --pdsh-group -n n0,n1
  ```

- Method 2 - Direct the cluster manager to create pdsh groups for each node group you create over time

  If you want to create pdsh groups for several node groups, you can direct the cluster manager to automatically create pdsh groups for each node group you create. This method lets you omit the --pdsh-group specification for each cm command you run to create each group.

  For example, assume the following sequence:

  ```
  # cm group custom set --pdsh-groups         # enable auto pdsh group creation for custom groups
  # cm group custom add -c fish -n n0,n1      # create custom group fish
  # cm group custom add -c whale -n n100,n101 # create custom group whale
  # cm group custom unset --pdsh-groups       # disable auto pdsh group creation for custom groups
  ```

  In this example, if you create a network group in this command sequence, the cluster manager does not create a pdsh group for that network group. To automatically create pdsh groups for network groups, specify another sequence of commands that start with the following command:

  ```
  # cm group network set --pdsh-groups
  ```

NOTE: If your job scheduler creates cluster manager custom groups, you might incur performance consequences if `pdsh` group creation is enabled at all times.

To determine which node groups are also `pdsh` groups, examine the files in the following directory:

```
/etc/dsh/group
```

Using the administrative interface

After you log into the admin node, you can use administrative commands such as the following:

- `cadmin`
- `cm node set`
- `cm node show`
- `cm node unset`
- `cm controller show`

For information about how to preserve custom configuration changes across `cm node update config` calls, see the following:

Preserving custom configuration changes

For information about the administrative commands, enter *`command`* `-h` and read the help output.

Displaying node information

From the admin node, enter the following command to display all nodes that are defined in the cluster database:

```
cm node show
```

As the following examples show, the command can show output in a compressed or expanded format.

Example 1:

The command in this example specifies `-z`, which displays all nodes in a compressed format all on one line. You can use your mouse to copy output from this command and use it as input to another command. The command is as follows:

```
# cm node show -z
leader[1-3],packer-login,r1c[1-4]t[1-8]n[1-4]
```

Example 2:

The command in this example specifies `-z`, which displays all nodes in a compressed format with nodes of specific types on individual lines. This style of output is useful for isolating node name sequences. You can use your mouse to copy output from one of the lines and use it as input to another command. The command is as follows:

```
# cm node show -Z
leader[1-3]
packer-login
r1c[1-4]t[1-8]n[1-4]
```

Example 3. The following command specifies `-n '*'` , which displays compute nodes, leader nodes, switches, and other entities.

```
# cm node show -n '*'
indeed
leader1
```

```
leader2
leader3
leader4
leader5
leader6
leader7
leader8
leader9
mgmtsw0
mgmtsw10
n1
n100
n2449
n2521
n289
n577
netgrouptest1
netgrouptest2
r999lead
```

Example 4.

The following command displays nodes in the custom group `comp` on an HPE Apollo 9000 cluster:

```
# cm node show -n @comp
r1c1t1n1
r1c1t1n2
r1c1t1n2v1
r1c1t1n2v10
r1c1t1n2v11
r1c1t1n2v12
r1c1t1n2v13
r1c1t1n2v14
r1c1t1n2v15
r1c1t1n2v2
r1c1t1n2v3
r1c1t1n2v4
r1c1t1n2v5
r1c1t1n2v6
r1c1t1n2v7
r1c1t1n2v8
r1c1t1n2v9
r1c1t1n3
r1c1t1n3v1
r1c1t1n3v10
r1c1t1n3v11
r1c1t1n3v12
r1c1t1n3v13
r1c1t1n3v14
r1c1t1n3v15
r1c1t1n3v2
r1c1t1n3v3
r1c1t1n3v4
r1c1t1n3v5
r1c1t1n3v6
r1c1t1n3v7
r1c1t1n3v8
r1c1t1n3v9
```

```
.
.
.
```

Example 5: The following command displays leader nodes and compute nodes on an HPE Apollo 9000 cluster:

```
# cm node show
leader1
leader2
leader3
packer-login
r1c1t1n1
r1c1t1n2
r1c1t1n2v1
r1c1t1n2v10
r1c1t1n2v11
r1c1t1n2v12
r1c1t1n2v13
r1c1t1n2v14
r1c1t1n2v15
r1c1t1n2v2
r1c1t1n2v3
r1c1t1n2v4
r1c1t1n2v5
r1c1t1n2v6
r1c1t1n2v7
r1c1t1n2v8
r1c1t1n2v9
r1c1t1n3
r1c1t1n3v1
r1c1t1n3v10
r1c1t1n3v11
r1c1t1n3v12
r1c1t1n3v13
r1c1t1n3v14
r1c1t1n3v15
r1c1t1n3v2
r1c1t1n3v3
r1c1t1n3v4
r1c1t1n3v5
r1c1t1n3v6
r1c1t1n3v7
r1c1t1n3v8
r1c1t1n3v9
.
.
.
```

Example 6. The following command displays controller information for each of the nodes:

```
# cm node show -C
NODE           TYPE                 IPADDRESS    MACADDRESS         PROTOCOL                        CHANNEL
x1000c0s0b0n0  cmm_node_controller  10.176.0.11  02:03:e8:00:30:00  Cray,NO_IPMI,None,redfish None  root
x1000c0s0b0n1  cmm_node_controller  10.176.0.11  02:03:e8:00:30:00  Cray,NO_IPMI,None,redfish None  root
x1000c0s0b1n0  cmm_node_controller  10.176.0.12  02:03:e8:00:30:10  Cray,NO_IPMI,None,redfish None  root
x1000c0s0b1n1  cmm_node_controller  10.176.0.12  02:03:e8:00:30:10  Cray,NO_IPMI,None,redfish None  root
x1000c0s1b0n0  cmm_node_controller  10.176.0.13  02:03:e8:00:31:00  Cray,NO_IPMI,None,redfish None  root
x1000c0s1b0n1  cmm_node_controller  10.176.0.13  02:03:e8:00:31:00  Cray,NO_IPMI,None,redfish None  root
x1000c0s1b1n0  cmm_node_controller  10.176.0.14  02:03:e8:00:31:10  Cray,NO_IPMI,None,redfish None  root
```

NOTE: In the preceding example, output was truncated from the right for inclusion in this documentation.

Example 7. The following command displays image information for each of the cluster nodes:

```
# cm node show -I
NODE          IMAGE.NAME        KERNEL            IMAGEPENDING CLONINGDATE                  CLONINGBLOCKDEVICE IMAGETRANSPORT
x1000c0s0b0n1 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
x1000c0s0b1n0 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
x1000c0s0b1n1 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
x1000c0s1b0n0 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
x1000c0s1b0n1 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
x1000c0s1b1n0 sles15spX-compute 5.3.18-22-default False        20XX-11-07T23:59:54.788+0000 default            rsync
.
.
.
```

Example 8. The following command displays information about management interfaces:

```
# cm node show -M
NODE       NETWORK.NAME  IPADDRESS   SUBNETMASK   MACADDRESS         MGMTSERVERIP  DEFAULTGATEWAY
fmn        None          172.23.0.5  255.255.0.0  48:df:37:bd:0c:40  default       default
leader1    None          172.23.0.2  255.255.0.0  48:df:37:bd:94:a0  default       default
leader2    None          172.23.0.3  255.255.0.0  48:df:37:bd:96:50  default       default
leader3    None          172.23.0.4  255.255.0.0  48:df:37:bd:96:00  default       default
river-gpu  None          172.23.0.6  255.255.0.0  18:c0:4d:13:d7:58  default       default
```

Example 9. The following command displays information about management BMC interfaces.

```
# cm node show -B
NODE           CARDIPADDRESS  CARDMACADDRESS     CARDTYPE  PROTOCOL                              USERNAME
leader1        172.24.0.2     b4:2e:99:fe:27:de  IPMI      NO_DCMI,ipmi,no_type,redfish          root
leader2        172.24.0.3     b4:2e:99:fe:28:6a  IPMI      NO_DCMI,ipmi,no_type,redfish          root
leader3        172.24.0.4     b4:2e:99:fe:14:54  IPMI      NO_DCMI,ipmi,no_type,redfish          root
pbs01          172.24.0.7     b4:2e:99:fe:27:aa  IPMI      NO_DCMI,ipmi,no_type,redfish          root
pbs02          172.24.0.8     b4:2e:99:fe:28:d2  IPMI      NO_DCMI,ipmi,no_type,redfish          root
warhawk-lm01   172.24.0.36    18:c0:4d:1c:ce:c0  IPMI      NO_DCMI_NM,NO_REDFISH,ipmi            root
warhawk-lm02   172.24.0.53    b4:2e:99:de:99:b2  IPMI      NO_DCMI,NO_DCMI_NM,NO_REDFISH,ipmi    root
warhawk-lm03   172.24.0.70    18:c0:4d:1c:ca:a8  IPMI      NO_DCMI,NO_DCMI_NM,NO_REDFISH,ipmi    root
warhawk-lm04   172.24.0.87    18:c0:4d:1c:cf:f4  IPMI      NO_DCMI,NO_DCMI_NM,NO_REDFISH,ipmi    root
warhawk-mla01  172.24.0.20    18:c0:4d:1a:cc:0c  IPMI      NO_DCMI,NO_DCMI_NM,NO_REDFISH,ipmi    root
.
.
.
```

NOTE: In the preceding example, output was truncated from the right for inclusion in this documentation.

Bringing a node online or setting a node offline

The following examples show how to bring a node online or set a node offline:

- To set `r1i0n0` offline, enter the following command:

  ```
  # cm node set --administrative-state offline -n r1i0n0
  ```

- To set `r1i0n0` online, enter the following command:

  ```
  # cm node set --administrative-state online -n r1i0n0
  ```

When you set the node administrative state to offline, the cluster manager changes the node status in the following:

- The cluster manager database
- Configuration files that depend on the database

The cluster manager ignores the node for all subsequent actions targeting online nodes.

Creating notes for nodes

You can create and display node-specific notes. For example, you might want to record why you placed a node offline. Use the following commands:

- `cm node set --node-notes 'text' -n node`
- `cm node show --node-notes -n node`

The variables are as follows:

Variable	Specification
text	A text string.
node	One or more node hostnames.

For example:

```
# cm node set --node-notes 'Booting problems May 3.  Put offline.' -n n1
# cm node show --node-notes -n n1
Booting problems May 3.  Put offline.
```

Changing compute node configuration elements

The following examples show how to change the hostname and IP address of a compute node.

- To change the hostname of `n0` to `myservice`, enter the following command:

  ```
  admin:~ # cm node set --name myservice -n n0
  ```

- To retrieve the IP addresses currently configured for `myservice`, enter the following command:

  ```
  admin:~ # cm node show --ips -n myservice
  IP Address Information for SMC node: n0

  ifname         ip                Network

  myservice-bmc  172.24.0.3        head-bmc
  myservice      172.23.0.3        head
  myservice-ib0  10.148.0.254      ib0
  myservice-ib1  10.149.0.67       ib1
  myhost         172.24.0.55       head-bmc
  myhost2        172.24.0.56       head-bmc
  myhost3        172.24.0.57       head-bmc
  ```

- To change the IP address on `myservice-ib0`, enter the following command:

  ```
  admin:~ # cm node set --update-ip 172.23.0.199 --ifname eth0 --net head -n myservice
  ```

- To set the boot order for compute node `myservice`, enter the following command:

  ```
  # cm node set --boot-order value -n myservice
  ```

  For *value*, specify any positive integer number. The default is 1. This value is the boot order specification.

You can use the `cm node set` command to control the boot order (boot sequence) of a group of compute nodes. You can implement this kind of control to ensure that the server nodes boot before clients nodes. For example, you might want to use this feature with NFS, CIFS, or SMB servers.

When you boot a group of compute nodes with varying boot order values, the cluster manager first boots all nodes with boot order 1. Then the cluster manager boots those nodes with boot order 2, and so on.

Some power-down operations honor a specified boot order. These operations power down the compute nodes starting with those operations that have the largest boot order number. The power-down operations that respect boot order are `off`, `shutdown`, and `halt`.

The `reboot`, `reset`, and `cycle` operations do not respect boot order. These operations act on all target compute nodes simultaneously.

For information about power-on operations and power-down operations, see the following:

Rebooting, halting, powering on, and powering off

To make large-scale configuration changes, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Assigning or reassigning a compute node to a scalable unit (SU) leader node

The following procedure explains how to assign one or more new compute nodes to an SU leader node IP alias. You can also use this command to reassign a compute node to a different SU leader IP alias.

Procedure

1. Log into the admin node as the root user.
2. Use the following command:

```
cm node set -n node --su-leader new_su_leader_ip
```

The variables are as follows:

Variable	Specification
node	One or more node hostnames.
su_leader_ip	The IP alias of the new SU leader for the node.

Changing the admin node hostname and IP address on the house network

The procedure in this topic explains the following:

- How to retrieve information about the admin node
- How to update the admin node hostname or IP address

The examples show how to change the address information for the admin node on the house network.

Procedure

1. Log into the admin node as the root user.
2. Use the `cadmin` command to retrieve information about the current house network IP address.

For example:

```
admin:~ # cadmin --show-house-network-info
-----Network Information-----
broadcast       :       137.38.82.255
base_ip :       137.38.82.0           # the IP of the house network
netmask :       255.255.255.0
gateway :       137.38.82.254
ip      :       137.38.82.166         # the IP address of the admin node
```

3. Use the `cadmin` command in the following format to assign a new IP address to the admin node:

   ```
   cadmin --set-house-network ip_addr,netmask,gateway_info
   ```

   The variables are as follows:

Variable	Specification
ip_addr	The new IP address that you want to assign to the admin node.
netmask	The network mask you want to assign to the new IP address.
gateway_info	Either the default gateway you want to assign to the new IP address or the keyword `no_gateway`.

   You can also use the `cadmin --set-house-network` command to specify a new network mask or new gateway information for the admin node. In that case, specify the existing admin node IP address, the new network mask, and/or the new default gateway.

4. Use the `service network restart` command to restart the network services.

   When you use the `--set-house-network` parameter to the `cadmin` command to change any of the networking information, restart network services.

   The following examples show how to use the `service network restart` command:

   Example 1. On a the admin node, enter the following:

   ```
   admin:~ # cadmin --set-house-network 137.38.82.165,255.255.255.0,137.38.82.253
   admin:~ # systemctl network restart
   ```

   Example 2. On the admin node, enter the following:

   ```
   admin:~ # cadmin --set-house-network 137.38.82.165,255.255.255.0,no_gateway
   admin:~ # systemctl network restart
   ```

Displaying network information

The following examples show how to use the `cadmin` command to display network information.

- To show the cluster domain, enter the following command:

  ```
  admin:~ # cadmin --show-cluster-domain
  The cluster domain is: cm.clusterdomain.com
  ```

- To show the admin node house network domain, enter the following command:

  ```
  admin:~ # cadmin --show-admin-domain
  The admin node house network domain is: clusterdomain.com
  ```

- To set the cluster domain, enter the following command:

  ```
  admin:~ # cadmin --set-cluster-domain domain
  ```

- To set the admin node house network domain, enter the following command:

  ```
  admin:~ # cadmin --set-admin-domain domain
  ```

Changing switch management network settings

The following examples show how to use the `cadmin` command to change the switch management network settings.

- To retrieve the current switch management value for a specified node, enter the following command:

  ```
  admin:~ # cm node show --switch-mgmt-network -n admin
  no
  ```

  In this example, returned value is `no`. This value means that there is no switch management network. This configuration is a nondefault configuration.

- To enable the switch management network for a specified node that is connected to managed top-level switches, enter the following command:

  ```
  admin:~ # cm node set --switch-mgmt-network -n admin
  ```

- To disable the switch management network for a specified node that is connected to managed top-level switches, enter the following command:

  ```
  admin:~ # cm node unset -switch-mgmt-network -n admin
  ```

Displaying controller information

Each cluster can include various controllers for chassis, nodes, switches, and other components. The `cm controller show` command displays information about the controllers.

The controller commands are the same for all clusters, however some clusters have more types of controllers than others.

Example 1. The following command shows all the cluster controllers:

```
# cm controller show
NAME         TYPE                            ADMINISTRATIVESTATUS  PROTOCOL  CHANNEL  MACADDRESS         IPADDRESS    IPV6ADDRESS
x5000c1r1b0  cmm_switch_controller           online                None      None     02:13:88:01:61:00  10.176.0.3   None
x5000c1r3b0  cmm_switch_controller           online                None      None     02:13:88:01:63:00  10.176.0.4   None
x5000c1r5b0  cmm_switch_controller           online                None      None     02:13:88:01:65:00  10.176.0.5   None
x5000c1r7b0  cmm_switch_controller           online                None      None     02:13:88:01:67:00  10.176.0.6   None
x5000c1s0b0  cmm_node_controller             online                None      None     02:13:88:01:30:00  10.176.0.7   None
x5000c1s0b1  cmm_node_controller             online                None      None     02:13:88:01:30:10  10.176.0.8   None
x5000c1s1b0  cmm_node_controller             online                None      None     02:13:88:01:31:00  10.176.0.9   None
x5000c1s1b1  cmm_node_controller             online                None      None     02:13:88:01:31:10  10.176.0.10  None
x5000c1s2b0  cmm_node_controller             online                None      None     02:13:88:01:32:00  10.176.0.11  None
x5000c1s2b1  cmm_node_controller             online                None      None     02:13:88:01:32:10  10.176.0.12  None
x5000cec0    cabinet_environment_controller  online                None      None     None               None         None
x5000cec1    cabinet_environment_controller  online                None      None     None               None         None
```

Example 2. The following command shows information for HPE Cray EX chassis environment controllers (CECs):

```
# cm controller show -t environment
NAME       TYPE                            ADMINISTRATIVESTATUS  PROTOCOL  CHANNEL  MACADDRESS  IPADDRESS  IPV6ADDRESS
x1000cec0  cabinet_environment_controller  online                None      None     None        None       None
x1000cec1  cabinet_environment_controller  online                None      None     None        None       None
x1001cec0  cabinet_environment_controller  online                None      None     None        None       None
x1001cec1  cabinet_environment_controller  online                None      None     None        None       None
x1002cec0  cabinet_environment_controller  online                None      None     None        None       None
x1002cec1  cabinet_environment_controller  online                None      None     None        None       None
x1003cec0  cabinet_environment_controller  online                None      None     None        None       None
x1003cec1  cabinet_environment_controller  online                None      None     None        None       None
.
.
.
```

Example 3. The following command shows location information for the controllers:

```
# cm controller show –l
NAME         RACK  CHASSIS  TRAY  NODE
x1000c0r3b0  None  0        3     1000
x1000c0r7b0  None  0        7     1000
x1000c0s0b0  None  0        0     1000
x1000c0s0b1  None  0        0     1000
x1000c0s1b0  None  0        1     1000
x1000c0s1b1  None  0        1     1000
x1000c0s2b0  None  0        2     1000
.
.
.
```

Example 4. The following command shows the node controllers in the cluster:

```
# cm controller show -t node
NAME         TYPE                 ADMINISTRATIVESTATUS  PROTOCOL                   CHANNEL  MACADDRESS         IPADDRESS
x1000c0s0b0  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:30:00  10.176.0.11
x1000c0s0b1  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:30:10  10.176.0.12
x1000c0s1b0  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:31:00  10.176.0.13
x1000c0s1b1  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:31:10  10.176.0.14
x1000c0s2b0  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:32:00  10.176.0.15
x1000c0s2b1  cmm_node_controller  online                Cray,NO_IPMI,None,redfish  None     02:03:e8:00:32:10  10.176.0.16
.
.
.
```

NOTE: The preceding output has been truncated from the right for inclusion in this documentation.

Example 5: The following command shows the names of node controllers and their associated nodes:

```
# cm controller show -c 'x1000c0*' -t node --nodes
x1000c0s0b0
   x1000c0s0b0n0
   x1000c0s0b0n1
x1000c0s0b1
   x1000c0s0b1n0
   x1000c0s0b1n1
x1000c0s1b0
   x1000c0s1b0n0
   x1000c0s1b0n1
x1000c0s1b1
   x1000c0s1b1n0
   x1000c0s1b1n1
x1000c0s2b0
   x1000c0s2b0n0
   x1000c0s2b0n1
x1000c0s2b1
   x1000c0s2b1n0
   x1000c0s2b1n1
x1000c0s3b0
   x1000c0s3b0n0
   x1000c0s3b0n1
x1000c0s3b1
   x1000c0s3b1n0
   x1000c0s3b1n1
x1000c0s4b0
   x1000c0s4b0n0
   x1000c0s4b0n1
x1000c0s4b1
```

```
   x1000c0s4b1n0
   x1000c0s4b1n1
x1000c0s5b0
   x1000c0s5b0n0
   x1000c0s5b0n1
x1000c0s5b1
   x1000c0s5b1n0
   x1000c0s5b1n1
x1000c0s6b0
   x1000c0s6b0n0
   x1000c0s6b0n1
x1000c0s6b1
   x1000c0s6b1n0
   x1000c0s6b1n1
x1000c0s7b0
   x1000c0s7b0n0
   x1000c0s7b0n1
x1000c0s7b1
   x1000c0s7b1n0
   x1000c0s7b1n1
```

Deleting a controller

The procedure in this topic explains how to delete a controller from the cluster database. If you want to delete a controller, the cluster manager requires you to delete the nodes associated with the controller, too. You cannot delete a controller if the controller is connected to nodes that are active in the cluster database.

You can delete both a controller and all the nodes associated with the controller with one command.

Procedure

1. Log into the admin node as the root user.

2. (Conditional) Obtain information about nodes and their associated controllers.

   Complete this step if you do not know the hostnames of the nodes and controllers you want to delete. Enter the following command to list nodes and their associated node controllers: For example:

   ```
   # cm node show -C
   x1000c0s0b0n0  cmm_node_controller  10.176.0.11  02:03:e8:00:30:00  Cray,NO_IPMI,None,redfish None  root
   x1000c0s0b0n1  cmm_node_controller  10.176.0.11  02:03:e8:00:30:00  Cray,NO_IPMI,None,redfish None  root
   .
   .
   .
   ```

3. Use the `cm controller` command in the following format to delete the node and the node controller from the cluster database.

   `cm controller delete -c` *`name`*`[,`*`name`*`] [--delete-nodes]`

   The parameters are as follows:

   - For *name*, enter the hostname of the controller. Wildcard characters are accepted.
   - Specify `--delete-nodes` if the nodes associated with the controller are still in the cluster database. If the nodes have already been deleted from the cluster database, there is no reason to specify `--delete-nodes`.

Changing console management settings

If you have hundreds of compute nodes connected to the system, console logging and the number of active IPMI processes can affect performance.

To avoid excessive console logging and `ipmitool` processing, you can suppress console logging and reduce the number of active IPMI processes.

The following `cadmin` command parameters control console logging and the number of active IPMI processes:

- Console logging. Console logging is enabled by default.

  The following parameters affect console logging:

  - On the `cm node show` command, use the `--conserver-logging` parameter to display logging settings.
  - On the `cm node set` command, use the `--conserver-logging` parameter. You can set this value on a global basis or on a per-node basis. This setting affects ICE compute nodes tied to leaders.

- Console on demand. Console on demand is disabled by default.

  This feature allows IPMI to connect to the node controller to access the console when there is an active console session through the `console` command. For this feature to work, console logging must be enabled.

  The following parameters affect the console on-demand setting:

  - On the `cm node show` command, use the `--conserver-ondemand` parameter to display settings.
  - On the `cm node set` command, use the `--conserver-ondemand` parameter to set the value.

Managing UDP multicast (UDPcast) provisioning

The cluster manager supports UDP multicast provisioning, which allows you to quickly install hundreds of compute nodes at once. UDPcast allows many nodes to join a multicast stream of the content being transported. With all the nodes sharing a single stream, the network is protected from being saturated by disjoint installations. Regardless of cluster type or node type, UDPcast is the default transport method for provisioning.

The following topics explain UDPcast provisioning:

- **UDPcast overview**
- **UDPcast configuration tuning**

For more information about various transport methods, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

UDPcast overview

UDPcast is the basic tool used for multicast installation. It has two primary commands:

- `udp-sender`. Sends a single image stream to one or more receivers.
- `udp-receiver`. Issued by the recipients to listen to the stream.

The following is additional information about UDPcast:

- Flamethrower

Flamethrower is a wrapper program. The cluster manager uses Flamethrower to manage UDPcast content when installing systems and pushing images.

It maps `udp-sender` commands to content to be transported. It starts a `udp-sender` on a unique port for each component to be transported. When `udp-sender` terminates (due to a transfer being complete), Flamethrower starts a new one.

The content managed by Flamethrower includes the Flamethrower directory itself, the system imager boot environment, and any available images. For each image, there are two components: the image itself and the overrides associated with the image.

On a system with three images, there are typically 10 different pieces of content to manage, each with a dedicated `udp-sender` process running on a unique port.

On the admin node, `udp-sender` is run in tar-pipe mode, which means the image is run through tar through a pipe. Separate tar files for each image do not need to be maintained. What is being transported is always the current image.

- Flamethrower directory

  All of the content managed by Flamethrower is listed in the Flamethrower directory. The directory contains a module file for each piece of content that is to be sourced by Bash.

  When a node is interested in multicast content, it first uses `udp-receiver` to transfer the Flamethrower directory. Once the node has the directory, it has the list of components to transport and the port numbers to use. It then uses `udp-receiver` to transfer the desired content.

- Management Ethernet

  The management Ethernet switches must be configured to properly handle multicast traffic. Switches that are supported and configured by the cluster manager are likely to be configured correctly automatically. Switches that are not configured by the cluster manager must be configured to transport multicast traffic.

  The multicast IP addresses are adjustable for the RDV address (the address used for nodes to find each other). The data transport IP addresses are not configurable. The admin node uses 239.0.0.1 by default for RDV, which often requires special switch configuration to work properly. The leader nodes serve the ICE compute nodes. The leader nodes use 224.0.0.1 for RDV by default.

  For more information about these IP addresses and configuration adjustments, see the following:

  UDPcast configuration tuning

- Node memory used for compute and leader nodes

  Compute nodes and leader nodes installed using UDPcast must have enough system memory to hold the image. The image is stored in to a `tmpfs` file system on the node during installation to make the transport more efficient. With hundreds of nodes listening to a stream, writing the data directly to disk would slow down the transfer for all nodes. For this reason, the data is saved to `tmpfs` first and then expanded onto the system disk. If you have nodes with little memory, UDPcast installation could fail for this reason.

- Node memory used for ICE compute nodes in `tmpfs` mode

  The UDP receiver is used in tar-pipe mode. That is, the files are expanded from a pipe directly to the `tmpfs` file system. The `tmpfs` file system is used as the root file system.

UDPcast configuration tuning

This topic describes settings you can fine-tune to optimize UDPcast performance. The goal is to get most nodes to listen to a stream at the same time. Various settings affect the wait time for neighbors to join. It is acceptable for nodes to join different streams. The UDP receiver waits for the current stream to complete and joins when a new stream starts. In this case, some nodes can grab the first stream and other nodes can join the second. You can tune the following attributes:

- `flamethrower-directory-portbase`

The `flamethrower_directory_portbase` attribute is the port number for the Flamethrower directory itself. This directory is important because all nodes need access to the Flamethrower directory to find the appropriate port number for pertinent content. This port number is provided as a kernel parameter for compute (service) and leader nodes when using the UDPcast transport as well as ICE compute nodes when in `tmpfs` mode. The default is 9000.

NOTE: A technical support representative can help you using the `cattr` command to adjust the value if necessary.

- `udpcast-min-receivers`

  This attribute defines the minimum number of receivers that must be present before a UDP sender can start a stream. The admin node uses this value when it serves compute and leader nodes. The leader nodes that serve ICE compute nodes use this global value in `tmpfs` mode. You can use the `cadmin` command to change this value.

- `udpcast-min-wait`

  The `udpcast-min-wait` attribute defines the minimum time that the UDP sender waits before starting a given stream. The UDP sender waits the minimum time for `udpcast-min-receivers` receivers (described earlier) to join the stream. The admin node uses this value when it serves compute and leader nodes. The leader nodes that serve ICE compute nodes use this global value in `tmpfs` mode.

- `udpcast-max-wait`

  The `udpcast-max-wait` attribute defines the maximum time a UDP sender waits before starting a stream. If the minimum number of receivers have not joined by this time, the stream starts anyway. The admin node uses this value when it serves compute and leader nodes. The leader nodes that serve ICE compute nodes use this global value in `tmpfs` mode.

- `udpcast-max-bitrate`

  The `udpcast-max-bitrate` attribute defines the stream bit rate that a UDP sender attempts to achieve. If the bit rate is too fast, the result is an excessive number of retransmits and retries. The default is 900m. The admin node uses this value when it serves compute and leader nodes. The leader nodes that serve ICE compute nodes use this global value in `tmpfs` mode.

- `udpcast-mcast-rdv-addr`

  The `udpcast-mcast-rdv-addr` attribute is an IP address. Senders and receivers use this IP address to find each other (rendezvous).

  This setting affects switch configuration. If the cluster includes switches that were not configured by cluster manager tools, ensure the following:

  - Multicast traffic must be properly routed inside the switches.
  - Multicast traffic must be properly routed between the spine switches and the leaf switches.

  The default RDV addresses are as follows:

  - 239.0.0.1. The admin node, compute nodes, and leader nodes use this address when pushing images for the first time. This address is used because 224.0.0.1 does not cross switch VLANs.
  - 224.0.0.1. Leader nodes that serve ICE compute nodes in `tmpfs` boot mode use this address, which is the default. The default is suitable in this case because VLAN crossing is not necessary.

NOTE: If you adjust the `udpcast-mcast-rdv-addr` value, you might need to adjust the `udpcast-rexmit-hello-interval` attribute.

The `udpcast-mcast-rdv-addr` value takes effect on the leader nodes after the `cimage` command pushes (or repushes) files from the admin node. The image push process reconfigures Flamethrower and the node boot files on leader nodes.

The udpcast-mcast-rdv-addr value resides in the network boot files of nodes that are being booted or installed with UDPcast. The udpcast-mcast-rdv-addr value on the nodes must match the value on the server. To adjust this value, use the cadmin command.

- udpcast-rexmit-hello-interval

  The udpcast-rexmit-hello-interval attribute defines how often a UDP sender process sends a hello packet. This value is especially important when the RDV address is not 224.0.0.1 Remember that the admin node, for example, defaults to 239.0.0.1 for UDP sender processes.

  When a UDP receiver process starts for an RDV address other than 224.0.0.1, the operating system sends an IGMP packet. The Ethernet switch detects this packet. The Ethernet switch then updates its tables with this information. This action allows the multicast packets to properly route through the switch. A problem can arise if the UDP receiver sends its connection packet before the switch updates the switch routing. In this case, the UDP receiver waits forever for a UDPcast stream.

  When you set a udpcast-rexmit-hello-interval value, the UDP sender sends a hello packet at regular intervals and UDP receivers respond to it. In this way, if the UDP receiver missed the initial packet, the UDP receiver sends a fresh request after seeing the hello packet from the UDP sender.

  By default, for admin node UDP senders, this value is 5000 (5 seconds). By default, on leader nodes, this value is 0 (disabled). On leader nodes, this value typically does not need to be set. The RDV address is 224.0.0.1, and there are no VLANs being crossed. If you change the RDV address used by leader nodes, also adjust the udpcast-rexmit-hello-interval value. To adjust this value, use the cadmin command.

NOTE: If you adjust the UDPcast settings, push the new images to the ICE leader nodes. This action ensures the following:

- Correct configuration of the Flamethrower utility on the leader nodes that serve ICE compute nodes, and launching of the needed UDP sender processes on the designated ports.
- Correct configuration information for the ICE compute node tmpfs network boot files.

For more information, see the following:

- The help output for individual commands. For example, enter one or more of the following commands to obtain more information about how to modify UDPcast:
  - cm node set -h
  - cm node unset -h
  - cm node show -h
- The cadmin(1) manpage.
- The udp-sender(1) manpage.
- The cattr(1) manpage.

Console management

The cluster manager uses the open-source console management package called conserver. The conserver package allows all consoles to be accessed from the admin node and facilitates the following functions:

- Manage the console devices of all managed nodes in a cluster
- Console logging

Console management differs depending on the type of cluster.

Console management for clusters without leader nodes or for clusters with ICE leader nodes

A `conserver` daemon runs on the admin node and the leader nodes. The admin node manages leader node and compute node consoles. The leader nodes manage ICE compute blade consoles. The `conserver` daemon uses `ipmitool` to connect to the consoles. Users connect to the daemon to access them. Multiple users can connect, but nonprimary users are read-only.

The `/etc/conserver.cf` file is the configuration file for the `conserver` daemon.

For both the admin node and the ICE leader nodes, the `generate-conserver-files` script generates `/etc/conserver.cf` for the racks and for the admin node. The `discover-rack` command calls the `generate-conserver-files` script as part of rack discovery or rediscovery. The script resides in the following location:

```
/opt/sgi/lib/generate-conserver-files
```

Console management for clusters with scalable unit (SU) leader nodes

The following information pertains to console management on clusters with SU leader nodes:

- The `conserver` file includes entries for the compute nodes. These entries reside in the following file:

  ```
  /etc/conserver.cf
  ```

  Each SU leader node has an IP address alias. Within the `conserver.cf` file, entries are based on the SU leader alias IP address. As compute node entries are added to `/etc/conserver.cf`, they are also added to the admin node `conserver` entries.
- Cluster manager console management includes support for failover and failback for SU leader nodes. If an SU leader node goes down, the IP address alias for the downed SU leader moves to another SU leader. At the same time, the cluster manager writes a relevant `conserver` entry for the new SU leader to the following:
  - The `conserver.cf` file
  - The proxy `conserver` entry on the admin node
- The following command opens a console to one of the compute nodes:

  ```
  cm node console -n node
  ```

  The preceding command reads the proxy SU leader name and connects. Then, the command contacts the `conserver` running on that SU leader and opens the console of the compute node.
- The console log files can be found on the admin node and the SU leader nodes at the following location:

  ```
  /var/log/consoles
  ```

  The preceding directory contains console log files for service nodes, SU leader nodes, and compute nodes that were booted by SU leader nodes.

Additional console management information

For information about `conserver`, see the following:

- `cm node console`, which connects to a node console
- `cm node set --console-timestamp`, which enables time stamps on a global basis.
- `console`(1), a console server client program.
- `conserver`(8), the console server daemon.
- `conserver.cf`(5), the console configuration file for `conserver`.
- `conserver.passwd`(5), user access information for `conserver`.
- **http://www.conserver.com/**

Starting `conserver` on a cluster without leader nodes or on an HPE SGI 8600 cluster

Procedure

1. Connect to the service console and enter the appropriate login when prompted.

   For example:

   ```
   admin:~ # cm node console -n n0
   ```

2. (Conditional) Connect to the ICE leader node console and enter the appropriate login when prompted.

   Complete this step on clusters with ICE leader nodes.

   For example:

   ```
   admin:~ # cm node console -n r1lead
   ```

3. Enter the following to activate system request commands (`sysrq`):

   ```
   Ctrl-e c l 1 8                    # sets log level to 8
   Ctrl-e c l 1 <sysrq cmd>          # sends sysrq command
   ```

4. Enter the following to display a list of `conserver` escape keys:

   ```
   Ctrl-e c ?
   ```

Accessing console log files on a cluster without leader nodes or on an HPE SGI 8600 cluster

The console log files can be found on the admin nodes and the leader nodes at the following location:

```
/var/log/consoles
```

The preceding directory contains console log files as follows:

- On the admin node, the preceding directory contains console log files for each leader node and compute node. The name of each console log file corresponds to the hostname of the server console that is logged.
- On the leader nodes, the directory contains console log files for ICE compute nodes.

An `autofs` configuration file lets you access leader-node-managed console log files from the admin node.

The following procedure enables access to all the ICE compute console log files from the admin node.

NOTE: The following procedure is not supported on clusters with scalable unit (SU) leader nodes.

Procedure

1. Enable the `autofs` service:

   ```
   admin # systemctl enable autofs
   Created symlink from /etc/systemd/system/multi-user.target.wants/autofs.service to
   /usr/lib/systemd/system/autofs.service.
   ```

2. Start the `autofs` service:

   ```
   admin # systemctl start autofs
   ```

3. Change to the directory where the log files reside:

   ```
   admin # cd /net/r1lead/var/log/consoles/
   ```

4. List the log files:

   ```
   r1i1n0  r1i1n2
   ```

Starting `conserver` on clusters with scalable unit (SU) leader nodes

Procedure

1. Connect to the service console and enter the appropriate login when prompted.

   For example:

   ```
   admin:~ # cm node console -n n0
   ```

2. Connect to the SU leader node console and enter the appropriate login when prompted.

   For example:

   ```
   admin:~ # cm node console -n leader1
   ```

3. Connect to one of the compute nodes and enter the appropriate login when prompted.

   For example:

   ```
   admin:~ # cm node console -n n5000
   ```

4. Enter the following to trigger system request commands `sysrq` (after you connect to a console):

   ```
   Ctrl-e c l 1 8                        # sets log level to 8
   Ctrl-e c l 1 <sysrq cmd>              # sends sysrq command
   ```

5. Enter the following to display a list of `conserver` escape keys:

   ```
   Ctrl-e c ?
   ```

Enabling per-line `conserver` time stamps

The following procedure explains how to configure the `conserver` to generate per-line time stamps on all cluster nodes.

Procedure

1. Log into the admin node as the root user.

2. Enter the following command:

   ```
   # cm node set --console-timestamp enabled
   ```

Booting leader nodes or compute nodes from a local disk

By default, compute nodes (including leaders) boot over the network using the GRUB 2 bootloader and miniroot from the admin node.

The cluster manager allows you to boot a compute node from a local disk. For example, you can use this feature in the following situations:

- The node has local images with kernels that are not registered on the admin node.
- You want to boot the node independently from the admin node.

Unless you have a compelling reason to do otherwise, Hewlett Packard Enterprise recommends that you use the default boot mode for compute nodes. There are multiple reasons for this recommendation. For one, regardless of the boot-from-disk mode, the cluster manager does not provide any kernel parameter management capability. You must boot as you would with a standalone system. Secondly, the boot-from-local-disk feature is not supported on MD RAIDs. The boot-from-local-disk feature is supported for the following disk configurations:

- Physical disks.
- Hardware-defined RAIDs.
- BIOS SW RAIDs where BIOS can find the boot sector storing the GRUB 2 boot loader.

The cluster manager supports disjoint boot and admin-assisted boot for booting from a local disk. The following topics describe these modes:

- **Disjoint boot mode**
- **Enabling admin-assisted boot mode**

For a general description of the default boot process for compute nodes, see the following:

Booting a leader node or a compute node

Disjoint boot mode

In a disjoint boot, the node boots without retrieving the bootloader or the miniroot from the admin node. This boot could be useful if the network or the admin node is down. By default, the node uses the on-disk GRUB 2 bootloader and boots the most recently installed slot. In an environment with multiple root slots, a menu lets you choose a different slot from the console.

To select disjoint boot mode, adjust the boot order in BIOS to select booting from a disk. If you reinstall the node, change the BIOS boot order back to select booting from the network. On some platforms, you can use the `ipmitool` command `chassis bootdev pxe`. On UEFI platforms, you can use the `efibootmgr` command.

Enabling admin-assisted boot mode

In admin-assisted boot mode, the cluster manager assumes that the node is using the GRUB 2 bootloader, not `ipxe-direct`. The cluster manager does not send or use the miniroot. The bootloader chain loads to the appropriate boot partition, and then it instructs the node to boot from disk.

If a node has been marked for installation, however, the cluster manager supersedes this boot mode with its normal over-the-network boot operation. On noninstallation boots, however, the cluster manager honors the boot mode specification.

Procedure

1. Log into the admin node as the root user.
2. (Optional) Display the current value of the disk bootloader:

   ```
   cm node show --disk-bootloader -n node
   ```

   For *node*, specify one or more node hostnames.
3. Edit the cluster definition file, and add the following attribute to the target nodes:

   ```
   disk_bootloader=yes
   ```

   The cluster definition file can reside on the cluster in any directory, under any name.

   To retrieve a copy of the cluster definition file, enter the following command:

   ```
   # discover --show-configfile > filename
   ```

   The actions in this step ensure that this feature is configured if the cluster manager is reinstalled at a later date.
4. Use the `cm node set` command to specify admin-assisted boot mode:

   ```
   cm node set --disk-bootloader --dhcp-bootfile grub2 -n node
   ```

   For *node*, specify the node hostname.

   At a later date, to disable admin-assisted boot mode, edit the cluster definition file again, and set `disk_bootloader=no`. Then run the following command:

   ```
   cm node unset --disk-bootloader --dhcp-bootfile grub2 -n node
   ```

Changing the size of `/tmp` on ICE compute nodes

The following procedure explains how to change the size of `/tmp` on ICE compute nodes.

Procedure

1. From the admin node, use the `cd` command to change to the following directory:

   ```
   /opt/sgi/share/per-host-customization/global
   ```
2. Open the `sgi-fstab.sh` file.
3. Change the `size=` parameter for the `/tmp` mount in both locations that it appears.
4. Push the image out to the racks to pick up the change.

   For example:

   ```
   # cimage --push-rack --customizations-only sgi-fstab.sh image_name "r*"
   ```

   For more information about using the `cimage` command, see the following:

   Managing ICE compute node images

Changing the root file system to `tmpfs` on ICE compute nodes

This procedure assumes that all the nodes are configured in the cluster manager. That is, the nodes are recognized in the cluster manager database. ICE compute nodes are configured with the NFS root file system.

NOTE: Do not change the root file system on any leader nodes.

Procedure

1. Log into the admin node as the root user.
2. Use the `cimage` command in the following format to specify that the nodes use the `tmpfs` file system:

   ```
   cimage --set --tmpfs image kernel 'nodes'
   ```

   The variables are as follows:

Variable	Specification
image	The image name. For example, `ice-rhel8.X`.
kernel	The kernel number. For example, `3.10.0-693.el7.x86_64`.
nodes	The hostnames of all of the ICE compute nodes in the rack. For example, `'r1i*n*'`.

3. Use the `cimage` command in the following format to propagate the new root file system image to all the nodes:

   ```
   cimage --push-rack image 'rack_name'
   ```

   The variables are as follows:

Variable	Specification
image	The image name. For example, `ice-rhel8.X`.
rack_name	`r` and the rack number. The syntax supports globbing. For example, specify one of the following: • `'r1'` • `'r*'` or `'r?'`

4. Power-down the nodes.

   Enter one of the following commands:

   - To power-down the nodes gracefully, use the following command:

     ```
     cm power shutdown -t node 'nodes'
     ```

   - To power-down the nodes abruptly, use the following command:

     ```
     cm power off -t node 'nodes'
     ```

5. Use the following command to power-up the nodes:

   ```
   cm power on -t node nodes
   ```

   For *nodes*, specify the node hostnames.

For example:

```
# cm power on -t node 'r1i*n*'
```

Changing the root file system to `tmpfs` on compute nodes

This procedure assumes that all the nodes are configured in the cluster manager. That is, the nodes are recognized in the cluster manager database. Compute nodes are configured with the disk file system.

NOTE: Do not change the root file system on any leader nodes.

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node set` command in the following format to change the configured nodes to use the new root file system:

   ```
   cm node set --rootfs tmpfs -n nodes
   ```

   For *nodes*, specify one or more node hostnames.
3. Power-down the nodes.

   Enter one of the following commands:

   - To power-down the nodes gracefully, use the following command:

     ```
     cm power shutdown -n nodes
     ```

   - To power-down the nodes abruptly, use the following command:

     ```
     cm power off -n nodes
     ```

4. Use the following command to power-up the nodes:

   ```
   cm power on -n nodes
   ```

   For example:

   ```
   # cm power on -n r1i1n0
   ```

5. (Conditional) Enable newly added nodes to be configured into the system with an NFS file system.

   Complete the following steps if the cluster is an HPE Cray EX cluster or an HPE Apollo 9000 cluster.

   a. Open the following file in a text editor:

      ```
      /opt/clmgr/etc/cmcinventory.conf
      ```

   b. Edit the `rootfs=` and `transport=` lines as follows to specify that new nodes added to the system be configured with an NFS file system:

      ```
      rootfs=tmpfs
      transport=udpcast
      ```

c. Save and close /opt/clmgr/etc/cmcinventory.conf.

d. Enter the following command to restart the cmcinventory service:

```
# systemctl restart cmcinventory.service
```

Configuring local storage space for swap and scratch disk space on an HPE SGI 8600 cluster

You can configure a cluster with ICE leader nodes to support local storage space on ICE compute nodes. The nodes are also known as **blades**. Solid-state drive (SSD) devices and 2.5" disks are available for this purpose.

HPE supports a set of parameters that you can use to configure partitions on your system. You can define the size and status for both swap and scratch partitions.

You can set the partition values on a global basis or on an individual basis. If you set a value on a global basis, the value applies to all ICE compute nodes. You can also set the value to apply to only one node.

By default, the disks are partitioned only if blank. Swap is off. Scratch is set to occupy the whole disk space. Scratch is mounted at /tmp/scratch.

You can use the cattr command to retrieve the status of a setting, to enable a setting, or to disable a setting. If you do not set any parameters, the system uses the defaults.

The cluster manager /etc/init.d/set-swap-scratch script configures the swap and scratch space based on the settings you specify with the cattr command.

The following list explains the local storage space settings:

blade_disk_allow_partitioning

Determines whether you can repartition and reformat the local storage disk. Specify on or off. Default is on.

To protect user data, the cluster manager prevents you from repartitioning a disk that is already partitioned. In this case, you need a blank disk to use for the swap and scratch partitions.

blade_disk_raid_level

Specifies whether you can enable RAID0 (striping) or RAID1 (mirroring) when you have two disks for swap and scratch. The values are as follows:

off

Does not enable RAID. Default.

0

Enables RAID0 (striping) for the swap and scratch partitions.

If you use blade_disk_scratch_size or blade_disk_swap_size to specify the partition size in megabytes, the cluster manager creates a partition of the specified size on each individual disk. This means that in the case of RAID 0, the actual space created is double the size. If a single disk, or RAID 1, is used, the space available is equal to the specified value.

1

Enables RAID1 (mirroring) for the swap and scratch partitions.

blade_disk_reformat_scratch_at_boot

Specifies whether you are allowed to format the scratch partition every time the ICE compute node boots. The values are as follows:

`off`

Prevents formatting of the scratch partition at boot. Default.

0

Enables formatting of the scratch partition every time the ICE compute node boots.

`blade_disk_reformat_swap_at_boot`

Specifies whether you are allowed to format the swap partition every time the ICE compute node boots. The values are as follows:

`off`

Prevents formatting of the swap partition at boot. Default.

0

Enables formatting of the swap partition every time the ICE compute node boots.

`blade_disk_scratch_mount_point`

Specifies the mount point for the scratch partition. Default is `/tmp/scratch`.

You can mount the disk to any mount point. If it does not exist, the cluster manager creates the mount point directory. The cluster manager needs permission to create the mount point at the mount point you specify. On the ICE compute nodes, the root mount point (/) is not writable. If you want to mount to `/scratch`, make sure to create that folder as part of the ICE compute node image.

`blade_disk_scratch_size`

Specifies the scratch size. Specify either an integer number of megabytes or one of the special values, as follows:

- When you specify the partition size in megabytes, the system creates a partition of that size on each individual disk. In the case of RAID 0, the actual size is double the size you specified. In the case of a single disk or RAID1, the space available is equal to the specified value.
- The special values are `-0` and `0`. When you use these values, the outcomes are as follows:
  - `-0`

    Uses all free space for scratch when partitioning. Default.
  - `0`

    Does not create a scratch partition on the local storage disk. Prevents the cluster manager from creating a scratch partition.

`blade_disk_scratch_status`

Determines whether the cluster manager creates a scratch partition on the local storage disk. Specify `on` or `off`. Default is `off`, which means that the cluster manager does not create a scratch partition.

The cluster manager assigns the label `SGI_SCRATCH` when it partitions the disk. It mounts the scratch on the partition labeled `SGI_SCRATCH`.

`blade_disk_swap_size`

Specifies the swap size. Specify either an integer number of megabytes or one of the special values, as follows:

- When you specify the partition size in megabytes, the system creates a partition of that size on each individual disk. In the case of RAID 0, the actual size is double the size you specified. In the case of a single disk or RAID1, the space available is equal to the specified value.
- The special values are `-0` and `0`. When you use these values, the outcomes are as follows:
  - `-0`

    Uses all free space for swap when partitioning. Default.
  - `0`

    Does not create a swap partition on the local storage disk. Prevents the cluster manager from creating a swap partition.

`blade_disk_swap_status`

Determines whether the cluster manager creates a swap partition on the local storage disk. Specify `on` or `off`. Default is `off`, which means that the cluster manager does not create a swap partition.

The cluster manager assigns the label `SGI_SWAP` when it partitions the disk. It enables the swap only if an `SGI_SWAP` label exists.

The following topics show the `cattr` commands you can use to configure the swap and scratch disk space:

- **Retrieving the status of a local storage space setting**
- **Enabling, disabling, or respecifying a local storage space setting on an HPE SGI 8600 cluster**

Retrieving the status of a local storage space setting

The following procedure explains how to display the status of a local storage space setting.

Procedure

1. Log into the admin node as the root user.
2. Enter the `cattr get` command, in the following format, to retrieve the current setting:

   ```
   cattr get setting [-N node] --default default
   ```

   The variables are as follows:

Variable	Specification
setting	One of the local storage space settings. For the list of settings, see the following: **Configuring local storage space for swap and scratch disk space on an HPE SGI 8600 cluster**
node	One ICE compute node hostname. Specify this argument only if you want to set one of the local storage space settings for an individual ICE compute node.
default	The default value for this setting.

Example 1. The following command returns on, which indicates that the setting is enabled and applies to all ICE compute nodes:

```
# cattr get blade_disk_allow_partitioning --default on
on
```

Example 2. Assume that you set the blade_disk_scratch_size to 2 megabytes. To retrieve the current scratch size, enter the following command:

```
# cattr get blade_disk_scratch_size --default -0
2
```

Enabling, disabling, or respecifying a local storage space setting on an HPE SGI 8600 cluster

The following procedure explains how to modify a local storage space setting.

Procedure

1. Log into the admin node as the root user.
2. Enter the cattr set command, in the following format, to enable, disable, or specify a value for a local storage space setting:

   ```
   cattr set [-N node] setting value
   ```

   The variables are as follows:

Variable	Specification
node	One ICE compute node hostname. Specify this argument only if you want to set one of the local storage space settings for an individual ICE compute node.
setting	One of the local storage space settings.
value	on, off, an integer value that represents megabytes, or a mount point. For information about possible values, see the individual setting information in the following topic: **Configuring local storage space for swap and scratch disk space on an HPE SGI 8600 cluster**

Example 1. The following command turns on the blade_disk_allow_partitioning setting for all ICE compute nodes:

```
# cattr set blade_disk_allow_partitioning on
```

Example 2. The following command turns on blade_disk_allow_partitioning for ICE compute node r1i0n0:

```
# cattr set -N r1i0n0 blade_disk_allow_partitioning on
```

Example 3. The following command sets the scratch partition mount point for the local disk associated with ICE compute node r1i0n0 to /tmp/scratch22:

```
# cattr set -N r1i0n0 blade_disk_mount_point /tmp/scratch22
```

Example 4. The following command enables the scratch space feature:

```
# cattr set -N r1i0n0 blade_disk_scratch_status on
```

Example 5. In this example, the cattr command directs the cluster manager to allocate 8192 megabytes of disk space as scratch space. In the case of RAID 0, the actual space available is double because both disks are combined in RAID 0.

```
# cattr set -N r1i0n0 blade_disk_scratch_size 8192
```

3. Use the cimage command to push the changes out to the desired nodes:

```
# cimage --push-rack image_name racks
```

Using the cattr command to modify system attributes

You can use the cattr command to assign attributes to cluster nodes. You can assign attributes either on a global basis, to the entire system, or on an individual node basis.

The cattr command can retrieve attribute settings, set attributes, remove attributes, and perform other functions. Enter the following to retrieve a cattr command help statement and the list of attributes you can manipulate:

```
# cattr -h
```

NOTE: When possible, use the cm node set command or the cadmin command, rather than the cattr command, to modify system attributes. When you use the cm node set command or the cadmin command, the cluster manager regenerates the configuration and eliminates the need for you to issue a cm node update config command. To make custom configuration changes that you want preserved across cm node update config calls, see the following:

Preserving custom configuration changes

For information about how to modify local storage space attributes, see the following:

Configuring local storage space for swap and scratch disk space on an HPE SGI 8600 cluster

Configuring quotas on admin nodes and scalable unit (SU) leader nodes

NOTE: The information in this topic applies to clusters with SU leader nodes and to clusters without leader nodes. This information does not pertain to clusters with ICE leader nodes.

The admin node and SU leader nodes are diskful nodes. By default, these nodes have XFS root file systems. Also by default, the cluster manager implements disk quotas to lessen the risk of the root file system becoming full.

On admin nodes, the quotas are set to 50% by default. On SU leader nodes, the quotas are set to 80% by default. Areas that are affected by these quotas include the following:

- /var/crash
- /var/lib/kafka
- /var/lib/elasticsearch
- /var/log
- /opt/clmgr/postgresql/var/lib/pgsql/*postgres_major_version*/data/

For example, in the HPE Performance Cluster Manager 1.8 release, the path is as follows:

```
/opt/clmgr/postgresql/var/lib/pgsql/14/data/
```

For more information, see the following:

- `/etc/opt/sgi/conf.d/80-xfs-prjquota`
- The `xfs_quota` manpage
- The `conf.d` script that enables quotas

Managing the writable disk space used by compute nodes with writable NFS file systems

NOTE: The information in this topic applies to clusters with SU leader nodes and to clusters without leader nodes. This information does not pertain to clusters with ICE leader nodes.

The default maximum size of the writable image file on each SU leader node is 500 megabytes. Because the overlay solution and the overmount solution are not compatible, a separate image name is used for the overlay solution and the overmount solution. You can change the global default writable image size, and you can also change the per-image size.

On a global basis, use the following commands to manage this size:

- `cm image set --perhost-size` *size*
- `cm image unset --perhost-size`
- `cadmin --show-image-perhost-size`

On a per-image basis, use the following commands to manage this size:

- `cm image set --perhost-size` *size* `-i` *image*
- `cm image unset --perhost-size -i` *image*
- `cadmin --show-image-perhost-size --image` *image_name*

NOTE: If you increase the size, the SU leader node automatically expands to the larger space on the next boot. However, the act of reducing the size is destructive. If you reduce the size, delete the per-host writable areas. For more information, enter the following command:

```
# cm image activate -h
```

Disk quotas on an HPE SGI 8600 cluster

Within the compute image for an ICE leader node, the cluster manager sets default per-directory disk **quotas**, which can also be called **project quotas**. The quota mechanism prevents a disk from filling up and inhibiting a node from booting.

Soft quotas and hard quotas apply to any entity that writes to disk. For example, quotas pertain to a user writing to disk actively or a user job that writes to disk.

Quotas prevent an ICE compute node from accidentally filling the disk space of the associated leader node over the network file system (NFS). Quotas apply when a ICE compute node is booted with NFS root directories, not `tmpfs` directories.

The cluster manager sets default quota settings in each software image, rather than in each node. You can adjust these quota settings at your site. The soft quotas and hard quotas are as follows:

- A soft quota is an initial limit. After an ICE compute node exceeds a soft quota, the ICE compute node can continue to use resources up until it reaches the upper hard limit.
- A hard quota is a firm limit.

The default quotas are as follows:

- Soft quota = 2048 minutes
- Hard quota = 2148 minutes
- Quota timer = 1 day

The ICE compute node might fail to boot properly in the following cases:

- If a hard quota is exceeded
- If a soft quota is exceeded past the time set in the timer

An ICE cluster prevents additional writes to a disk when either of the following events occur:

- A disk reaches its hard limit.
- A disk reaches its soft limit and the timer has expired.

The following topics provide more information about quotas:

- **Retrieving quota information on an HPE SGI 8600 cluster**
- **Setting quotas on an HPE SGI 8600 cluster**
- **Viewing the ICE compute node read/write quotas on an HPE SGI 8600 cluster**

Retrieving quota information on an HPE SGI 8600 cluster

The following procedure explains how to retrieve quota values for a specific image.

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to retrieve a list of the images on the system:

```
# cm image show
ice-rhel8.X
ice-rhel8.X-kdump
ice-sles15spX-mofed
lead-rhel8.X
lead-rhel8.X-ha
rhel8.X
sles15spX-mofed
```

The example output shows the following ICE compute node images:

- `ice-rhel8.X`
- `ice-rhel8.X-kdump`
- `ice-sles15spX-mofed`

3. Enter one of the following commands to retrieve information about one of the quotas or the quota timer:

   ```
   cadmin --show-soft-quota --image image_name
   cadmin --show-hard-quota --image image_name
   cadmin --show-quota-timer --image image_name
   ```

   For *image_name*, enter one of the names from the `Image Name` column in the previous step.

   For example:

   ```
   # cadmin --show-soft-quota --image ice-rhel8.X
   2048m
   # cadmin --show-hard-quota --image ice-rhel8.X
   2148m
   # cadmin --show-quota-timer --image ice-rhel8.X
   1d
   ```

   The `cadmin` command output displays the quotas using the format of the underlying tool, which is the XFS file system project quota infrastructure. For information about the format, see the `xfs_quota`(8) manpage.

4. (Optional) Set site-specific quotas.

   Proceed to the following:

   Setting quotas on an HPE SGI 8600 cluster

Setting quotas on an HPE SGI 8600 cluster

The following procedure explains how to change a quota or the quota timer.

Procedure

1. Log into the admin node as the root user.

2. Verify the current value for the quota setting you want to change.

   For information about how to verify quota settings, see the following:

   Retrieving quota information on an HPE SGI 8600 cluster

3. Modify the quota setting.

   - To set a site-specific *value*, use one of the following commands:

     ```
     cm image set -i image_name --soft-quota value
     cm image set -i image_name --hard-quota value
     cm image set -i image_name --quota-timer value
     ```

     The variables are as follows:

Variable	Specification
image_name	One of the image names in the output from the `cm image show` command.
value	An integer value followed by a unit specification. For `--soft-quota` or `--hard-quota` operations, specify the following: ◦ `k` for kilobytes ◦ `m` for megabytes ◦ `g` for gigabytes ◦ `t` for terabytes For the `--quota-timer` operation, specify the following: ◦ `m` for minutes ◦ `d` for days ◦ `h` for hours ◦ `w` for weeks

The following examples specify site-specific values for the quotas associated with the `ice-sles15spX` compute image:

```
# cm image set -i ice-sles15spX --soft-quota 4200m
# cm image set -i ice-sles15spX --hard-quota 4196m
# cm image set -i ice-sles15spX --quota-timer 3d
```

- To set a site-specific value back to the factory default value, use one of the following commands:

  ```
  cm image unset -i image_name --soft-quota
  cm image unset -i image_name --hard-quota
  cm image unset -i image_name --quota-timer
  ```

  The following examples reset site-specific values back to the factory default values:

  ```
  # cm image unset -i ice-sles15spX --soft-quota
  # cm image unset -i ice-sles15spX --hard-quota
  # cm image unset -i ice-sles15spX --quota-timer
  ```

4. Push out the changes to the ICE compute nodes.

For information about how to push changes, see the following:

Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Viewing the ICE compute node read/write quotas on an HPE SGI 8600 cluster

You can retrieve the read quota and the write quota for each ICE compute node.

The following procedure explains how to retrieve current usage.

Procedure

1. Log into the admin node as the root user.

2. Enter the following command to retrieve a list of leader nodes:

   ```
   # cm node show -t system leader
   r1lead
   r2lead
   ```

3. Use the `ssh` command to log into one of the leader nodes.

   ```
   # ssh r1lead
   ```

4. Enter the following command to retrieve a list of projects:

   ```
   # less /etc/projects
   1:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n0
   2:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n1
   3:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n2
   4:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n3
   5:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n4
   6:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n5
   7:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n6
   8:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n7
   9:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n8
   10:/var/lib/sgi/per-host/ice-rhel8.X/1/i0n9
   .
   .
   .
   ```

   The project numbers are the leftmost integers in the output. Enter `q` to exit the `less` command.

5. Use the `xfs_quota` command, in the following format, to retrieve the current usage values:

   ```
   xfs_quota -x -c 'quota -ph project_num'
   ```

   For *project_num*, specify one of the project numbers you retrieved in the preceding step.

   For example:

   ```
   r1lead:~ # xfs_quota -x -c 'quota -ph 1'
   Disk quotas for Project #1 (1)
   Filesystem   Blocks  Quota  Limit Warn/Time    Mounted on
   /dev/disk/by-label/sgiroot
                 64.6M      0     1G  00 [------] /
   ```

Creating custom partitions

By default, the cluster manager provides partition layouts. These layouts allow one or more instances of the cluster manager to be installed on the same root drive or drives. Each instance of the cluster manager is housed in a slot. A **slot** is a disjoint subset of partitions. The cluster manager documentation refers to the default partitioning scheme as **slot partitioning**. The cluster manager supports 1 to 10 slots. By default, there are 2 slots.

If the default partition scheme does not work for your cluster, you can create a custom partitioning scheme.

For more information about the default partition layout and the role of slots, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Custom partitioning notes, constraints, and cautions

Before you implement custom partitioning, consider the following:

- Reserving disk space for partitions

  If you want scratch space reserved on a disk for your partitions, use the disk reservation feature rather than custom partitioning. Use custom partitioning when core operating system partitions are in play. The disk reservation feature and custom partitioning are not compatible.

  For information about the disk reservation feature, see one of the following:

  - **Configuring scratch disk space on an admin node**
  - **Configuring scratch disk space on leader nodes and on compute nodes**

- RAID configurations

  Custom partitioning does not affect RAID configurations: MD RAID, BIOS SW RAID, or single disks (including HW RAIDs). Custom partitioning is activated after any RAIDs are created or assembled. Custom partitioning does not define the RAIDs.

- Applicable cluster nodes

  The cluster manager supports custom partitioning on admin, leader, and compute nodes. However, the manner by which you configure the custom partitioning varies.

  The cluster manager does not support custom partitioning on ICE compute nodes.

- High Availability (HA) configurations

  Custom partitioning is not supported on physical admin nodes in HA configurations.

- Slots

  The slot partitioning scheme of the admin node determines the initial slot partitioning of all other nodes. Custom partitioning does not support slot partitioning (install slots). Hence, custom partitioning of the admin node precludes the use of slots on any nodes in the cluster.

  The use of slots, especially on the admin node, allows you to more smoothly handle cluster manager and Linux distribution updates. Slots provide convenient fallback locations when upgrading. Even if you desire custom partitions on other node types, consider your future upgrade plans before deciding to use custom partitions on the admin node itself.

- Data loss

  If a node is configured for custom partitioning and you reinstall the node, the following occurs:

  - Its partitions are erased.
  - All data on the root drive is lost.

  Similarly, assume the following:

  - A node is configured with custom partitioning.
  - You configure it for default slot partitioning.

  In this case, data loss occurs on the root drive when the node is reinstalled.

- Custom partitioning does not apply to compute nodes configured with an NFS root file system or a `tmpfs` root file system.

Configuring custom partitioning for leader nodes and for compute nodes

Procedure

1. Describe the custom partition layout in a specially formatted configuration file.

   You can choose any file name as long as it begins with an alphabetic character and has a file extension of `.cfg`.

   Write the file to the `/opt/clmgr/image/scripts/pre-install` directory.

   Within the file, create a table. In this table, each row defines a partition. Use a vertical bar (`|`) character to separate the columns. The columns contain the partition specifications. In order, the columns contain the following partition specifications:

   - Partition number.
   - Mount point.
   - Size.
   - File system type. Can be one of the following:
     - XFS, which is the default
     - `ext4`
     - `ext3`
   - File system label.
   - EFI-only option, which is a file system that can only be created and mounted on a UEFI platform.
   - Mount/file system options.
   - A `mkfs` command specification, which includes substitutions for a file system label and the partition device.

   For an example configuration file, see the following:

   `/opt/clmgr/image/scripts/pre-install/custom_partitions.cfg`

   The example file contains comments that describe requirements. For example, the file explains how to order the partitions and explains the constraints on various specifications. The following is a sample partition layout:

   ```
   #part| mount_point | sgdisk_size | fs  | fs_label | efi_only | mount_opts | mkfs_command
   1    |             | 50M         |ext4 | sgidata  | no       |            | mke2fs -F -L %fs_label -j %dev
   2    | swap        | 2048M       |swap | sgiswap  | no       |            | mkswap -f -L %fs_label %dev
   3    |             | 50M         |vfat | ADMINEFI | yes      | defaults   | mkdosfs -I -n %fs_label %dev
   4    | /           | 16384M      |xfs  | sgiroot  | no       | defaults   | mkfs.xfs -f -L %fs_label %dev
   5    | /boot       | 400M        |ext4 | sgiboot  | no       | defaults   | mke2fs -F -L %fs_label -j %dev
   6    | /boot/efi   | 150M        |vfat | SGIEFI   | yes      | defaults   | mkdosfs -I -n %fs_label %dev
   7    | /var        | 32768M      |ext4 | myvar    | no       | defaults   | mke2fs -F -L %fs_label -j %dev
   8    | /scratch    | FILL        |xfs  | scratch  | no       | defaults   | mkfs.xfs -f -L %fs_label %dev
   ```

2. Edit the cluster definition file, and set the `custom_partitions=`*`filename`* attribute.

   The cluster definition file can reside on the cluster in any directory, under any name.

   To retrieve a copy of the cluster definition file, enter the following command:

   ```
   # discover --show-configfile > filename
   ```

   The actions in this step ensure that custom partitioning is configured if the cluster manager is reinstalled at a later date.

NOTE: The order in which you list file systems is important. As in an `fstab` file in Linux, list base mounts before mounts that reside on base mounts. For example, if you plan to have a file system for `/var` and a file system for `/var/log`, list `/var` before `/var/log`.

Many versions of Linux require that the root file system (/) contain `/usr/lib/systemd/system`. For this reason, do not make `/usr` a separate mount point. If you make `/usr` a separate mount point, the node cannot boot properly.

3. Use one of the following methods to set the `custom-partitions` system attribute to the name of this configuration file:

   - Use the `cm node set` command in the following format:

     ```
     # cm node set --custom-partitions filename -n node
     ```

     The variables are as follows:

Variable	Specification
filename	The name of the configuration file you created.
node	One or more node hostnames.

     When the node reboots, the repartitioning takes place.

     Or

   - Use the `cm image set` command in the following format:

     ```
     # cm image set -i image --custom-partitions filename
     ```

     The variables are as follows:

Variable	Specification
image	The image name.
filename	The name of the configuration file you created.

     Any node that is assigned this image receives the custom partitioning scheme at its next reboot. If there is a conflict in the partitioning specification between the node and image, the node specification takes precedence.

Configuring custom partitioning for admin nodes

As with leader and compute nodes, you must describe your custom partition layout in a specially formatted configuration file. However, with admin nodes, you must do the configuring at cluster installation time. To get a template of this configuration file to use during installation, access the file named `custom_partition_example` on the installation media. For help and support, contact your customer support representative.

Managing custom partitions

The following are some management actions you might need to take when you have nodes with custom partitions:

- Display information about an image or node regarding custom partitioning.

Use the following commands:

```
cadmin --show-custom-partitions --image image
```

```
cm node show --custom-partitions -n node
```

The commands return the name of the custom partition configuration file or `Disabled`, which means one of the following:

- For images, `Disabled` means that the cluster manager uses slot partitioning when the image is installed on a node. Slot partitioning is the default.
- For nodes, `Disabled` indicates one of the following:
- The node currently has slot partitioning.
- The node is going to be repartitioned with slot partitioning the next time the node boots.

- Clear the `--unset-custom-partitions` attribute for selected nodes.

  Use the `cm node unset` command in the following format:

  ```
  cm node unset --custom-partitions -n node
  ```

  The command disables custom partitioning for the nodes effective at the next reboot. A node specification takes precedence over a specification associated with an image.

- Clear the `custom-partitions` attribute for an image.

  Use the `cm image unset` command in the following format:

  ```
  cm image unset --custom-partitions -i image
  ```

  The command disables custom partitioning for the image effective at the next reboot. A node specification takes precedence over a specification associated with an image.

Backing up and restoring the cluster database

The cluster manager database is critical to the operation of your cluster. The database includes relevant data for the managed objects in a cluster. Make sure you back up the database on a regular basis.

Managed objects on a cluster include the following:

- The cluster itself

  The whole cluster is a managed object. A cluster with leader nodes is modeled as a meta-cluster. This meta-cluster contains the racks each modeled as a subcluster.

- Nodes

  The admin node, leader nodes, compute nodes, ICE compute nodes (blades), and chassis controllers are modeled as nodes.

- Networks

  The preconfigured and potentially customized IP networks.

- NICs

  The network interfaces for Ethernet and fabric adapters.

- The node images installed on each particular node.

By default, the cluster manager backs up the cluster database automatically every 24 hours, compresses the backup file, and retains the backup file for 365 days. The cluster manager writes the compressed backup files to the following directory:

```
/opt/clmgr/database/backup/
```

The backup files all have a `.xz` suffix.

To disable the backups or to customize the backup schedule, see the following file:

```
/etc/cron.d/backup_cm_db
```

Backing up the cluster database manually

Use the following procedure when you want an immediate backup. For example, if you change the cluster configuration, complete the procedure to back up the new configuration.

Procedure

1. Log into the admin node as the root user.
2. Enter the following commands to stop the cluster manager:

   ```
   # systemctl stop config_manager.service
   # systemctl stop clmgr-power.service
   # systemctl stop cmdb.service
   ```

3. Back up the cluster database:

   ```
   # sqlite3 /opt/clmgr/database/db/cmu.sqlite3 ".backup file"
   ```

   For *file*, specify a name for the backup file.

   For example:

   ```
   # sqlite3 /opt/clmgr/database/db/cmu.sqlite3 ".backup cmu.backup.sqlite3"
   ```

4. Enter the following commands to start the cluster manager:

   ```
   # systemctl start cmdb.service
   # systemctl start clmgr-power.service
   # systemctl start config_manager.service
   ```

5. Write the backup file to another computer system at your site for safekeeping.

   The cluster database is the internal database that hosts information about each cluster component. A copy of the original cluster database can be valuable when performing a disaster recovery. Make sure to take additional, periodic database backups in the future as you modify your system.

Restoring the cluster database

Procedure

1. Log into the admin node as the root user.
2. Enter the following commands to stop the cluster manager:

   ```
   # systemctl stop config_manager.service
   # systemctl stop clmgr-power.service
   # systemctl stop cmdb.service
   ```

3. Locate the backup file.

To restore a backup file that the cluster manager created automatically, find the file in the following directory:

```
/opt/clmgr/database/backup/
```

To restore a backup file that you created manually, locate the file in the directory you used.

4. (Conditional) Expand the backup file.

   Complete this step if the backup file is compressed.

   By default, the cluster manager compresses the daily backup files that it creates. For example, to expand backup file `cmu-backup-1641445201991.sqlite3.xz`, enter the following command:

   ```
   # unxz /opt/clmgr/database/backup/cmu-backup-1641445201991.sqlite3.xz
   ```

5. Enter the following command to restore the cluster database:

   ```
   # cp -i file /opt/clmgr/database/db/cmu.sqlite3
   ```

   For *file*, specify the name of the backup file.

   For example, the following lines show how to restore the database. Answer **y** when prompted to affirm the database overwrite.

   ```
   # cp -i cmu-backup-1641445201991.sqlite3 /opt/clmgr/database/db/cmu.sqlite3
   cp: overwrite '/opt/clmgr/database/db/cmu-backup-1641445201991.sqlite3'? y
   ```

6. Enter the following commands to start the cluster manager:

   ```
   # systemctl start cmdb.service
   # systemctl start clmgr-power.service
   # systemctl start config_manager.service
   ```

Linux shell commands

The cluster manager provides a Linux shell API interface. Most functions provided from the GUI and CLI have their equivalent in the API interface. The API interface is easily called from a shell script.

For information about commands, see one of the following:

- **Manpages**
- The manpages themselves

Specifying descriptive alias names for nodes

The cluster manager enables you to specify descriptive node names to compute nodes or any other kind of node. These descriptive node names are called *aliases* or *node name aliases*. If you work consistently with a group of nodes, you can assign each node one or more customized aliases.

Many cluster manager commands require you to specify node hostnames, and many cluster manager commands accept node alias names in the place of node hostnames. You can specify the node name aliases in the commands that begin with `cm`. Commands that do not begin with `cm` require you to specify the node hostnames.

For example:

- You can specify either node name aliases or hostnames on the `cm node` command line, the `cm image` command line, and other command lines.
- You can specify hostnames, but not node name aliases, on the command lines that do not begin with `cm`. For example, on many legacy SGI Management Suite commands, such as `cinstallman`, you can only specify hostnames.

You can use the following parameters on a `cm node show` command line to specify that you want the cluster manager to display node name aliases in output:

- `-D` *`alias_group`*
- `--convert-to-aliases` *`alias_group`*

Assigning alias names to nodes

You can assign alias names to any kind of node. The steps and examples in this documentation address only how to assign aliases to compute nodes.

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node show` command to display the hostnames for each cluster node.

   For example:

   ```
   # cm node show
   .
   .
   .
   node42
   node43
   node44
   node45
   .
   .
   .
   ```
3. Use the `cm node set` command in the following format to specify aliases for one or more nodes.

   ```
   cm node set --alias-name new_name --alias-group group_name -n node
   ```

   The variables are as follows:

Variable	Specification
new_name	Your customized alias name for this node. This name must be unique across the cluster. You cannot specify the same alias name for two or more nodes.
group_name	A group name for the node. Subsequent commands can add aliases for other nodes to this group.
node	One or more node hostnames.

For example, the following commands specify a node group named `gnodes` and new alias names for the nodes with hostnames `node44` and `node45`:

```
# cm node set --alias-name gpu-node1 --alias-group gnodes -n node44
# cm node set --alias-name gpu-node2 --alias-group gnodes -n node45
```

4. Enter the following command to verify the aliases:

```
# cm node show --alias-names
NODE      ALIAS
node42
node43
node44    gpu-node1
node45    gpu-node2
```

5. (Optional) Create additional alias names for the nodes and display information about the nodes.

   Example 1. Assume that you want to assign names `work1`, `work2`, `work3`, and `work4` to nodes `node42`, `node43`, `node44`, and `node45`. The group name for these new alias names is `wnodes`. The commands are as follows:

```
# cm node set --alias-name work1 --alias-group wnodes -n node42
# cm node set --alias-name work2 --alias-group wnodes -n node43
# cm node set --alias-name work3 --alias-group wnodes -n node44
# cm node set --alias-name work4 --alias-group wnodes -n node45
```

   Example 2. The following command displays the nodes with their hostnames and their new aliases:

```
# cm node show --alias-names
NODE      ALIAS        ALIAS
node42                 work1
node43                 work2
node44    gpu-node1    work3
node45    gpu-node2    work4
```

   Example 3: The following command displays the nodes with their alias names and their group names:

```
# cm node show --alias-names --alias-groups
NODE      GROUP   ALIAS        GROUP   ALIAS
node42                         wnodes  work1
node43                         wnodes  work2
node44    gnodes  gpu-node1    wnodes  work3
node45    gnodes  gpu-node2    wnodes  work4
```

Removing alias names from nodes

Procedure

1. Log into the cluster manager as the root user.

2. Enter the following command to display nodes name aliases:

```
# cm node show --alias-names
NODE      ALIAS        ALIAS
node42                 work1
node43                 work2
node44    gpu-node1    work3
node45    gpu-node2    work4
```

3. Use the following command to clear node name aliases you no longer want:

   ```
   cm node unset --alias-name alias
   ```

   For *alias*, enter one or more node name aliases. If you specify more than one, use a comma to separate alias names.

   For example:

   ```
   # cm node unset --alias-name gpu-node1,gpu-node2
   ```

4. Use the following command to confirm that the node name aliases you specified are no longer associated with any nodes.

   ```
   # cm node show --alias-names
   NODE      ALIAS
   node42    work1
   node43    work2
   node44    work3
   node45    work4
   ```

Using alias names on commands lines and displaying alias names in output

Assume that you specified alias names for the following nodes:

Hostname	Group	Alias	Group	Alias
node42			wnodes	work1
node43			wnodes	work2
node44	gnodes	gpu-node1	wnodes	work3
node45	gnodes	gpu-node1	wnodes	work4

You can use the alias names in the `cm` commands in places where you might otherwise have specified node hostnames. The following examples show how alias names appear in output:

- The following command returns the node hostnames and the node alias names:

  ```
  # cm node show --alias-names
  NODE      ALIAS        ALIAS
  node42                 work1
  node43                 work2
  node44    gpu-node1    work3
  node45    gpu-node2    work4
  ```

  The following command returns the node hostnames, the node alias names, and the alias group names:

  ```
  # cm node show --alias-names --alias-groups
  NODE      GROUP   ALIAS        GROUP   ALIAS
  node42                         wnodes  work1
  node43                         wnodes  work2
  ```

```
node44    gnodes  gpu-node1    wnodes  work3
node45    gnodes  gpu-node2    wnodes  work4
```

- The following command does not specify that the cluster manager display information using aliases, so the command returns the node hostnames and the management network information for each node:

```
# cm node show -M
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
node42          None            192.168.61.2    255.255.255.0   00:9c:02:99:1b:3c    default         default
node43          None            192.168.61.3    255.255.255.0   00:9c:02:a5:29:06    default         default
node44          None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
node45          None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
```

- The following command requests network information for two nodes, and the two node hostnames are specified:

```
# cm node show -M -n node44,node45
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
node44          None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
node45          None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
```

  The following command requests network information for the same two nodes; the node aliases are specified; and the node alias names appear in the output:

```
# cm node show -M -n gpu-node1,gpu-node2
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
gpu-node1       None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
gpu-node2       None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
```

- As the following command shows, you can use wildcard characters in commands with alias specifications:

```
# cm node show -M -n 'gpu-*'
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
gpu-node1       None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
gpu-node2       None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
```

  The following is another command that shows how to use wildcard characters in a command with alias names:

```
# cm node show -M -n 'work[3-4]'
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
work3           None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
work4           None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
```

- To display aliases instead of hostnames in the output of the `cm node show` command, specify either `--convert-to-aliases` or `-D`.

  For example, the following command specifies output that displays hostnames for this one command and only for the group `gnodes`:

```
# cm node show -D gnodes -M
NODE            NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS           MGMTSERVERIP    DEFAULTGATEWAY
gpu-node1       None            192.168.61.4    255.255.255.0   00:9c:02:99:1b:b0    default         default
gpu-node2       None            192.168.61.5    255.255.255.0   00:9c:02:99:2f:28    default         default
node42          None            192.168.61.2    255.255.255.0   00:9c:02:99:1b:3c    default         default
node43          None            192.168.61.3    255.255.255.0   00:9c:02:a5:29:06    default         default
```

- The following command specifies the compute nodes aliased to `work1`, `work2`, `work3`, and `work4` on the `cm node run` command:

```
# cm node run -n 'work*' --confirm ethtool eno1

This command will include the following node(s): work[1-4]

continue [y|n]: y
```

- The following command opens a console to the node with node name alias `work4`:

```
# cm node console -n work4
[Enter `^Ec?' for help]

Red Hat Enterprise Linux 8.X (Ootpa)
Kernel 4.18.0-240.el8.x86_64 on an x86_64
```

```
node45 login:
```

The Linux `console` command does not recognize node name aliases.

- The following command specifies the node with node name alias `work4` on the `cm node provision` command:

  ```
  # cm node provision -i rhel8.X -n work4

  Assigning image "rhel8.X" and kernel "4.18.0-240.el8.x86_64" to the nodes...

  Configuration manager initiating node configuration.
  1 of 1 nodes completed in 1.9 seconds, averaging 0.2s per node
  Node configuration complete.

  Checking node power status...
  Halting the nodes that are not down...

  direct node node45 has been issued a halt command

  Waiting 15 seconds for nodes to halt...

  Checking node power status...

  Setting non-autoinstall nodes to provision on their next boot...


  Checking node power status...
  Issuing node reset to non-autoinstall nodes that are "On"...

  direct node node45 has been issued a reset command
  ```

  The output from the command contains a mix of the node name alias and the node hostname. This mix occurs because some of the commands that run underneath `cm node provision` do not recognize node name aliases.

Using commands to provision nodes and manage software images

Each node in the cluster can have one software image running at a time. The topics in this chapter explain how to manage software images.

NOTE: Hewlett Packard Enterprise recommends that you back up an image before you modify an image. For information about how to back up an image to VCS, see the following:

Saving a copy of an image

Node types and default image names

Clusters include different types of nodes, and there is a unique software image for each individual node type. When you install additional software on your cluster, you might need to modify the software on some of the nodes.

To get a profile of the node types on your cluster, use the `cm node show` command. The command can display the profile of all nodes (enter `cm node show`), or you can retrieve profile information for a specified node type. Enter `cm node show -h` to see the various specifications.

The `cm node show` output does not include `admin`, but you can specify `-n admin` on imaging commands and some other commands, such as the following:

- `cm node update config`
- `cm node zypper`
- `cm node dnf`

The following figure shows example image names for the nodes on cluster systems. The master images for the compute nodes, ICE compute nodes, and leader nodes reside on the admin node.

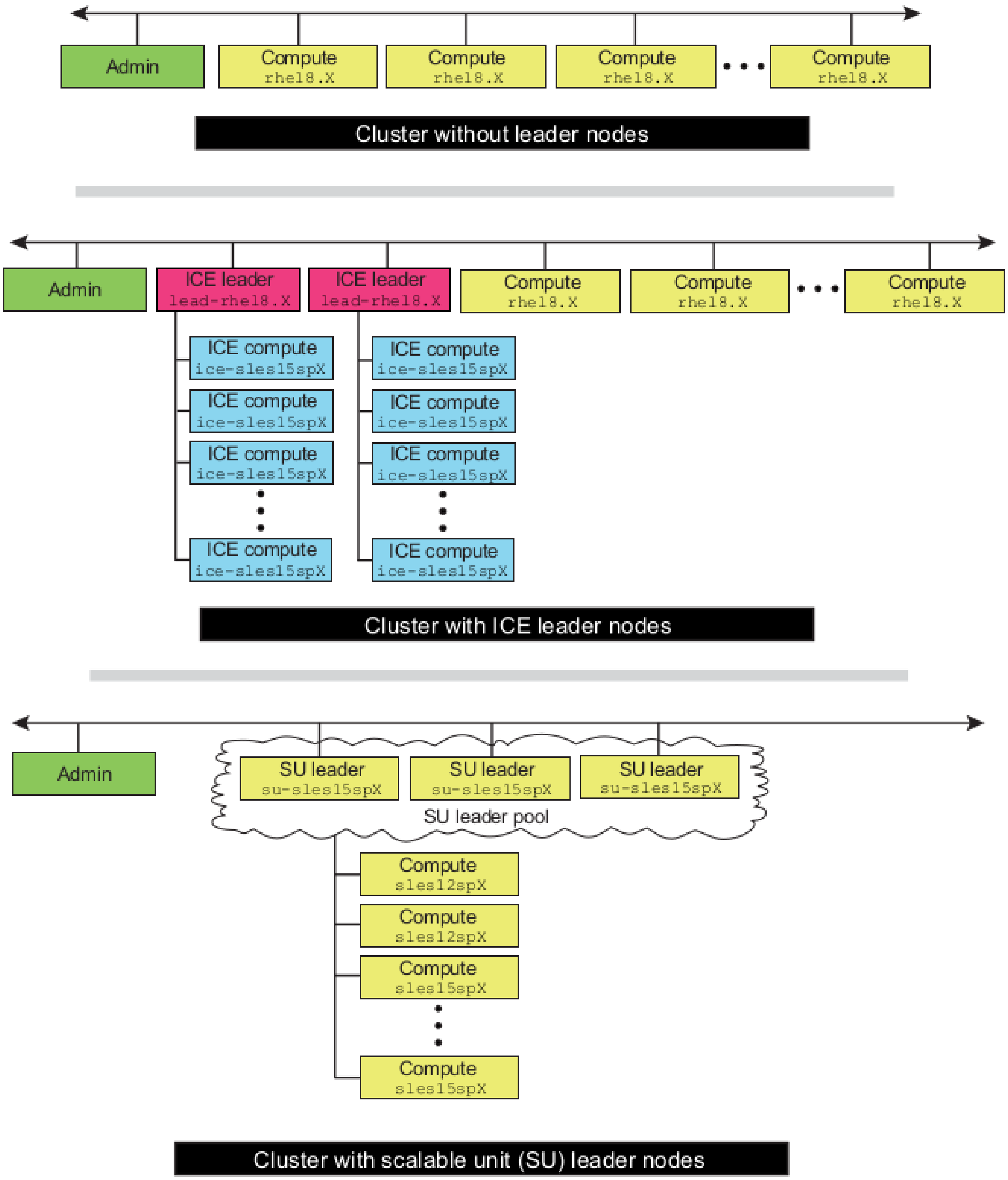


Figure 7: Image types

Image management commands

The following list shows the primary image management commands:

`cm image` ***`action`***

Depending on the *action* you specify, the `cm image` command performs various imaging functions.

For example:

`cm image show`:

- Lists available images.
- Lists images currently on nodes.

`cm image set` or `cm image unset`:

- Sets image characteristics such as quotas, kernel parameters, and other characteristics.
- Clears image characteristics

Other *action* specifications help you complete other image tasks such as the following:

- Creates an image from scratch.
- Recreates an image. Any nodes associated with the image before you run the command are associated with the image after the command runs.
- Uses existing images that might have been created by some other means.
- Deletes images.
- Shows available images.
- Updates or manages images.
- Formally tracks revisions to images.

`cimage`

Used with ICE cluster images only. Performs the following functions:

- Assigns images to nodes.
- Pushes images to racks.

`cm repo`

Manages repositories. Performs the following functions:

- Adds, deletes, and displays repositories.
- Selects and clears repositories for RPM list generation.
- Creates, deletes, and displays logical groups of repositories.

`cm node provision`

Deploys images on nodes.

On HPE Cray EX clusters, HPE Apollo clusters, and SGI Rackable clusters, the `cm node provision` command does the following:

- Assigns images.
- Instructs nodes regarding how to install themselves or how to NFS boot themselves.

The cluster manager does not support the `cm node provision` command on clusters with ICE leader nodes.

`cm node set` (HPE Apollo clusters and SGI Rackable clusters)

Sets the root file system type. Can be one of the following:

- `--rootfs disk`
- `--rootfs tmpfs`
- `--rootfs nfs`

  When you specify an NFS root file system, you can specify any of the following properties:

  - `--writable nfs-overmount`
  - `--writable nfs-overlay`
  - `--writable tmpfs-overmount`
  - `--writable tmpfs-overlay (default)`

- `--rootfs custom`

`cm image activate` (HPE Apollo clusters and SGI Rackable clusters)

Activates the NFS file system capabilities.

To retrieve a list of parameters for each command, enter the command name and `-h` on the command line.

Image management flowcharts

The following flowcharts show image management creation and customization commands.

Customization flow step 1 shows the following ways to create an image:

- The first method uses an RPM list to build the image on the admin node.
- The second method uses an existing image and brings that image into the cluster manager.
- The third method uses autoinstall (autoyast or kickstart) to build the image on the admin node.
- The fourth method captures an image from another node to bring that image into the cluster manager.

All methods enable you to create an image and store the image on the admin node. For more information, see the `cm image` manpage.

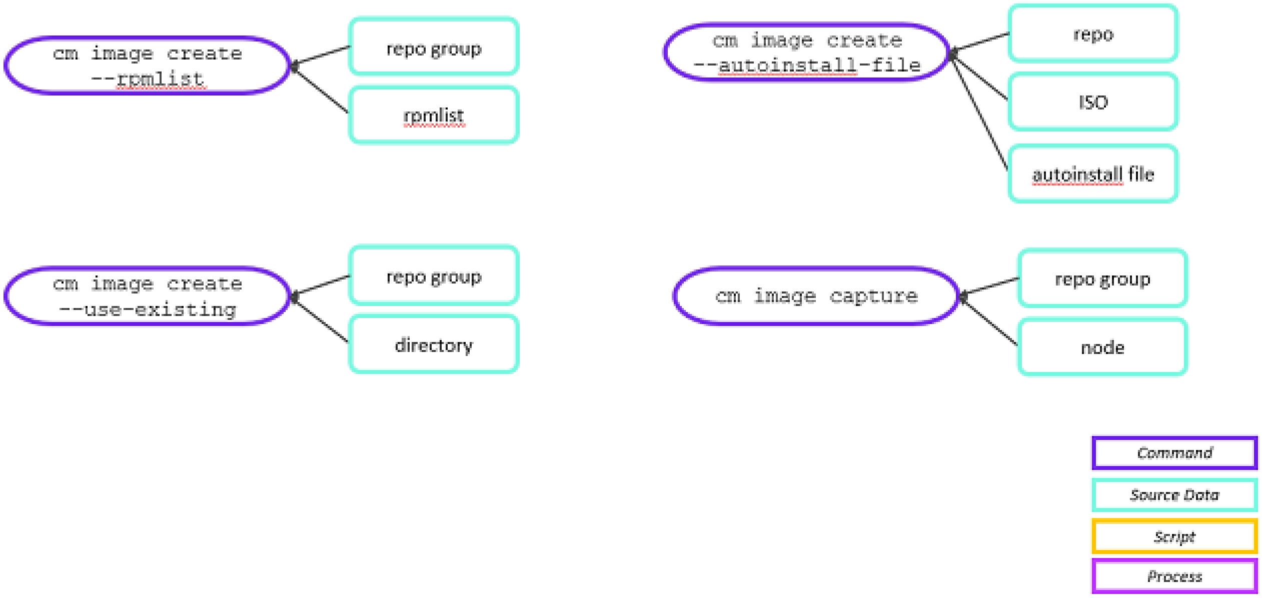


Figure 8: Customization flow step 1: creation

Customization flow step 2 shows that you can run a script before you create the image. This flow also shows that some image management commands offer you a `--pre-script` parameter and a `--post-script` parameter. The ability to run a script assumes that you based the image on an RPM list or on an existing image. After you create the image, you can use the `cm image set` command to further customize the image. For more information, see the `cm image` manpage and the `cm image set` manpage.

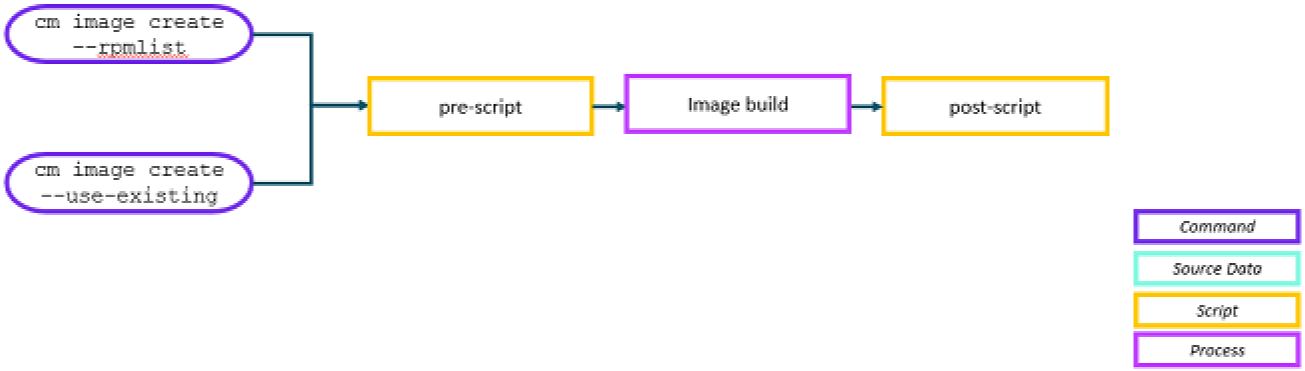


Figure 9: Customization flow step 2: creation options

Customization flow step 3 shows that you can use the `cm image set` command to further customize an image based on an RPM list or on an existing image.

Figure 10: Customization flow step 3: customization options after the image exists

Customization flow step 4 shows the following image-provision flows:

- On the left, for HPE SGI 8600 clusters, the cluster manager pushes the image to the rack of ICE compute nodes. As the image is pushed, the per-host customization scripts run.
- On the right, for clusters without leader nodes or for clusters with scalable unit (SU) leader nodes, the provisioning step runs the preinstall scripts and the postinstall scripts as the node is provisioned. For more information, see the following file:

  `/opt/clmgr/image/scripts/post-install/README`

For all clusters, when the node boots, the cluster manager runs the scripts that reside in the following directory within the image:

`/etc/opt/sgi/conf.d`

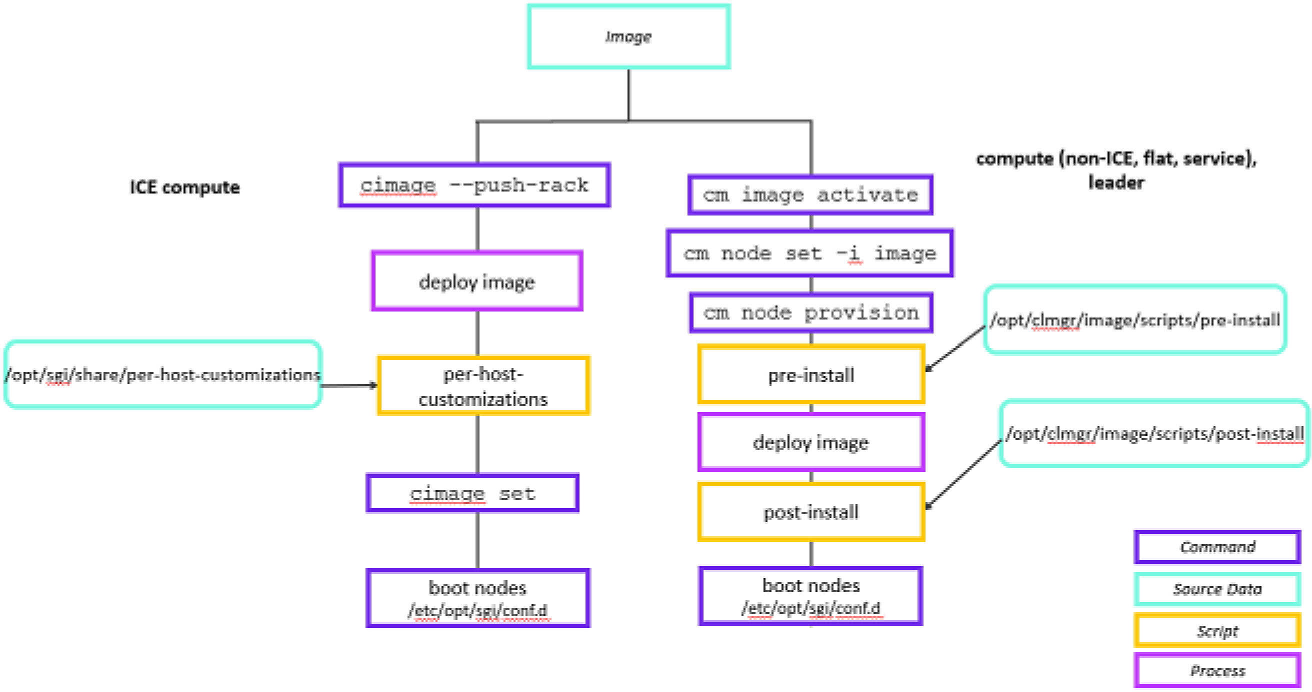


Image
ICE compute
cimage --push-rack
deploy image
/opt/sgi/share/per-host-customizations
per-host-customizations
cimage set
boot nodes /etc/opt/sgi/conf.d
compute (non-ICE, flat, service), leader
cm image activate
cm node set -i image
cm node provision
/opt/clmgr/image/scripts/pre-install
pre-install
deploy image
/opt/clmgr/image/scripts/post-install
post-install
boot nodes /etc/opt/sgi/conf.d
Command
Source Data
Script
Process

Figure 11: Customization flow step 4: deployment

Installation repositories

You can use the `cm image` command and the `cm repo` command to manage the cluster manager software and the Linux distribution software. You can use these commands to manage custom repositories that you create yourself or to add media.

Often, the repositories for the cluster manager and for the Linux distribution reside in the following directory on the admin node:

`/opt/clmgr/repos`

On other clusters, the repositories could reside on a remote server.

The following topics explain concepts related to installation repositories:

- **Repository metadata**
- **Creating a remote repository**

- **General repository management parameters**
- **Updating RPM lists**

Repository metadata

The cluster manager associates the following with each repository:

- A name

  The repository name is a metadata item that you supply to the `cm repo` command. Except for custom repositories, the `cm repo` command extracts the repository name from the media when adding a repository. The `cm repo show` command displays the repository name and the repository location in the following format:

  name : *location*

  The following topic contains an example:

  Installation repositories
- A directory
- Selection status
- Suggested package lists

The repository information is determined from the media itself when you add the following types of software:

- HPE Performance Cluster Manager
- HPE Cray operating system (COS)
- Linux distributions (RHEL, SLES, Rocky Linux, or CentOS)
- Any other YaST-compatible software

For customer-supplied repositories, supply information to the `cm repo add` command when adding the repository.

The `cm repo add` command uses items like the selection status and the suggested package lists to build the RPM lists required for software installations.

Creating a remote repository

The cluster manager supports the use of both local repositories and remote repositories. You can access the repositories by using `http` and `https`.

To use remote repositories for HPE distribution media or Linux distribution media, the repositories must contain the complete, expanded media including any dot files (`.`*filename*). If the remote media is HPE distribution media or Linux distribution media, the cluster manager processes the default RPM lists the same way it processes locally hosted media. You can use remote repositories on cluster nodes that are routed to the servers upon which the remote repositories reside. If necessary, establish the correct routing.

For remote Linux distribution repositories, if the distribution spans one source (for example, the distribution includes both DVD1 and DVD2), take care when expanding DVD2. If you overwrite DVD1 files, the overwrite breaks distribution detection.

The following is an example of a correct copy:

```
# cp -a /dvd-mnt-rhel8-dvd1 /web-export/rhel8
# cp -a /dvd-mnt-rhel8-dvd2/* /web-export/rhel8
```

General repository management parameters

You can use the cm repo command to manage your software repositories. The following table summarizes various management actions.

Repository Action	Description
Adding	Use cm repo add. For information, see the following: **Adding software to the cluster manager repository database**
Displaying	Use cm repo show to display all available repositories. Use cm repo show --distros to display repositories for Linux distributions.
Grouping	For information, see the following: **Repository groups**
Deleting	Use cm repo del *repo_name* to delete a repository. This command also deletes the associated /opt/clmgr/tftpboot directory, if present.
Refreshing metadata	Use cm repo refresh *repo_name* to refresh metadata. You cannot refresh HPE distribution repositories or Linux distribution repositories. Instead, put any updated or additional RPMs in a custom repository.

Updating RPM lists

The cm repo command constructs default RPM lists based on the repository metadata. These lists include suggested package lists and selection status. The cm image create command can use RPM lists when creating images. The RPM lists are generated only if a single distribution is selected. You can find the lists in /opt/clmgr/image/rpmlists/generated. The lists match the form generated-*.rpmlist. The cm repo command displays its actions when it updates or removes generated RPM lists.

For example:

```
# cm repo select SLE-15-SPX-Full-x86_64
Updating: /opt/clmgr/image/rpmlists/generated-ice-sles15spX.rpmlist
Updating: /opt/clmgr/image/rpmlists/generated-sles15spX.rpmlist
```

NOTE: When you select a repository, make sure to select an appropriate repository. The cluster manager expects you to select only repositories that are appropriate to the image.

When generating RPM lists, the crepo command combines a list of distribution RPMs with suggested RPMs from every other selected repository. The following directory contains the generated RPM lists:

/opt/clmgr/image/rpmlists/generated

For information about generating RPM lists for repository groups, see the following:

Repository groups

Procedure

1. Copy the generated list.
2. Edit the RPM list copy.

   To override a suggested RPM list, create an override RPM list associated with a media type in the following directory:

   ```
   /opt/clmgr/image/rpmlists/override/
   ```

   For example, to change the default RPM list for the SLES 15 SPX HPE Performance Cluster Manager media, create the following directory:

   ```
   /opt/clmgr/image/rpmlists/override/Cluster-Manager-1.8-sles15spX-x86_64
   ```

   Inside the new directory, create the RPM lists that you want to override. For example:

   ```
   /opt/clmgr/image/rpmlists/override/Cluster-Manager-1.8-sles15spX-x86_64/cm.rpmlist
   /opt/clmgr/image/rpmlists/override/Cluster-Manager-1.8-sles15spX-x86_64/cm-ice.rpmlist
   /opt/clmgr/image/rpmlists/override/Cluster-Manager-1.8-sles15spX-x86_64/cm-lead.rpmlist
   /opt/clmgr/image/rpmlists/override/Cluster-Manager-1.8-sles15spX-x86_64/cm-admin.rpmlist
   ```

3. Pass the updated RPM list copy to the `cm image refresh` command.

   For example:

   ```
   # cm image refresh -i sles15spX \
   --rpmlist /opt/clmgr/image/rpmlists/override/ \
   Cluster-Manager-1.8-sles15spX-x86_64/cm.rpmlist
   ```

Applying image overrides

You can specify that the cluster manager install additional files into an image during the post-install deployment process. You can use this procedure, for example, if there is a script that you want to install on all nodes after you deploy the node image. By storing this script in the `/opt/clmgr/image/overrides/` directory, you have them all in one place.

The following procedure explains how to modify an installed image.

Procedure

1. Log into the admin node as the root user.
2. Change to the following directory:

   `/opt/clmgr/image/overrides/`*`image_name`*`/`

   For *image_name*, specify the name of an installed image. This directory must be the top of the target file system.

3. Enter the following command:

   ```
   # mkdir -p -usr/local/bin
   ```

4. Create one or more files that you want to install into the deployed image, or move files to here from another location.

Repository groups

Your site might have a compute environment that requires many repositories or sufficiently interesting sets of repositories. In this case, you might want to define repository groups.

To create a repository group, use the `cm repo group add` command in the following format on the admin node:

`cm repo group add [--repos `*`list`*`] `*`group_name`*

The variables are as follows:

Variable	Specification
list	Specify zero or more repository names, space-delimited. Subsequent `cm repo` commands that specify a group name add members to the group.
group_name	A collective name for the listed repositories.

When you create a repository group that includes a supported Linux distribution and the necessary components, the `cm repo` command generates RPM lists for the default images of the various node types. For example, for group `xyz` that includes the Linux distribution RHEL 8.X, it creates the following:

```
/opt/clmgr/rpmlists/generated-group-xyz-rhel8.X.rpmlist
/opt/clmgr/rpmlists/generated-group-xyz-ice-rhel8.X.rpmlist
/opt/clmgr/rpmlists/generated-group-xyz-lead-rhel8.X.rpmlist
```

Related group actions:

- The following command displays groups or group members of specified groups:

  ```
  cm repo group show [group_names]
  ```

  If no group names are specified, the cluster manager displays all groups names.

- The following command deletes groups or group members:

  ```
  cm repo group del [--repos list] group_name
  ```

  If you do not specify a `[--repos list]` of repository names, the system deletes *group_name*.

  If you delete a repository group member from the repository, the group membership remains intact. Hence, you can recreate a repository using the same name and have its group membership already established.

- The following command assigns repository groups to images for use by the `cm node provision` command:

  ```
  cm image set -i image --repo-group group
  ```

- The following command removes an assignment of a repository group to an image:

  ```
  cm image unset -i image --repo-group
  ```

- The following example command displays repository groups and their assignments:

  ```
  cm image show --settings -i rhel8.X | grep repo-group
  ```

The following topic contains information and examples about repository groups:

Creating an NFS server on a compute node and mounting a file system into the cluster

Building an HPE Cray operating system (COS) image

The following procedure explains how to create a COS compute node image as a series of basic steps. For more complete information, see the COS operating system information contained in the COS documentation on the HPE website.

Procedure

1. Log into the admin node as the root user.
2. Download the COS operating system `.iso` file from the customer portal and write it to a location on the admin node:

3. Use the following command to import the COS RPM:

   ```
   cm repo add /path_to_iso/cos-2.X.30.iso
   ```

   For *path_to_iso*, specify the path to the COS `.iso` file you downloaded.

4. Use the `cm repo group` command to create a repository group that contains the following:

   - COS
   - SLES
   - The HPE Performance Cluster Manager

   For example:

   ```
   # cm repo group add cos-2.X.30-hpcm \
   --repos Cluster-Manager-1.8-sles15spX-x86_64 SLE-15-SPX-Full-x86_64 COS-2.X.30-x86_64
   ```

   The command in this step creates the following RPM list:

   ```
   /opt/clmgr/image/rpmlists/generated/generated-group-cos-2.X.30-hpcm.rpmlist
   ```

5. Use the `cm image create` command to create a COS image.

   Use the RPM list and your repository group.

   For example:

   ```
   # cm image create -i cos-2.X.30-hpcm \
   --rpmlist /opt/clmgr/image/rpmlists/generated/generated-group-cos-2.X.30-hpcm.rpmlist \
   --repo-group cos-2.X.30-hpcm
   ```

   The command displays output to the screen and creates files in the following directory:

   ```
   /var/log/cinstallman
   ```

6. Enter the following command to verify that the image was created:

   ```
   # cm image show
   ```

Adding software to the cluster manager repository database

To install new software, first use the `cm repo add` command to add the software to the repository. Log into the admin node as the root user, and use the `cm repo add` command in the following format:

```
cm repo add [--custom 'repository_name'] location
```

The following subsections describe the command options for various repositories:

- **Standard repository**
- **Custom repository**
- **Multiple media sources**
- **Creating nested repositories**

Standard repository

To build a repository for an OS distribution or HPE software media, use the `cm repo` command in the following format:

```
cm repo --add location
```

For *location*, specify one of the following:

- If the new software resides on a remote web server, specify a URL.

  Make sure that the host name in the URL is a fully qualified domain name (FQDN). If you do not specify an FQDN, the `crepo` command might not properly resolve the host location.

  The cluster manager adds the new repository in the remote web location.

- If the new software resides on a server that is NFS-mounted to the admin node, specify the full path to the ISO file on that server. Use the following format for *location*:

  ```
  host:/path_to_ISO_file
  ```

  The cluster manager adds the new repositories to the following directory:

  ```
  /opt/clmgr/repos
  ```

- If the new software resides on the admin node, specify one of the following:

  - The path to the ISO file on the admin node
  - The path to mounted media

  The cluster manager adds the new repositories to the following directory:

  ```
  /opt/clmgr/repos
  ```

Custom repository

When the `cm repo add` command includes the `--custom` option, the value *location* must point to an existing directory (with existing RPMs). Repository metadata is created inside the existing directory. Make sure that *location* resides under the following directory:

```
/opt/clmgr/repos
```

The *location* can be a remote repository.

For more information about custom repositories, see the following:

Example workflow: adding a kernel debug package using a custom repository

Multiple media sources

For Linux distribution media that has more than one DVD, enter a `cm repo add` command for the first DVD and enter a separate command for the second DVD. The `cm repo add` command combines the two DVDs into a single repository.

Creating nested repositories

Some Linux distributions have subdirectories with additional separate repositories in them. The `cm repo add` command searches for these additional repositories and includes the subdirectory repositories when the media is selected. The result is that certain distribution media may have more than one URL associated with the media.

The following procedure explains how to set up nested repositories on a remote server.

Procedure

1. Use the `cm repo add` command to register the remote Linux distro URL.

   Example using a remote repository for RHEL 8.X media:

   ```
   # cm repo add http://updateserver.example.com//rhel/8.X/x86_64/os/
   ```

2. Add a separate custom repository that points to each separate subdirectory you want to include.

   The following example uses the required `ScalableFileSystem` component:

   ```
   # cm repo add  --custom rhel8.X-scalable-fs \
   http://updateserver.example.com//rhel/8.X/x86_64/os/ \
   ScalableFileSystem
   ```

Using the cluster manager version control system (VCS)

Using VCS to create an image is similar to the cloning method. The difference is that when you use VCS, you do not need to employ a second name for the changed image. VCS does the implicit cloning and lets you change the clone. The changed image retains the same name as the original, but VCS assigns different version numbers to the images. Depending on your image management needs, you might find the VCS scheme sufficient.

In this scheme, you move to the image environment on the admin node, `/opt/clmgr/image/images`, and update the original image.

For more information about VCS, see the following:

Using the version control system (VCS)

Capturing an image from a running compute node

The procedure in this topic explains how to capture the operating environment from a running compute node. You can capture this environment in an image.

Procedure

1. Log into the admin node as the root user.

2. (Optional) Use the `ssh` command to make sure that the operating system repositories currently selected match that of the image that you want to capture from the running compute node.

   For example:

   ```
   # ssh n0 "grep -e 'PRETTY_NAME' /etc/os-release"
   PRETTY_NAME="SUSE Linux Enterprise Server 15 SP4"
   ```

   The `ssh` command output indicates that you should select the `SLE-15-SP4-Full-x86_64` distribution repository.

3. (Optional) Enter the following command to display the list of available distribution repositories:

   ```
   # cm repo show --distros
   ```

4. Use the `cm repo group` command to create a SLES repository group.

For example:

```
# cm repo group add sles15sp4 --repos SLE-15-SP4-Full-x86_64 Cluster-Manager-1.8-sles15sp4-x86_64
```

5. Enter the `cm image capture` command in the following format:

   `cm image capture -i new_image -n node [options]`

   The variables are as follows:

Variable	Specification
new_image	A name for the new image. If the image name *new_image* exists, the cluster manager overwrites or refreshes the existing image. Otherwise, the cluster manager adds image *new_image* to the list of available images in `/opt/clmgr/image/images`.
node	The hostname of the compute node upon which the image is running. If needed, use the `cm node show` command to identify a possible *node*.
options	Specify zero or more options. The command passes any specified *options* arguments to the `rsync` command. Examples of such arguments are as follows: • `-g repo_group` • `--exclude` • `--exclude-file` • `--one-file-system` For more information, see the `rsync` manpage.

   For example:

```
# cm image capture -i new_image -n node -g sles15sp4
```

NOTE: The cluster manager does not support capturing an image from a running ICE compute node.

Due to the size of the captured image, Hewlett Packard Enterprise does not recommend capturing an image from a leader node.

For related information, see the following:

Comparing the image on a running node with images on the admin node

Installing new software into new images

Generally, you want to create new images for software updates. You might want to do so for scalability, change tracking, and reusability. When installing software into cluster images, there are various cases to consider.

Some scenarios involve installing the software update directly on a running compute or leader node. Depending on your site requirements, the direct installation onto a running node might be preferred because it avoids the overhead associated with reimaging nodes.

The following topics describe various scenarios:

- **Installing packages from repositories into an image**
- **Installing miscellaneous RPMs into an image**
- **Installing packages from repositories onto running admin nodes, leader nodes, or compute nodes**
- **Installing miscellaneous RPMs onto running leader nodes or compute nodes**

Installing packages from repositories into an image

Procedure

1. Select a repository.
2. Back up the current image.

   For information, see the following:

   Saving a copy of an image
3. Add one or more packages from the repository to the current image.

   Use one of the following commands:

   - On RHEL 9 platforms and RHEL 8 platforms, use the following command:

     ```
     cm image dnf [--duk] -i image install packages
     ```

   - On RHEL 7 platforms, use the following command:

     ```
     cm image yum [--duk] -i image install packages
     ```

   - On SLES platforms, use the following command:

     ```
     cm image zypper [--duk] -i image install packages
     ```

   The variables are as follows:

Variable	Specification
`--duk`	This parameter is conditional. Use this parameter if you want to update an image, but you do not want to update the kernel. When you specify `--duk`, you might save some processing time. By default, the cluster manager updates the kernel in an image when you add a package to an image.
image	The image into which you want to install the packages.
packages	One or more packages from the selected repositories.

For example:

```
# cm image dnf --duk -i ice-rhel8.X install additional_new_package
```

Installing miscellaneous RPMs into an image

The following procedure explains how to install a miscellaneous set of RPMs into an existing image.

Procedure

1. Back up the current image.

   For information, see the following:

   Saving a copy of an image

2. Log into the admin node as the root user, and enter the following command to switch to the images directory:

   ```
   # cd /opt/clmgr/image/images/
   ```

3. Copy the RPMs you want to add to the image.

   For example, if your RPMs are in `/tmp/newrpm.rpm` and you want to update the `rhel8.X-new` image, enter the following:

   ```
   # cp /tmp/newrpm.rpm rhel8.X-new/tmp
   ```

4. Change to the image environment.

   For example:

   ```
   # chroot rhel8.X-new
   ```

5. Install the RPMs.

   For example:

   ```
   # rpm -Uvh /tmp/newrpm.rpm
   ```

6. Enter the following command to exit the `chroot` environment:

   ```
   # exit
   ```

Installing packages from repositories onto running admin nodes, leader nodes, or compute nodes

Rather than installing a software package in an existing image on the admin node, you can install the package directly on a running admin node, leader node, or compute node.

Procedure

1. Select a repository.

2. Back up the current image.

   For information, see the following:

   Saving a copy of an image

3. Use the `cm node` command in one of the following formats:

- On RHEL 9 or RHEL 8 platforms, use the following format:

  ```
  cm node dnf -n node install packages
  ```

  NOTE: The cluster manager supports RHEL 9 only on compute nodes. The cluster manager does not support RHEL 9 on leader nodes.

- On RHEL 7 platforms, use the following format:

  ```
  cm node yum -n node install packages
  ```

  NOTE: The cluster manager supports RHEL 7 only on compute nodes. The cluster manager does not support RHEL 7 on leader nodes.

- On SLES platforms, use the following format:

  ```
  cm node zypper -n node install packages
  ```

The variables are as follows:

Variable	Specification
node	One or more node hostnames. Specify the nodes to receive the packages. The *node* specification can include an at sign (@) to represent a custom group of nodes.
packages	One or more package names from the selected repositories.

Example 1. The following command installs package `zlib_devel` on SLES compute node `n0`:

```
# cm node zypper -n n0 install zlib_devel
```

Example 2. The following command installs a package on the admin node and specifies a repository group:

```
# cm node zypper -n admin --repo-group cm1.7-sles15sp3-aiops install emacs
```

4. (Optional) Capture the image from the running node and store the image on the admin node for further use.

   For related information, see the following:

   Capturing an image from a running compute node

Installing miscellaneous RPMs onto running leader nodes or compute nodes

You can install any set of RPMs onto running leader nodes or compute nodes. Later, you can capture and store the images. In this case, use standard Linux tools to manually install the RPMs.

The following general procedure explains how to install the RPMs:

Procedure

1. Log into the node.
2. Copy the RPMs you need to the following directory:

   ```
   /tmp/your_rpm_directory
   ```

3. Use the following command to Install the RPMs:

   ```
   rpm -Uvh /tmp/your_rpm_directory
   ```

Associating nondefault images with targeted nodes

Unless otherwise instructed, the cluster manager uses the default images for cluster nodes at boot time. The default images are determined by node type.

The following topics explain how to change the default image-node association for one or more nodes:

- **Associating a nondefault image with leader nodes and compute nodes**
- **Associating a nondefault image with ICE compute nodes**

For related information, see the following:

Node types and default image names

Associating a nondefault image with leader nodes and compute nodes

To associate a nondefault image with one or more leader nodes or compute nodes, use the `cm node provision` command in the following format:

```
cm node provision -s -n node -i new_image
```

The variables are as follows:

Variable	Specification
node	One or more node hostnames. The *node* specification can include an at sign (@) to represent a custom group of nodes.
new_image	The image to be associated with the *node*. You do not need to specify a kernel. If two kernels are available, and one is a rescue kernel, the cluster manager selects the nonrescue kernel. If three or more kernels are available, the command displays the available kernels and prompts you to select a kernel. For information about how to specify a kernel on the command line, see the command help.

Example 1. The command in this example assigns image `sles15spX-new` to all compute nodes. The `-s` parameter indicates that you do not want to reboot and reprovision the nodes at this time.

```
# cm node provision -s -n 'n*' -i sles15spX-new
```

Example 2. The command in this example assigns image `lead-rhel8.X-nis` to all leader nodes in custom group `leader`. The `-s` parameter indicates that you do not want to reboot and reprovision the nodes at this time. The `--kernel` parameter indicates that you want to specify a particular kernel.

```
# cm node provision -s -n '@leader' -i lead-rhel8.X-nis --kernel 2.6.32-504.el6.x86_64
```

Associating a nondefault image with ICE compute nodes

To associate a nondefault image with one or more ICE compute nodes, use the `cimage` command in the following format:

```
cimage --set new_image kernel_version node
```

The variables are as follows:

Variable	Specification
new_image	The image to be associated with the nodes.
kernel_version	The kernel to be used.
node	One or more node hostnames.

The following command assigns image `ice-sles15spX-new` to all ICE compute nodes:

```
# cimage --set ice-sles15spX-new 4.4.21-81 'r*i*n*'
```

To change the default kernel association, use the `cimage --show-images` command to see the available image-kernel combinations.

Enabling federal information processing support (FIPS) in an image

The cluster manager enables you to create a new, FIPS-compliant images that strengthen the security and encryption capabilities of a cluster.

Enabling federal information processing support (FIPS) in a RHEL 9.X image or in a RHEL 8.X image

The procedure in this topic explains how to create a FIPS-compliant image that includes a compliant `initrd`. While you can obtain the same results with a node-based approach, using the `cm node set` command, this procedure takes an image-based approach.

Procedure

1. Log into the admin node as the root user.
2. Ensure that the FIPS content is installed in the image.

   Use the following command:

   ```
   cm image dnf --duk -i image install -t pattern fips
   ```

   For *image*, specify the image name.
3. Enter the following commands to initialize the image for FIPS:

   ```
   # chroot /opt/clmgr/image/images/image mount -t proc proc /proc
   # chroot /opt/clmgr/image/images/image fips-mode-setup --enable
   # chroot /opt/clmgr/image/images/image umount /proc
   ```
4. Update the kernels.

   Use the following command:

   ```
   cm image update -i image --kernels
   ```

For *image*, specify the image name.

5. Verify whether or not kernel parameters have been set in the image.

   For example, to verify whether there are custom parameters set in the image named `rhel8`, enter the following command:

   ```
   # cm image show -s -i rhel8
   custom-partitions    = Undefined
   crashkernel          = crashkernel=410M
   hard-quota           = Undefined
   kernel-extra-params  = cgroup_disable=memory
   kernel-distro-params = ro root=dhcp selinux=0 biosdevname=0 numa_balancing=disable
   kernel-leader-params =
   nfsroot-extra-params = Undefined
   quota-timer          = Undefined
   repo-group           = Undefined
   soft-quota           = Undefined
   ```

6. Enable FIPS in the kernel command line.

   Use the following command:

   ```
   cm image set -i image --kernel-extra-params "fips=1 [other_params]"
   ```

   The variables are as follows:

Variable	Value
image	The name of the image that you want to be FIPS-compliant.
other_params	Conditional. If the image already was configured already with custom parameters, specify the additional parameters here in a space-separated list.

7. Use the `cm node provison` command to provision the node with the FIPS-compliant miniroot.

Enabling federal information processing support (FIPS) in a SLES 15 SPX image

The procedure in this topic explains how to create a FIPS-compliant image that includes a compliant `initrd`. While you can obtain the same results with a node-based approach, using the `cm node set` command, this procedure takes an image-based approach.

Procedure

1. Log into the admin node as the root user.

2. Verify whether or not kernel parameters have been set in the image.

   For example, to verify whether there are custom parameters set in the image named on `sles15`, enter the following command:

   ```
   # cm image show -s -i sles15
   custom-partitions    = Undefined
   crashkernel          = crashkernel=410M
   hard-quota           = Undefined
   kernel-extra-params  = cgroup_disable=memory
   kernel-distro-params = ro root=dhcp selinux=0 biosdevname=0 numa_balancing=disable
   kernel-leader-params =
   ```

```
nfsroot-extra-params = Undefined
quota-timer          = Undefined
repo-group           = Undefined
soft-quota           = Undefined
```

3. Enable FIPS in the kernel command line.

   Use the following command:

   ```
   cm image set -i image --kernel-extra-params "fips=1 [other_params]"
   ```

   The variables are as follows:

Variable	Value
image	The name of the image that you want to be FIPS-compliant.
other_params	Conditional. If the image already was configured already with custom parameters, specify the additional parameters here in a space-separated list.

4. Ensure that the FIPS content is installed in the image:

   ```
   cm image zypper --duk -i image install -t pattern fips
   ```

   For *image*, specify the image name.

5. Update the kernel in the image:

   ```
   cm image update -i image --kernels
   ```

6. Use the `cm node provison` command to provision the node with the FIPS-compliant miniroot.

Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Compute nodes can reside in any type of cluster. They are the dominant compute node type in a cluster without leader nodes or in a cluster with scalable unit (SU) leader nodes. All compute nodes are configured to use the Preboot eXecution Environment (PXE) specification. The nodes PXE boot with the kernel from the operating system software distribution and `initrd`.

Provisioning is the act of installing an image on the disks inside the target compute nodes. During this process, the cluster manager transfers a copy of the image from an infrastructure node to the compute node. The infrastructure nodes include the admin node or, if leader nodes are present, one of the leader nodes.

The default transfer method for the image is BitTorrent. The `cm node provision` command supports other image-transfer methods.

The cluster manager takes different actions to provision compute nodes depending on your preferences and the following node characteristics:

- Whether the node is diskless or diskful.
- Whether you want to install an on-disk root file system, an on-disk `tmpfs` root file system, or an NFS file system.

For information about the transfer methods, enter the following command:

```
# cm node provision -h
```

Notes for provisioning a compute node with an on-disk root file system on an HPE Cray EX cluster or an HPE Apollo cluster

If the compute node has internal disks, you can provision the node with the root file system on those disks. **Provisioning** is the act of installing an image on the local disks inside the target compute nodes. During this process, the cluster manager transfers a copy of the image from an infrastructure node to the compute node. The infrastructure nodes include the admin node or, if leader nodes are present, one of the leader nodes.

If you choose to implement a disk root file system, first determine whether or not the nodes were previously installed by using autoinstall.

Notes for provisioning a compute node with an on-disk `tmpfs` file system on an HPE Cray EX cluster or an HPE Apollo cluster

If the compute node has internal disks, you can provision the node with a `tmpfs` root file system on those disks. A `tmpfs` root file system, is a system-memory-based file system. Every time the node boots, the node installs itself with a root file system. System memory hosts the root file system instead of a disk drive.

The default transfer method for the image is UDPcast. The `cm node provision` command supports other image-transfer methods. For information about the transfer methods, enter the following command:

```
# cm node provision -h
```

Notes for provisioning a compute node with an NFS root file system on an HPE Cray EX cluster or an HPE Apollo cluster

When a compute node use NFS for its root file system, the root file system resides on the cluster manager infrastructure nodes, as follows:

- On a system with an admin node, but with no scalable unit (SU) leader nodes, the admin node is the only infrastructure node. The admin node is the NFS server for the network file system.
- On a system with SU leader nodes, the admin node and the SU leader nodes are the infrastructure nodes. The SU leaders are the NFS servers for the NFS file system.

An NFS root file system is never physically resident on a compute node. The compute node mounts the file system from one of the infrastructure nodes. If you want the network to host the root file system, configure an NFS root file system for the compute node.

For compute nodes with an NFS root file system, **provisioning** is a two-step process. The `cm image activate` command specifies that compute node can use the image, and the `cm node provision` command specifies the node to receive the image. If requested, provisioning also sets up areas for NFS writing. During provisioning, the compute node mounts its root file system from one of the infrastructure nodes.

NFS root file system - provisioning a compute node with a condensed image on an HPE Cray EX cluster or an HPE Apollo cluster (default)

For NFS root file systems, the default node image exists as a file system image contained in a single file. The cluster manager uses a SquashFS model to pack the image into the single file. By encapsulating the whole image into a single SquashFS file, nodes that use NFS for their root file systems provision more quickly than other methods. With a condensed image, the node can use the buffer cache locally, and NFS metadata traffic between the compute node and the server is reduced.

Hewlett Packard Enterprise recommends this method for all clusters with NFS root file systems, particularly those with 2,000 or more nodes. A disadvantage to this method is that you cannot edit the image. The node itself mounts the NFS export, and loopback mounts the file system image. The node uses that image for the read-only root component when booting. The image becomes the root file system on the node. In the online help and in documentation, the terms **condensed image** and **image bundle** are interchangeable.

NFS root file system - provisioning with an expanded tree image on an HPE Cray EX cluster or on an HPE Apollo cluster

An expanded tree image exists as a hierarchical set of files and directories. This image, is unlike a single archive, condensed image.

The advantage of an expanded tree image is that you can provision a node, browse the images, edit the content, and change the content while the node is booted. If you are developing a node image, you can provision a few nodes with the expanded tree image as part of your iterative image development efforts. A disadvantage is that this provisioning method is slower than the default, condensed-image method, especially as the node count rises. When many compute nodes boot with this method, the admin node or leader nodes incur an intense load.

NFS root file system - specifications for the writable area of a compute node on an HPE Cray EX cluster or on an HPE Apollo cluster

The following figure shows characteristics you can specify for the writable area of an NFS root file system. The figure puts various specifications in context, introduces the `rwtab` file, explains writable areas, and includes other information.

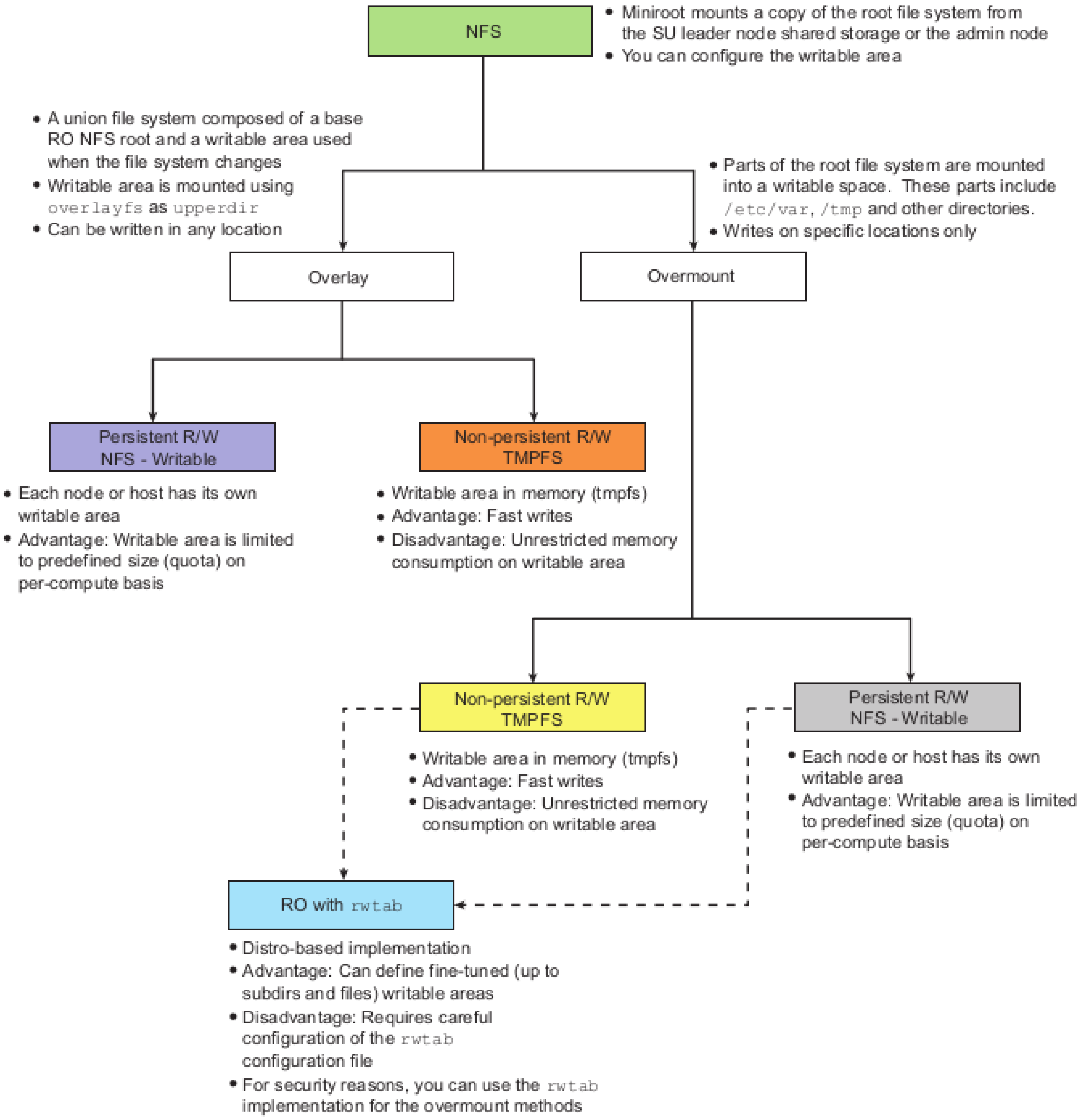

NFS
• Miniroot mounts a copy of the root file system from the SU leader node shared storage or the admin node
• You can configure the writable area
• A union file system composed of a base RO NFS root and a writable area used when the file system changes
• Writable area is mounted using overlayfs as upperdir
• Can be written in any location
• Parts of the root file system are mounted into a writable space. These parts include /etc/var, /tmp and other directories.
• Writes on specific locations only
Overlay
Overmount
Persistent R/W NFS - Writable
• Each node or host has its own writable area
• Advantage: Writable area is limited to predefined size (quota) on per-compute basis
Non-persistent R/W TMPFS
• Writable area in memory (tmpfs)
• Advantage: Fast writes
• Disadvantage: Unrestricted memory consumption on writable area
Non-persistent R/W TMPFS
• Writable area in memory (tmpfs)
• Advantage: Fast writes
• Disadvantage: Unrestricted memory consumption on writable area
Persistent R/W NFS - Writable
• Each node or host has its own writable area
• Advantage: Writable area is limited to predefined size (quota) on per-compute basis
RO with rwtab
• Distro-based implementation
• Advantage: Can define fine-tuned (up to subdirs and files) writable areas
• Disadvantage: Requires careful configuration of the rwtab configuration file
• For security reasons, you can use the rwtab implementation for the overmount methods

Figure 12: Specifications for NFS root file systems for compute nodes when provisioning with a condensed image

Provisioning diskless compute nodes with an NFS image on an HPE Cray EX cluster or on an HPE Apollo cluster

For more information about the different ways you can configure the node image, see the following:

NFS root file system - specifications for the writable area of a compute node on an HPE Cray EX cluster or on an HPE Apollo cluster

Procedure

1. Log into the admin node as the root user.
2. Use the `cm image` command to modify an image or create a new image.

The cluster manager implements several defaults when it creates images.

Example 1. Assume that you want to configure NFS with NFS overlay, which is the default. In this case, the cluster manager sets the image size for the per-host writable areas used in writable NFS solutions. The size is 500M. You can use the `cm image set` command with the `--perhost-size` to change this value. If you want to configure `tmpfs` in the writable area, you do not need to specify a custom `--perhost-size`.

NOTE: If you want to decrease an image size, delete the per-host writable image files. If you reduce the per-host writable image size, that action also destroys the file system image.

Example 2: Assume that you want to specify the granularity of the writable areas. Specify `rwtab` with either the `tmpfs-overmount` or the `nfs-overmount` option. These options let you specify the files and/or directories that can be written.

For more information, see the following file on any booted node:

```
/opt/clmgr/image/etc/rwtab.example
```

Alternatively, on the admin node, in the image root directory, see the following file:

```
/opt/clmgr/image/images/image_name/opt/clmgr/image/etc/rwtab.example
```

3. (Conditional) Power off any nodes that use the image you updated.

   Complete this step if you modified an existing image and an existing version of this image is running on nodes at this time.

   For example, to power off `n0`:

   ```
   # cm power off -n n0
   ```

4. (Optional) Manually delete the read/write area of the existing image on specific nodes.

   ```
   cm image activate --delete-rw-content [-n node] -i image
   ```

   The parameters are as follows:

Parameter	Meaning
node	One or more node hostnames. By default, this command operates on all nodes. The *node* specification can include an at sign (`@`) to represent a custom group of nodes.
image	The image name.

   For example, the following command deletes the read/write area of the development image on `n0` and `n1`:

   ```
   # cm image activate --delete-rw-content -t development.image -n n0,n1
   ```

5. Use the `cm image activate` command to declare to the cluster manager that the image you want to provision is suitable for NFS clients.

   The format is as follows:

   ```
   cm image activate --expanded [--force] [--no-inplace] [--no-msrsync] [--disable-kdump] -i image
   ```

   The parameters are as follows:

Parameter	Meaning
`--force`	Activates a previously activated image. Optional.
`--no-inplace`	Prevents the use of `--inplace` with the `rsync` transport method. Optional.
`--no-msrsync`	Prevents the use of `msrsync`, which is parallel `rsync`. Forces the use of `rsync`. Optional.
`--expanded`	Creates an expanded tree image that is suitable for use during the image development phase. You can edit this image, but it takes longer to deploy this image on a node. Optional.
`--disable-kdump`	Disables `kdump`. If you specify this parameter, image activation can be faster, but it leaves the activated image without `kdump` support. Optional.
image	The image name.

There are additional options that you can specify. For information, see the help output for the `cm image activate` command.

NOTE: This step can be destructive if nodes are booted on the image you activate. Avoid changing the root file system of a node out from under the node.

6. Use the `cm node provision` command to provision the node or nodes with an image.

   The format is as follows:

   ```
   cm node provision -n node -i image [--rootfs nfs] [--writeable area] [--stage]
   ```

   The parameters are as follows:

Parameter	Meaning
node	One or more node hostnames. The default is all nodes in the cluster.
image	The image name.
`--rootfs nfs`	Specifies an NFS root file system. Optional. By default, the cluster manager creates an NFS file system.

Table Continued

Parameter	Meaning
area	Specify one of the following designs for the writable area for the NFS root file system: • `nfs-overmount`, which designs the writable area as a read/write NFS mount point. • `nfs-overlay`, which designs the writable area as an NFS overlay with a read/write NFS mount point. • `tmpfs-overmount`, which designs the writable area as a `tmpfs` mount point. • `tmpfs-overlay`, which designs the writable area as an NFS overlay with a `tmpfs` mount point. Default.
`--stage`	Prevents the cluster manager from automatically power-cycling the node at this time. You can power cycle the node manually later. Optional.

For a graphic description of the NFS solutions, see the following:

Figure 12: Specifications for NFS root file systems for compute nodes when provisioning with a condensed image

7. (Optional) Repeat the preceding steps if you need to adjust settings.

   After testing and experimentation, you might need to adjust some settings in the image.

(Optional) Changing the compute node boot order before provisioning on an HPE Cray EX cluster or an HPE Apollo cluster

Before the cluster manager provisions a compute node, it boots the node over the network. This is the default action, and it is called **network booting**.

The network boot process is as follows:

1. The compute node BIOS and network controller perform a PXE boot.
2. The admin node or leader nodes respond to requests and facilitate transfer of the kernel and `initrd`.
3. The compute node starts the Linux kernel and begins the boot process.
4. The `initrd` loads the cluster manager miniroot, which contains setup and provisioning tools for all root types.
5. If the node is set up to boot from disk and does not have an install pending, the miniroot mounts the necessary file systems and passes control to Linux startup.

   Otherwise, the miniroot facilitates the provisioning, which could be NFS-based or image-transfer based.

   The transport type is the method used to transfer the image when NFS provisioning is not used. The transport types include `rsync`, UDPcast, and BitTorrent. Even when NFS is used, the transport type can influence how the cluster manager loads the miniroot and install scripts. Typically, `rsync` transport is suggested for NFS provisioning methods and UDPcast is the default otherwise.

   To change the default transport method for a node, specify the `--transport` parameter on the `cm node provision` command.

If the node uses local disk drives for its root file system, you can adjust the boot order for the node and avoid network booting.

The following procedure explains how to update the boot order for a node.

Procedure

1. Log into the admin node as the root user.
2. Use the `ssh` command to log into the compute node.
3. Update the boot order.

   Use the compute node hardware documentation to access the BIOS and to modify the boot order.

Updating an expanded NFS image with a small change on an HPE Cray EX cluster or an HPE Apollo cluster

Assume that on a cluster with scalable unit (SU) leader nodes, activating an NFS root image is taking too long. This topic explains how to directly modify an activated NFS root image. Use the information in this topic if you made a small change to an NFS image when you activated it using `--expanded`, and you do not want to wait for the activation to complete.

Procedure

1. Edit the image file directly in the activation path.

   The activation path on the node is as follows:

   ```
   /opt/clmgr/image/images_ro_nfs/image_name
   ```

2. Update the image copy on the admin node in the following directory:

   ```
   /opt/clmgr/image/images/image_name
   ```

Setting a new default image on an HPE Cray EX cluster or an HPE Apollo cluster

The procedure in this topic explains how to set a default image for specific node type. For example, if you create a new image and want to set it as the default image for all compute nodes, you can use the procedure in this topic.

Procedure

1. Log into the admin node as the root user.
2. Show the default images:

   ```
   # cm image show --settings
   Image Type           Image Name           Kernel Version
   Flat Compute         rhel8.X              4.18.0-193.14.2.el8_2.x86_64
   Leader               lead-rhel8.X         4.18.0-193.14.2.el8_2.x86_64

   Global Image Attributes:

   copy-admin-ssh-config = yes
   ```

3. Set the `Flat Compute` image as the default image.

   This step sets the default image for all the compute nodes in the cluster today and all the nodes to be configured into the cluster in the future.

   ```
   cm image set -i image --default [kernel_version]
   ```

   The variables are as follows:

Variable	Specification
image	The name of the compute node image shown in the output from Step **2**.
kernel_version	Optional. One of the kernels shown in the output from Step **2**. If you do not specify a *kernel version*, the cluster manager processes the command as follows: • If only one kernel is available, the cluster manager selects that kernel. • If two kernels are available, and one is a rescue kernel, the cluster manager selects the nonrescue kernel. • If multiple kernels are available, the cluster manager displays the list and asks you to select one.

Provisioning compute nodes on an HPE Cray EX cluster or on an HPE Apollo cluster with scalable unit (SU) leader nodes

The following procedure provisions compute nodes with an image that resides on the admin node. This procedure assumes that image name on the admin node is the same as the image name on the compute node. For example, if you updated the image, but did not change its name, use the procedure in this topic.

Prerequisites

Verify that the compute nodes you want to provision have disks.

Verify that the image you want to provision creates a `tmpfs` root file system.

This procedure applies only to compute nodes that meet the preceding requirements.

Procedure

1. Log into the admin node as the root user.
2. Use the `cm image` command to update the image as needed.

   The cluster manager implements several defaults when it creates images.

   For example, assume that you want to configure one of the following:

   - NFS with read/write NFS overmounts
   - NFS with read/write NFS overlay (default)

   In this case, the cluster manager sets the image size for the per-host writable areas used in writable NFS solutions. This size is 500M. You can use the `cm image set` command with the `--perhost-size` to change this value. If you want to configure `tmpfs` in the writable area, you do not need to specify a custom `--perhost-size`.

   NOTE: If you want to decrease an image size, delete the per-host writable image files. Failure to alter the per-host writable image files destroys the image. For more information, enter the following command:

   ```
   # cm image activate -h
   ```

3. Use the `cm node provision` command in the following format to push the image, reboot, and reprovision the nodes:

   ```
   cm node provision -n node
   ```

   For *node*, specify one or more node hostnames.

   For example, the following command specifies all compute nodes in the custom group `comp`:

   ```
   # cm node provision -n @comp
   ```

4. Run the following command to synchronize the image with the image on the SU leader nodes:

   ```
   su-sync-image image
   ```

   For *image*, specify the image name.

Reprovisioning scalable unit (SU) leader nodes on an HPE Cray EX cluster or on an HPE Apollo cluster

For information about how to reprovision an SU leader nodes, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Clearing the per-host writable space on compute nodes with writable NFS file systems on an HPE Cray EX cluster or on an HPE Apollo cluster

Complete this procedure if you update the node images or if you want to provide a clean start for the nodes. In these cases, Hewlett Packard Enterprise recommends that you clear the per-host writable areas so that the nodes start afresh the next time the nodes boot.

The commands in this topic target an entire image including all nodes. To specify only certain nodes, use the `-n` parameter.

Procedure

1. Log into the admin node as the root user.
2. Use the following command to clear all content:

   ```
   cm image activate --delete-rw-content -i image
   ```

   For *image*, specify the image name.

   This command clears all read/write NFS content in this image for all nodes. When you specify `--delete-rw-content`, you remove the whole directory that is rooted at the image name. This specification makes it easier to work around certain kinds of NFS hangs.

Provisioning ICE compute nodes that have NFS root file systems on HPE SGI 8600 clusters

Procedure

1. Use the `cm power` command in the following format to stop the ICE compute nodes:

   ```
   cm power halt -n node
   ```

For *node*, specify one or more node hostnames.

To specify all ICE compute nodes, specify `'r*i*n*'`.

To specify only one node, specify the hostname in `rxixnx` format.

For example, the following command stops all ICE compute nodes:

```
# cm power halt -n 'r*i*n*'
```

2. Use the `cimage` command in the following format to push the new image to the affected rack leaders:

   ```
   cimage --push-rack image rack
   ```

   The variables are as follows:

Variable	Specification
image	The name of the ICE compute node image that you updated.
rack	One or more rack hostnames. Specify the rack or racks that host the target node or nodes. You can use wildcard characters.

   For example, the following command pushes changes to all the ICE compute nodes:

   ```
   # cimage --push-rack ice-rhel8.X-new 'r*'
   ```

3. Use the `cimage` command in the following format to ensure that the nodes boot with new image the next time the nodes boot:

   ```
   cimage --set --nfs image kernel_version node
   ```

   For example:

   ```
   # cimage --set --nfs ice-rhel8.X-new 2.2.14-504 'r*i*n*'
   ```

4. Power up the targeted ICE compute nodes.

   ```
   cm power on -n node
   ```

   For example, enter the following command to power up all ICE compute nodes:

   ```
   # cm power on -n 'r*i*n*'
   ```

Provisioning ICE compute nodes that have `tmpfs` root file systems on HPE SGI 8600 clusters

Procedure

1. Use the `cimage` command in the following format to push the new image to the affected rack leaders:

   ```
   cimage --push-rack image rack
   ```

   The variables are as follows:

Variable	Specification
image	The name of the ICE compute node image that you updated.
rack	One or more rack hostnames. Specify the hostname or hostnames of the rack or racks that include the nodes. You can use wildcard characters.

For example, the following command pushes changes to all the ICE compute nodes:

```
# cimage --push-rack ice-rhel8.X-new 'r*'
```

For more information about wildcard characters, see the following:

Using the `cm` commands

2. Use the `cimage` command in the following format to ensure that the nodes boot with new image the next time the nodes boot:

   ```
   cimage --set --tmpfs image kernel_version node
   ```

   The variables are as follows:

image	The name of the ICE compute node image that you updated.
kernel_version	The identifier for the kernel in the image.
node	One or more node hostnames. Specify the hostname or hostnames of the nodes that you want to receive the image. You can use wildcard characters.

   For example:

   ```
   # cimage --set --tmpfs ice-rhel8.X-new 2.2.14-504 'r*i*n*'
   ```

3. Use the `cm power` command in the following format to reboot the nodes:

   ```
   cm power reboot -n node
   ```

Running a post-installation script or a pre-installation script automatically

You might have scripts that you want to run before or after an image is created or updated. The cluster manager can run these scripts for you if you install a new image on a node or if you update an existing node image. The procedure in this topic explains how to direct the cluster manager to a script and run the script. The cluster manager runs the script in the package install environment with `/proc`, `/sys`, and other standard directories mounted.

For example, if you have a script that installs software on a node after the node is imaged, the cluster manager can run that script.

Procedure

1. Log into the admin node as the root user.

2. Create the script and write it to the image directory:

The script name determines how the cluster manager handles the script, as follows:

- If the script name starts with the slash (/) character, the cluster manager copies the script to `/tmp` in the image and runs the script from the `/tmp` directory.
- If the script name starts with a character other than the slash (/) character, the cluster manager assumes that the script exists in the image. The cluster manager runs scripts named in this manner in a `chroot` environment relative to the root (/) directory.

3. Run the `cm image create` command or the `cm image update` command.

   For the `cm image create` command or the `cm image update` command, you can run the script after image creation or after image update. Specify the script on the command line as follows:

   `-p` *`scriptname`*

   For the `cm image create` command, you can run the script after the image bootstraps, but before the packages in the list are added to the image. Specify the script on the command line as follows:

   `-P` *`scriptname`*

   Enter one of the following commands to display the complete syntax for the commands that accept the `-p` and `-P` script parameters:

   - `cm image create -h`
   - `cm image update -h`

Example workflow: installing packages onto running nodes or into images

This topic introduces an example workflow that shows how to install a few packages onto running nodes and into images. This workflow creates a new repository to house the packages. Alternatively, you could use an existing repository. In that case, you can modify an existing repository instead of creating a new repository.

A typical scenario is the need to install or update packages. Often, this is done in two steps:

1. Updating running nodes so that the updates are reflected right away
2. Updating the node images so that they are current for future boots

---

NOTE: In certain configurations, you cannot install the packages onto the running nodes themselves. For example, if you have an NFS root file system on ICE compute nodes, you cannot install the packages onto the running ICE compute nodes.

The following topics explain the workflows for installing new software packages:

- **Creating a repository group and adding to a repository group**
- **Installing packages onto running nodes**
- **Installing a package into an existing image on the admin node**

Creating a repository group and adding to a repository group

This example workflow adds packages to running nodes and images. The example uses repositories and repository groups to help manage your software.

To install raw RPMs, put them in a separate repository so that the RPMs are available if you want to create images from scratch later. If you maintain many repositories with different roles, you can add the repository to a repository group.

The workflow in this topic assumes that you want to use a new repository for the packages. The repository group might or might not already exist.

Procedure

1. Choose a name for the repository.

   If you want the repository to be specific to a version of Linux or a type of node, consider including that version or node information in the repository name. If necessary, include the architecture type (x86_64) in the repository name, too.

2. Create the repository.

   The following are the general steps:

   a. Create a directory for the repository. For example:

   ```
   # mkdir /opt/clmgr/repos/other/hpe-maintenance-tools-x86_64
   ```

   b. Copy the packages to the directory. For example:

   ```
   # cp /root/rpm/hp-ams-2.8.2-3017.4.rhel8.x86_64.rpm \
   /root/rpm/ssaducli-3.30-14.0.x86_64.rpm \
   /opt/clmgr/repos/other/hpe-maintenance-tools-x86_64
   ```

   c. Ensure permissions. For example:

   ```
   # chmod a+r /opt/clmgr/repos/other/hpe-maintenance-tools-x86_64/*
   ```

   d. Make a custom repository out of the copied packages. For example:

   ```
   # cm repo add --custom hpe-maintenance-tools-x86_64 /opt/clmgr/repos/other/hpe-maintenance-tools-x86_64/
   ```

3. Add the new repository to a repository group.

   For example, assume that the RPMs are destined for I/O servers and that you can add them to an existing repository group for I/O servers. The following are the general steps that add the new repository to the existing repository group:

   - Display the existing repository groups. For example:

     ```
     # cm repo group show
     ```

   - Add the new repository to one of the existing repository groups, or make a new repository group. For example:

     ```
     # cm repo group add rhel8X-lustre-x86_64 --repos hpe-maintenance-tools-x86_64
     ```

   - Confirm the addition. For example:

     ```
     # cm repo group show rhel8X-lustre-x86_64
     ```

Installing packages onto running nodes

This procedure explains how to install the two new packages on a group of nodes. For this to work, the nodes need to have a root file system type that allows for adding packages. NFS root file systems for ICE compute nodes do not let you change packages live. You can perform a live update for other root file system types, including `tmpfs` and nodes with disks.

Procedure

1. Install the new packages from the repository group onto one or more running nodes.

   For example:

   ```
   # cm node dnf -n gw[1-3],mds[1-2],oss[1-40] \
   --repo-group rhel8X-lustre-x86_64 \
   install hp-ams ssaducli
   ```

   Notice that the updated repository group appears in the preceding command.

2. Verify that the packages installed properly.

   For example:

   ```
   # cm node run -n "gw[1-3],mds[1-2],oss[1-40]" rpm -q hp-ams ssaducli
   ```

Installing a package into an existing image on the admin node

This topic shows how to install a package into an existing image on the admin node. The example uses revision control, which enables you to roll back the change if needed. This operation is not destructive to running nodes.

In this example, the image name is as follows:

```
rhel8.X-mlnx-io
```

The following are notes regarding the parameters on the command lines:

- The commands use the `--duk` parameter to avoid rebuilding the `initrd` process and to avoid updating the miniroot. Do not use `--duk` if you are updating a kernel.
- The `--noplugins` parameter causes the cluster manager to omit the step in which the cluster manager typically contacts the nodes after an image update.

Procedure

1. (Optional) Commit the current image to revision control so you can revert (if necessary).

   For example:

   ```
   # cm image commit --image rhel8.X-mlnx-io --msg 'before installing hp-ams and ssaducli'
   ```

2. Install the package into the image.

   For example:

   ```
   # cm image dnf --image rhel8.X-mlnx-io --repo-group rhel8X-lustre-x86_64 \
   --duk install --noplugins hp-ams ssaducli
   ```

3. (Optional) Commit the updated revision of this image to revision control and supply a commit message.

   For example:

   ```
   # cm image update -i rhel8.X-mlnx-io -c 'installed hp-ams and ssaducli'
   ```

   Now the image is updated.

Miscellaneous image management tasks

Example workflow: adding a kernel debug package using a custom repository

You can maintain software packages specific to your site and have them available to the cluster manager commands. Hewlett Packard Enterprise recommends that you put site-specific packages in a separate location. Do not store site-specific packages in the same location as cluster manager packages or operating system packages.

The following procedure describes how to create a custom repository.

Procedure

1. Create a distinct directory for site-specific packages under the `/opt/clmgr/repos` directory, and move the custom RPMs into that directory.
2. Use the `crepo` command in the following format to create a site-specific repository and add the contents of the new directory to the repository:

   ```
   cm repo add --custom 'repository_name' location
   ```

   The variables are as follows:

Variable	Specification
repository_name	A name you create for the new site-specific repository.
location	The location of the site-specific packages. This location is the location you created in step **1**. The *location* can be one of the following: • A URL • An NFS path specified as *host*:/*path* • A path local to the admin node

   For example:

   ```
   # cm repo add  --custom 'Site-RPMs' /opt/clmgr/repos/site-local/site-rpms
   ```

3. Make the new site-specific repository available to the `cm image` command:

   ```
   cm repo select repository_name
   ```

   For example:

   ```
   # cm repo select Site-RPMs
   ```

4. (Optional) Add the new RPM base names to an existing RPM list.

   This step makes your site-specific RPMs available by default when you create node images in the future.

   Complete the following steps:

   a. Use the `cp` command to copy an existing generated RPM list.

   b. Open the new RPM list file with a text editor.

c. Add each new RPM to the file on individual lines.

d. Save and close the file.

For example, assume that you want to add the following site-specific RPMs to the RPM list called `generated-rhel8.X.rpmlist`:

```
kernel-debug-debuginfo-2.6.32-431.el6.x86_64.rpm
kernel-debuginfo-2.6.32-431.el6.x86_64.rpm
kernel-debuginfo-common-x86_64-2.6.32-431.el6.x86_64.rpm
```

Complete the following steps:

- Enter the following commands:

  ```
  # cp /opt/clmgr/image/rpmlists/generated-rhel8.X.rpmlist \
  /opt/clmgr/image/rpmlists/site-rhel8.X.rpmlist
  # vi /opt/clmgr/image/rpmlists/site-rhel8.X.rpmlists
  ```

- Use the `vi` editor to add the following lines to file `site-rhel8.X.rpmlists`:

  ```
  kernel-debug-debuginfo
  kernel-debuginfo
  kernel-debuginfo-common
  ```

- Save and close the file.

5. Install the new packages into an existing image, onto a node, or into a new image that contains these packages.

- To install the new packages into an existing ICE compute node image, use one of the following formats:
  - For RHEL 9.X or RHEL 8.X images:

    ```
    cm image dnf -i image install package package ...
    ```

  - For RHEL 7.X images:

    ```
    cm image yum -i image install package package ...
    ```

  - For SLES images:

    ```
    cm image zypper -i image install package package ...
    ```

- To install the new packages onto a running compute node, use one of the following formats:
  - For RHEL 9.X or RHEL 8.X:

    ```
    cm node dnf -n node install package package ...
    ```

  - For RHEL 7.X:

    ```
    cm node yum -n node install package package ...
    ```

The following is a RHEL example:

```
# cm node dnf -n n0 install \
kernel-debuginfo kernel-debug-debuginfo kernel-debuginfo-common
```

- For SLES SPX:

```
cm node zypper -n node install package package ...
```

Variable	Specification
node	One or more node hostnames.
image	The name of image that you want to install into the packages.
package	One or more of the packages you wrote to the repository.

For example:

```
# cm image dnf -i ice-rhel8.X install \
kernel-debuginfo kernel-debug-debuginfo kernel-debuginfo-common
```

If necessary, enter the `cm image show` command to retrieve a list of existing images.

- To create an image that includes the packages, use the following format:

```
cm image create -i image -l path_to_rpmlist
```

The variables are as follows:

Variable	Specification
image	A name for the new image.
path_to_rpmlist	The location of the RPM list.

For example:

```
# cm image create -i my-image \
-l /opt/clmgr/image/rpmlists/site-rhel8.X.rpmlists
```

6. (Conditional) Push the changes to the appropriate nodes.

For information, see the following:

Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Comparing the image on a running node with images on the admin node

You might want to compare the image on a running node with the image on the admin node. The following are some situations in which this comparison can be helpful:

- You suspect that the image on the running node is somehow corrupted or has undesirable content.
- You want to confirm that you have added required packages to the running node.

- You are looking for a node that hosts an image that you want to use to reimage other nodes.
- Over a time, you are unsure about untracked changes to the image on the running node.

The cluster manager can display RPM and file differences in the following file:

```
/var/log/cinstallman-idiff.log
```

Additionally, you can customize the comparison by supplying a **comparison file**. A comparison file consists of files and directories to include in the comparison.

NOTE: This feature applies only to leader nodes and compute nodes. However, you can use any image from the image directory on the admin node for the comparison. The node type of the image can be different from the node type of the running node.

To perform the comparison, use the following command:

```
cm image revision diff -i base_image [--rev n] -n node \
[--file comparison_file]
```

The variables are as follows:

Variable	Specification
base_image	The name of the image to be used from the admin node image directory.
n	The image revision (an integer) to be used for the comparison. By default, the command uses the highest revision level of the image. To display existing revisions, use the following command: `cm image revision history -i image`
node	The hostname of the running node used for the comparison.
comparison_file	The name of the comparison file to be used. By default, the command uses the following file: `/etc/opt/sgi/idiff-selection-criteria.conf` To build a custom comparison file, copy the default file to a second file and modify its contents.

Example 1. The following command compares the image on running node `n0` to the default image for compute nodes for SLES 15 SPX:

```
# cm image revision diff -i sles15spX -n n0
```

Example 2. The following command runs the same comparison as example 1, but it uses custom comparison file `/tmp/comparison1`:

```
# cm image revision diff -i sles15spX -n n0 -f /tmp/comparison1
```

Creating RPM lists from images and nodes

After you create an image or deploy an image to a node, it might be necessary to upgrade or modify the image. It is often difficult to know what packages have been added or removed if the version control system (VCS) was not used.

The procedure in this topic explains how to use the `cm image rpmlist` to create a new RPM list from an image or from a running node. You can use the new RPM list to create a new image that reflects the addition or removal of packages.

The `cm image rpmlist` command can also generate an image and compare that image to the image in a repository or the image on a running node.

For information about VCS, see the following:

Using the version control system (VCS)

Procedure

1. Log into the admin node as the root user.
2. Use the `cm image rpmlist` command in one of the following formats to create an RPM list from an existing image or from a running node:

    - To create the RPM list from an existing image, use the command in the following format:

      ```
      cm image rpmlist -W output_file -i image
      ```

    - To create the RPM list from the image on a running node, use the command in the following format:

      ```
      cm image rpmlist -W output_file -n node
      ```

    The variables are as follows:

Variable	Specification
output_file	The name of an output file. The command writes the RPM list to this file.
image	The name of an existing image.
node	The hostname of the node that hosts the image for which you want the RPM list.

3. Create a new image from the RPM list that the `cm image rpmlist` command generated:

    ```
    cm image create -i image --rpmlist location
    ```

Variable	Specification
image	A name for the new image.
location	The full path to the RPM list that the `cm image rpmlist` command created.

    Example 1. The following commands create an RPM list from an existing image named `sles15spX-with-changes`:

    ```
    # cm image rpmlist -W /tmp/my_rpmlist -i sles15spX-with-changes
    # cm image create -i my_image --rpmlist /tmp/my_rpmlist
    ```

    Example 2. The following commands create an RPM list from the image on `node001`:

    ```
    # cm image rpmlist -W /tmp/my_node_rpmlist -n node001
    # cm image create -i node_image --rpmlist /tmp/my_node_rpmlist
    ```

Example 3. The following command displays the differences between two images:

```
# cm image rpmlist -i sles15spX -c sles15spXcopy
Comparing image sles15spX with image sles15spXcopy
There are no rpms in image sles15spX which are not in image sles15spXcopy
The following rpms exist only in image sles15spXcopy:
    libubsan0
    libasan4
    libtsan0
    gcc
    libmpxwrappers2
    linux-glibc-devel
    libmpx2
    liblsan0
    libxcrypt-devel
    libitm1
    libatomic1
    libcilkrts5
    glibc-devel
    gcc7
```

Example 4. The following command compares an image in a repository to the image on a running node:

```
# cm image rpmlist -i sles15spX -n node1
Comparing image sles15spX with node node1
The following rpms exist only in image sles15spX:
    gpg-pubkey
The following rpms exist only on node node1:
    libquadmath0
    libgfortran4
```

Configuring ICE compute node per-host customization on an HPE SGI 8600 cluster

You can add per-host ICE compute node customization to the compute node images. To accomplish the customization, add scripts to one of the following directories on the admin node:

- `/opt/sgi/share/per-host-customization/global/`
- `/opt/sgi/share/per-host-customization/mynewimage/`

Scripts in the global directory apply to all ICE compute nodes images. Scripts under the image name apply only to the image in question. The scripts are cycled through once per host when being installed on the leader nodes.

To run all or a subset of the customization scripts upon demand, using the `cimage` command with the `--customizations-only` parameter.

For more information about the customization scripts, see the following file:

```
/opt/sgi/share/per-host-customization/README
```

The following file contains an example global script:

```
/opt/sgi/share/per-host-customization/global/sgi-fstab.sh
```

NOTE: When creating custom images for ICE compute nodes, make sure to clone the original cluster manager images. You can fall back to the original images if necessary.

For more information about the `cimage` command, see the `cimage` manpage.

Managing ICE compute node images

On clusters with ICE leader nodes, the `cm image show` command and the `cimage` command allow you to list, modify, and set software images on the ICE compute nodes.

The following examples show some typical command operations.

Example 1. The following command lists the available images and their associated kernels:

```
# cm image show -d
name:                 ice-sles15spX
  architecture:       x86_64
  target:             ice_compute
  baseOsName:         sles
  baseOsVersion:      15
  revision:           1
  kernels:            4.12.14-150.14-default

name:                 lead-sles15spX
  architecture:       x86_64
  target:             leader
  baseOsName:         sles
  baseOsVersion:      15
  revision:           1
  kernels:            4.12.14-150.14-default

name:                 sles15spX
  architecture:       x86_64
  target:             compute
  baseOsName:         sles
  baseOsVersion:      15
  revision:           1
  kernels:            4.12.14-150.14-default
```

Example 2. Assume that you entered the following command:

```
# cimage --show-nodes r1
r1i0n0: ice-sles15spX 4.4.21-69-default nfs
r1i0n8: ice-sles15spX 4.4.21-69-default nfs
```

The command lists the following:

- The compute nodes in rack 1
- The image and kernel they are set to boot
- The root file system type (NFS or `tmpfs`)

Example 3. Assume that you entered the following commands:

```
# cimage --set ice-sles15spX 4.4.21-81 r1i0n0
# cimage --show-nodes r1
r1i0n0: ice-sles15spX 4.4.21.81
r1i0n1: ice-sles15spX 4.4.21.69-default-smp
r1i0n2: ice-sles15spX 4.4.21.69-default-smp
.
.
.
```

The commands set the `r1i0n0` compute node to boot the `4.4.21-81` kernel from the `ice-sles15spX` image. The commands also display new node information.

Example 4. The following command sets all nodes in all racks to boot the 4.4.21.81 kernel from the ice-sles15spX image:

```
# cimage --set ice-sles15spX 4.4.21.81 "r*i*n*"
```

Example 5. The following command sets two ranges of nodes to boot the 4.4.21.81 kernel:

```
# cimage --set ice-sles15spX 4.4.21.81 "r1i[0-2]n[5-6]" "r1i[2-3]n[0-4]"
```

Example 6. The following commands clone the ice-sles15spX image to a new image and modify the new image:

```
# cm image copy -o ice-sles15spX -i mynewimage
                                                          # cloning
Cloning ice-sles15spX to mynewimage ... done              # adds the image and its
                                                          #   kernels to the database
# cp *.rpm /opt/clmgr/image/images/mynewimage/tmp         # copy needed RPMS to
                                                          #    a temp directory
# chroot /opt/clmgr/image/images/mynewimage/ bash         # enter the directory
# rpm -Uvh /tmp/*.rpm                                     # install the RPMs
# exit                                                    # exit the chroot
```

Example 7. If you change the kernels in an image, refresh the kernel database entries for your image.

If you do not change the kernels in the cloned image created, you do not have to refresh the kernel database entries.

The following command refreshes the kernel database entries:

```
# cimage --update-db mynewimage
```

Example 8. The following command pushes new software images to compute blades in a rack or rack set:

```
# cimage --push-rack mynewimage "r*"
r1lead: install-image: mynewimage
r1lead: install-image: mynewimage done.
```

Example 9. The following command specifies that a set of compute nodes to boot an image:

```
# cimage --set mynewimage 4.4.21.81-smp "r1i3n*"
```

Reboot the compute nodes to run the new images.

Example 10. The following command completely removes an image you no longer use. The cluster manager removes the image from the admin node and from all compute nodes in all racks:

```
# cimage --del-image mynewimage
r1lead: delete-image: mynewimage
r1lead: delete-image: mynewimage done.
```

Example 11. You can use the cimage command with its --push-rack option to specify that the command push out only customization scripts to nodes. When you use the --customizations-only option, the command does not include images in the push. In these cases, make sure that you pushed a new image to the nodes before you push the customization scripts. You can specify a list of scripts to run, or you can specify the keyword all to direct the cluster manager to run all scripts.

The command format is as follows:

```
cimage --push-rack --customizations-only [script,script, ...]
```

Or

```
cimage --push-rack --customizations-only all
```

For more information about customization, the following:

Configuring ICE compute node per-host customization on an HPE SGI 8600 cluster

Installing graphics processing unit (GPU) software from Advanced Micro Devices, Inc (AMD)

The example in this topic shows how to download and install the following GPU monitoring software repositories from AMD onto an admin node that is running SLES. This example assumes the following:

- The ROCm repository, which contains the driver and developer tools.
- The AMD GPU repository, which contains additional `gfx` tools. This software is packaged as a compressed archive file. For example, the SLES file, `amdgpu-21.40.2-sle15-main-x86_64.tar.gz`, is available at the following location:

  https://repo.radeon.com/amdgpu/21.40.2/sle/15/main/x86_64/

- The `dkms` repository, which contains compatible Dynamic Kernel Module Support packages. The download file is called `dkms-plus.tar`. It contains a `dkms` package and Perl helper packages.

You can install the software into an image or onto a node.

NOTE: The examples in this topic assume SLES 15 and the AMD ROCm software. If you have different software, modify the instructions as needed.

Procedure

1. Log into the admin node as the root user.
2. Navigate to a directory on the admin node that you can use for downloading the software.

   For example, create directory `/var/tmp` and then change to that directory.

3. Create the ROCm repository.

   For example, if you downloaded the software to `/var/tmp`, enter the following commands:

   ```
   # mkdir -p /opt/clmgr/repos/other/ROCm-4.4-sles15spX
   # cd /opt/clmgr/repos/other/ROCm-4.4-sles15spX
   # tar xvzf /var/tmp/ROCm-4.4-sles15spX.tar.gz
   # cm repo add --custom ROCm-4.4-sles15spX /opt/clmgr/repos/other/ROCm-4.4-sles15spX
   # cm repo select ROCm-4.4-sles15spX
   ```

4. Create the AMD GPU repository.

   For example, if you downloaded the software to `/var/tmp`, enter the following commands:

   ```
   # mkdir -p /opt/clmgr/repos/other/amdgpu-21.40.2-sle15spX
   # cd /opt/clmgr/repos/other/amdgpu-21.40.2-sle15spX
   # tar xvzf /var/tmp/amdgpu-21.40.2-sle15-main-x86_64.tar.gz
   # cm repo add --custom --custom amdgpu-21.40.2-sle15spX
   # cm repo select amdgpu-21.40.2-sle15spX
   ```

5. Create the `dkms-plus` repository.

   For example, if you downloaded the software to `/var/tmp`, enter the following commands:

   ```
   # mkdir -p /opt/clmgr/repos/other/dkms-plus
   # cd /opt/clmgr/repos/other/dksm-plus
   # tar xvf /var/tmp/dksm-plus.tar
   # cm repo add --custom dkms-plus /opt/clmgr/repos/other/dkms-plus
   # cm repo select dkms-plus
   ```

6. Install the AMD GPU software into an image or onto a node.

Example 1. The following command installs the software into a SLES 15 image:

```
# cm image zypper -i <image name> install dkms rocm-dkms rocm-libs \
rocm-validation-suite gcc glibc freeglut-devel kernel-source kernel-default-devel
```

Example 2. The following command assumes that a COS repository is selected. In this case, the installation command is as follows:

```
# cm image zypper -i install cray-rocm-meta gcc glibc freeglut-devel
```

For more information about how to create a COS image, see the following:

Building an HPE Cray operating system (COS) image

Installing graphics processing unit (GPU) software from NVIDIA Corporation

The example in this topic shows how to download the CUDA software from NVIDIA Corporation, create a repository, and install the software. You can install the software into an image or onto a node.

NOTE: The examples in this topic assume SLES 15 and CUDA 11-5. If you have different software releases, modify the instructions as needed.

Procedure

1. Navigate to the website that hosts the CUDA software that you want to install.

   Some software vendors require an account, but to download CUDA from NVIDIA Corporation, you do not need to create an account.

2. Download the software to a central download server at your site or to the admin node.

   For example:

   ```
   # wget https://developer.download.nvidia.com/compute/cuda/11.5.1/\
   local_installers/cuda-repo-sles15-11-5-local-11.5.1_495.29.05-1.x86_64.rpm
   ```

3. Use the `rpm` command to expand the RPM you downloaded into a directory on the admin node.

   For example:

   ```
   # rpm -ivh cuda-repo-sles15-11-5-local-11.5.1_495.29.05-1.x86_64.rpm
   ```

4. Create a directory for the new cluster manager repository.

   For example:

   ```
   # mkdir -p /opt/clmgr/repos/other/cuda/11-5
   ```

5. Copy the RPMs from the NVIDIA repository directory to the HPCM repository directory.

   For example:

   ```
   # cd /var/cuda-repo-sles15-11-5-local/
   # cp -a *.rpm /opt/clmgr/repos/other/cuda/11-5/
   ```

6. Add the data center GPU manager (DCGM) RPM to the repository.

   For example:

   ```
   # cd /opt/clmgr/repos/other/cuda/11-5
   # wget https://developer.download.nvidia.com/compute/cuda/repos/sles15/x86_64/datacenter-gpu-manager-2.3.1-1-x86_64.rpm
   ```

7. Use the `cm repo add` command to create a repository for the new software.

For example:

```
# cm repo add /opt/clmgr/repos/other/cuda/11-5 --custom cuda-11-5
# cm repo select cuda-11-5
```

8. Install the new software into an image or onto a node.

   Example 1. The following command installs the new software onto a node:

```
# # cm node zypper -n gfx1 install gcc glibc freeglut-devel kernel-devel \
kernel-default-devel nvidia-glG0* nvidia-gfxG0*-kmp-default \
nvidia-computeG0* cuda-toolkit-11-5 datacenter-gpu-manager
```

   Example 2. The following command installs the new software into an existing SLES image:

```
# cm image zypper -i sles15spX-cuda install gcc glibc freeglut-devel kernel-devel \
kernel-default-devel nvidia-glG0* nvidia-gfxG0*-kmp-default \
nvidia-computeG0* cuda-toolkit-11-5 datacenter-gpu-manager
```

Using the version control system (VCS)

Node-specific software resides on the admin node. When you install and configure the cluster software, the installer pushes the node-specific software to each node in the cluster. Over time, you might modify the software images. For example, you might add a workload manager or file system software.

If you modify the images frequently, your image repository eventually contains several different versions and becomes difficult to manage. As an alternative to managing these images manually, you can use VCS. The cluster manager version control system (VCS) archives, tracks, and manages the various versions of an image. The cluster manager includes an implementation of VCS that uses the `cm image revision` command.

NOTE: Before you add or modify software, Hewlett Packard Enterprise recommends that you back up the original, default software images.

Do not modify the files within the version control system repository. If you edit files in the VCS repository, the integrity of VCS becomes compromised.

The following topics explain how to use VCS to manage system images:

- **VCS terminology**
- **Image directory**
- **Managing clones**
- **Committing the working copy**
- **Reverting the working copy to a specified revision**
- **Reviewing revision history**
- **Reviewing changes between revisions and the working copy**
- **Amending a commit message**
- **Removing revisions**
- **VCS examples**

VCS terminology

The following terminology pertains to the image files:

- The **working copy** of an image is the copy that is stored on the admin node in the following directory:

  `/opt/clmgr/image/images/`*`image_name`*

  The *image_name* directory contains additional subdirectories and files. The system image includes all the subdirectories and files. The format for the *image_name* directory name is one of the following:

  - `ice-`*`os_name`*. For example, `ice-rhel8.X`. This name of the image for the ICE compute nodes.
  - `lead-`*`os_name`*. For example, `lead-rhel8.X`. This name of the image for on the rack leader controllers.
  - *`os_name`*. For example, `rhel8.X`. This name of the image for compute nodes.
  - `su-lead-`*`os_name`*. For example, `su-lead-rhel8.X`. This name is the name of the image for scalable unit (SU) leader nodes.

When you install cluster software, the installer pushes the working copy image from the admin node to the appropriate nodes in the cluster.

- A **committed copy** of an image is a copy that resides in the VCS repository. It is best to check in, or **commit**, copies of images as you modify them to ensure that modifications are not lost.

Image directory

When you create an image, it resides in the following directory:

```
/opt/clmgr/image/images/image_name
```

In most cases, after the new image is created, the cluster manager sends a copy to the VCS repository and sets the revision number to 1. This set of events is true, for example, after you run the `configure-cluster` command during installation and configuration. However, if you capture an image from a node, the cluster manager does not automatically check in that image.

Managing clones

With VCS, cloning works like creating an image. When cloning, you can use the `cm image revision` command parameter called `--rev` to specify that you want a revision of the image to be the source for the clone.

For information about creating images, see the following:

Image directory

Committing the working copy

After you change the working copy of an image, use the `cm image revision commit` command to commit your changes to VCS. The working copy of an image resides in the following directory:

```
/opt/clmgr/image/images/image_name
```

The commit requires you to enter a log message. You can specify the log message with the `-m` parameter or you can let the command read in the message from the terminal.

Reverting the working copy to a specified revision

To revert the working copy of an image to a specified revision, use the `cm image revision revert` command. This parameter accomplishes the following:

- The parameter removes the working copy of the image.
- The parameter replaces the working copy with a copy of the revision you specify from the VCS repository.

Reviewing revision history

Each time you commit, the cluster manager adds a log message that notes the associated change. Use the following command to list the revision history of an image or of a range of images:

```
cm image revision history -i image
```

Reviewing changes between revisions and the working copy

To display what has changed in an image, use the `cm image revision diff-contents` command. When you do not specify a revision, the cluster manager compares the working copy to the highest version checked in to VCS. The working copy resides in the following directory:

```
/opt/clmgr/image/images/image_name
```

Preceding each file in the list of changed files is an 11-character summary of the differences. For an explanation of the 11-character summary, see the description of the `itemize-changes` option on the `rsync`(1) manpage.

The following information describes various ways to compare revisions:

- If you specify a single revision, the cluster manager compares the working copy to the specified revision.
- If you specify a revision range, the cluster manager displays changes between the two revisions.
- To display file changes in `diff` format, use one of the following parameters:
  - `--file`. The `--file` parameter targets a specific file.
  - `--diff-tool`. The `--diff-tool` parameter lets you specify a specific `diff` tool. Make sure that the tool you specify processes arguments the same way that `diff` command processes arguments.

Amending a commit message

To adjust the commit message of a committed change, use the following command:

```
cm image update -i image --commit-message [--message msg] --revision revision
```

If you do not specify the *msg* parameter, the command reads the message from the terminal.

Removing revisions

The `cm image revision delete` command deletes all stored revisions but leaves the working copy. This parameter can be useful if you want to free space used by revisions or want to start over with the revision history.

The following two commands free all space used in the revision history and then commit a new first revision:

```
# cm image revision delete -i myimage
# cm image revision commit -i myimage -m 'Initial commit'
```

The working copy remains intact and the two revisions effectively collapse into one.

VCS examples

The VCS examples assume that you logged in as the root user. Long output lines are wrapped for inclusion in this documentation.

The following are the example topics:

- **Adding a revision and querying changes**
- **Reverting to a previous revision**
- **Cloning an image**

- **Deleting all revisions permanently**
- **Saving a copy of an image**

Adding a revision and querying changes

The following example shows how to add the file `test_file` to the compute node image `sles15spX`.

Procedure

1. Enter the following command to view the status of image `sles15spX` in the VCS repository:

   ```
   icicle:~ # cm image revision history -i sles15spX
   Revision history for image sles15spX, revisions 1 through 1
   -------------------------------------------------------------------------------
   Revision: 1, Commit Time: Mon 18 September 20xx 11:40:24 AM CDT
   ===============================================================================
   Image created using cinstallman.
   ```

   For each image that you create, the cluster manager automatically adds the image to VCS as revision 1.

2. Enter the following command to add `test_file` to the working copy of the image:

   ```
   icicle:~ # echo "test file" > \
   /opt/clmgr/image/images/sles15spX/tmp/test_file
   ```

3. Enter the following command to show the differences between the working copy and revision 1.

   ```
   icicle:~ # cm image revision diff-contents -i sles15spX
   icicle: cinstallman: Comparing revision 1 and working copy for image sles15spX...
   cmd: rsync -avHix --dry-run --delete /opt/clmgr/image/images//sles15spX/ \
   /opt/clmgr/image/vcs/sles15spX/1/
   sending incremental file list
   .d..t...... tmp/
   >f+++++++++ tmp/test_file

   sent 4961524 bytes  received 16926 bytes  1991380.00 bytes/sec total size is 4015834620
   speedup is 806.64 (DRY RUN)
   ```

   The preceding output shows one difference and the addition of file `test_file`.

4. Enter the following command to commit the new image that contains file `test_file`:

   ```
   icicle:~ # cm image revision commit -i sles15spX -m 'Added test_file to /tmp'
   icicle: cinstallman: vcs: Using rsync to commit image sles15spX...
   cmd: rsync -aqHx --link-dest=/opt/clmgr/image/vcs/sles15spX/1\
    /opt/clmgr/image/images/sles15spX/ /opt/clmgr/image/vcs/sles15spX/2/
   icicle: cinstallman: image sles15spX committed to vcs, rev: 2
   ```

5. Enter the following command to verify that there are no differences between the working copy and the committed copy:

   ```
   icicle:~ # cm image revision diff-contents -i sles15spX
   icicle: cinstallman: Comparing revision 2 and working copy for image sles15spX...
   cmd: rsync -avHix --dry-run --delete /opt/clmgr/image/images//sles15spX/ \
   /opt/clmgr/image/vcs/sles15spX/2/
   sending incremental file list

   sent 4961516 bytes  received 16918 bytes  1991373.60 bytes/sec total size is 4015834611
   speedup is 806.65 (DRY RUN)
   ```

6. Enter the following command to retrieve the revision history of the image:

```
icicle:~ # cm image revision history -i sles15spX
Revision history for image sles15spX, revisions 1 through 2
-------------------------------------------------------------------------------
Revision: 1, Commit Time: Mon 18 September 20xx 11:40:24 AM CDT
===============================================================================
Image created using cinstallman.

Revision: 2, Commit Time: Wed 20 September 20xx 08:17:08 AM CDT
===============================================================================
Added test_file to /tmp

Done
```

7. Enter the following command to display the list of all files changed between revision 1 and revision 2:

```
icicle:~ # cm image revision diff-contents -i sles15spX -rev 1..2
icicle: cinstallman: Comparing revisions 1 and 2 for image sles15spX...
cmd: rsync -avHix --dry-run --delete /opt/clmgr/image/vcs/sles15spX/2/ \
/opt/clmgr/image/vcs/sles15spX/1/
sending incremental file list
>f.st...... etc/opt/sgi/vcs-log-entry
.d..t...... tmp/
>f+++++++++ tmp/test_file

sent 5135898 bytes  received 16931 bytes  3435219.33 bytes/sec total size is 4015834611
speedup is 779.35 (DRY RUN)
```

Notice the presence of the `vcs-log-entry` file, which is always modified upon commits.

Reverting to a previous revision

Assume that you have revised and checked in an image. Later, you decide that you want to revert to a previous image. Use the `cm image revision revert` command to perform the revert.

Procedure

1. Enter the following command to declare that you want `sles15spX` image version 1 software image to be the working copy:

```
icicle:~ # cm image revision revert -i sles15spX --rev 1
icicle: cinstallman: Removing image work dir: /opt/clmgr/image/images/sles15spX
icicle: cinstallman: vcs: Syncing revision 1 in to place...
cmd: rsync -aqHx /opt/clmgr/image/vcs/sles15spX/1/ /opt/clmgr/image/images/sles15spX/
icicle: cinstallman: Working copy of sles15spX now at revision 1
```

2. Enter the following command to retrieve the revision history:

```
icicle:~ # cm image revision history -i sles15spX
Revision history for image sles15spX, revisions 1 through 2
-------------------------------------------------------------------------------
Revision: 1, Commit Time: Mon 18 September 20xx 11:40:24 AM CDT
===============================================================================
Image created using cinstallman.

Revision: 2, Commit Time: Wed 20 September 20xx 08:17:08 AM CDT
===============================================================================
Added test_file to /tmp

Done
```

As the preceding output shows, reverting the working copy does not affect the revision history.

3. To make the working copy correspond to the highest revision (the normal order of things), you can use the following command to commit the current working image (same content as revision 1):

```
icicle:~ # cm image revision commit -i sles15spX \
-m 'Copy of image minus test_file in /tmp.'
icicle: cinstallman: vcs: Using rsync to commit image sles15spX...
cmd: rsync -aqHx --link-dest=/opt/clmgr/image/vcs/sles15spX/2\
 /opt/clmgr/image/images/sles15spX/ /opt/clmgr/image/vcs/sles15spX/3/
icicle: cinstallman: image sles15spX committed to vcs, rev: 3
```

4. Enter the following command to retrieve the revision history:

```
icicle:~ # cm image revision history -i sles15spX
Revision history for image sles15spX, revisions 1 through 3
-------------------------------------------------------------------------------
Revision: 1, Commit Time: Mon 18 September 20xx 11:40:24 AM CDT
===============================================================================
Image created using cinstallman.

Revision: 2, Commit Time: Wed 20 September 20xx 08:17:08 AM CDT
===============================================================================
Added test_file to /tmp

Revision: 3, Commit Time: Wed 20 September 20xx 08:25:39 AM CDT
===============================================================================
Copy of image minus test_file in /tmp.

Done
```

Cloning an image

The following example shows how to clone an image based on one of the previous revision images.

Procedure

1. Enter the following command to clone an image based on revision 2, which includes `test_file`:

```
icicle:~ # cm image copy -o sles15spX -i sles15spX+test_file --rev 2
About to use mksiimage --Copy to clone the image...
icicle: cinstallman: vcs: Syncing revision 2 of sles15spX to new sles15spX+test_file ...
cmd: rsync -aqHx /opt/clmgr/image/vcs/sles15spX/2/ \
/opt/clmgr/image/images/sles15spX+test_file/
icicle: cinstallman: Working copy of sles15spX+test_file now at revision 2 of image sles15spX
Ran sgi-mkautoinstallscript
.
.
.
```

Notice that the `--rev 2` parameter in the preceding command directs the command to create the clone from revision 2.

2. Enter the following command to verify the images that exist on the admin node after the cloning operation:

```
icicle:~ # cm image show
sles15spX
sles15spX+test_file
```

```
lead-sles15spX
ice-sles15spX
```

3. Enter the following command to retrieve the revision history:

```
icicle:~ # cm image revision history -i sles15spX+test_file
Revision history for image sles15spX+test_file, revisions 1 through 1
-------------------------------------------------------------------------------
Revision: 1, Commit Time: Wed 20 September 20xx 08:29:07 AM CDT
===============================================================================
Image clone of sles15spX by cinstallman

Done
```

Deleting all revisions permanently

The following procedure explains how to use the `cm image` command to permanently delete all revisions.

Procedure

1. Enter the following command to delete all revisions:

```
icicle:~ # cm image delete -i sles15spX
Removing all revisions of sles15spX, leaving the working copy...
```

2. Enter the following command to retrieve the revision history and verify the deletion:

```
icicle:~ # cm image revision history -i sles15spX
icicle: cinstallman: There are no checked in revisions of this image.
Image history failed. See above for error messages
```

3. (Optional) Enter the following command to commit the current image to the version control system:

```
icicle:~ # cm image revision commit -i sles15spX -m 'A new beginning :)'
icicle: cinstallman: vcs: First revision, rsync the image work dir to the first revision...
cmd: rsync -aqHx /opt/clmgr/image/images/sles15spX/ /opt/clmgr/image/vcs/sles15spX/1/
```

4. (Optional) Enter the following command to verify the commit:

```
icicle:~ # cm image revision history -i sles15spX
Revision history for image sles15spX, revisions 1 through 1
-------------------------------------------------------------------------------
Revision: 1, Commit Time: Wed 20 September 20xx 08:32:55 AM CDT
===============================================================================
A new beginning :)

Done
```

Saving a copy of an image

Hewlett Packard Enterprise recommends that you back up or save a copy of images before you upgrade or alter an image.

Procedure

1. Retrieve the list of images on the admin node:

```
# cm image show
ice-rhel8.X
lead-rhel8.X
```

```
rhel8.X
rhel8.X-aarch64
```

2. Use the following command to save a copy of the image you want to preserve:

   ```
   cm image revision commit -i current_image -m 'message'
   ```

   For example:

   ```
   # cm image revision commit -i rhel8.X -m 'updated image for new OS'
   ```

3. Verify the revision history.

   For example:

   ```
   # cm image revision history -i rhel8.X
   Revision history for image rhel8.X, revisions 1 through 2
   -------------------------------------------------------------------------------
   Revision: 1, Commit Time: Wed 01 September 20XX 11:03:11 AM CDT
   ===============================================================================
   Image created using cinstallman.

   Revision: 2, Commit Time: Wed 08 September 20XX 02:52:44 PM CDT
   ===============================================================================
   backup_image

   Done
   ```

Image management example

Creating an NFS server on a compute node and mounting a file system into the cluster

This topic provides a fully worked example. The example assumes that all compute nodes run the RHEL operating system. The example creates one compute node to act as an NFS server. The example shows how to enable the other compute nodes in the cluster to mount the `/home` directory through NFS.

The major steps in the example are as follows:

- Creating an image for the NFS server
- Customizing the existing node image for the NFS mount of the `/home` directory
- Provisioning the nodes

Procedure

1. Create an `rpmlist` for the NFS server.

   That is, copy the generated `rpmlist` to an `rpmlist` with a functional name. For example:

   ```
   # cp /opt/clmgr/image/rpmlists/generated/generated-group-rhel8.rpmlist \
   nfsserver-rhel8.rpmlist
   # echo "nfs-utils" >> nfsserver-rhel8.rpmlist
   ```

2. Create the image for the NFS server on one of the compute nodes.

   Use the `cm image create` command with the `rpmlist` from the preceding steps. For example:

   ```
   # cm image create -i nfsserver-rhel8 \
   --rpmlist nfsserver-rhel8.rpmlist --repo-group rhel8
   Repo group rhel8 specified, using repos: HPE-MPI-1.9-rhel8X Cluster-Manager-1.8-rhel8X Red-Hat-Enterprise-
   Linux-8

   An exact bootstrap tarball match was not found.
   Instead, the closest match without being newer than distro was found.

   .
   .
   .
   ```

3. Display the new images.

   For example:

   ```
   # cm image show
   rhel8
   nfsserver-rhel8
   ```

4. Initiate a `chroot` environment and enable the following:

   - The NFS server service
   - The NFS service

For example:

```
# chroot /opt/clmgr/image/images/nfsserver-rhel8/
Error, do this: mount -t proc proc /proc
/bin/basename: missing operand
Try '/bin/basename --help' for more information.
# systemctl enable nfs-server.service
Created symlink /etc/systemd/system/multi-user.target.wants/nfs-
server.service,
pointing to /usr/lib/systemd/system/nfs-server.service.
# systemctl enable nfs.service
```

NOTE: You can ignore the error message that the system generates in response to the preceding command.

5. Change to the `/etc` directory.

   For example:

   ```
   # cd /etc
   ```

6. Create a file called `exports` that lists the servers that you want the NFS server to serve.

   For example:

   ```
   # vi exports
   ```

   The `exports` file can include the following line:

   ```
   /home
   172.23.0.0/255.255.0.0(rw,insecure,sync,no_subtree_check,insecure_locks,no_root_squash)
   ```

7. Exit the `chroot` environment.

   For example:

   ```
   # exit
   exit
   ```

8. Display the difference between the following:

   - The active version of the NFS server image
   - The checked-in version of the NFS server image

   For example:

   ```
   # cm image revision diff-versions -i nfsserver-rhel8
   cb12: cinstallman: Comparing revision 1 and working copy for image nfsserver-rhel8...
   cmd: rsync -avHix --dry-run --delete /opt/clmgr/image/images//nfsserver-rhel8/ /opt/clmgr/image/vcs/nfsserver-
   rhel8/1/
   sending incremental file list
   .d..t...... etc/
   >f.st...... etc/exports
   .d..t...... etc/systemd/system/multi-user.target.wants/
   cL+++++++++ etc/systemd/system/multi-user.target.wants/nfs-server.service -> /usr/lib/systemd/system/nfs-
   server.service
   .d..t...... root/
   >f+++++++++ root/.bash_history
   cd+++++++++ root/.cache/
   cd+++++++++ root/.cache/abrt/
   >f+++++++++ root/.cache/abrt/lastnotification
   cd+++++++++ root/.config/
   cd+++++++++ root/.config/abrt/

   sent 1,861,494 bytes  received 7,821 bytes  1,246,210.00 bytes/sec
   total size is 1,735,609,785  speedup is 928.47 (DRY RUN)
   ```

9. Commit the active version of the NFS server image to VCS.

For example:

```
# cm image revision commit -i nfsserver-rhel8
No --msg specified, reading commit message from STDIN
Enabled nfs on nfsserver image
Created /etc/exports with /home to head network
cb12: cinstallman: vcs: Using rsync to commit image nfsserver-rhel8...
cmd: rsync -aqHXx --link-dest=/opt/clmgr/image/vcs/nfsserverrhel8/1 /opt/
clmgr/image/images/nfsserver-rhel8/ /opt/clmgr/image/vcs/nfsserver-
rhel8/2/
cb12: cinstallman: image nfsserver-rhel8 committed to vcs, rev: 2
```

10. Initiate a `chroot` environment in the compute image, and enable mounting the `/home` directory that the NFS server is exporting.

    For example:

    ```
    # chroot /opt/clmgr/image/images/compute-rhel8/
    Error, do this: mount -t proc proc /proc
    /bin/basename: missing operand
    Try '/bin/basename --help' for more information.
    ```

    NOTE: You can ignore the error message that the system generates in response to the preceding command.

11. In the `/etc` directory, create an `fstab` file.

    For example:

    ```
    # vi /etc/fstab
    ```

    The `fstab` file can include the following line:

    ```
    n0:/home /home nfs defaults 0 0
    ```

12. Exit the `chroot` environment.

    For example:

    ```
    # exit
    exit
    ```

13. Retrieve the current information for the compute image.

    For example:

    ```
    # cm image revision diff-versions -i rhel8
    cb12: cinstallman: Comparing revision 1 and working copy for image rhel8...
    cmd: rsync -avHix --dry-run --delete /opt/clmgr/image/images//rhel8/ /opt/clmgr/image/vcs/
    rhel8/1/
    sending incremental file list
    .d..t...... etc/
    >f+++++++++ etc/fstab
    .d..t...... root/
    >f+++++++++ root/.bash_history

    sent 1,401,838 bytes  received 6,633 bytes  938,980.67 bytes/sec
    total size is 1,466,086,828  speedup is 1,040.91 (DRY RUN)
    ```

14. Use the `cm image revision commit` command to check both images into VCS.

15. Provision all the nodes.

Configuring ICE compute nodes to mount an NFS server on an HPE SGI 8600 cluster

Prerequisites

The following procedure creates custom mount points for the ICE compute nodes.

Procedure

1. Ensure that one of the nodes in the cluster is configured as an NFS server.

   To configure an NFS server, complete the following procedure:

   Creating an NFS server on a compute node and mounting a file system into the cluster

   If you use the procedure in the preceding link, change the `exports` line from `172.23/16` to `InfiniBand/OmniPath` to improve performance.

2. Log into the admin node as the root user.

3. Use the `cd` command to change to the following directory:

   ```
   /opt/sgi/share/per-host-customization/global
   ```

   In the following steps, you complete the following actions:

   - You add the new file system to the `sgi-fstab.sh` file.
   - You ensure that the new file system mounts.

4. Use a text editor to open the following file on the admin node:

   ```
   sgi-fstab.sh
   ```

5. Within file `sgi-fstab.sh`, add a line for a file system mount point, and then save and close the file.

   For example:

   ```
   n0-ib0:/mnt/data /shared nfs defaults 0 0
   ```

6. Enter the following command to push the customization to all the ICE compute nodes:

   ```
   # cimage --push-rack --force --update-only --customizations-only sgi-fstab.sh image_name "r*"
   ```

   For *image_name*, specify the name of the ICE compute image name.

7. Enter the following command to propagate the changes:

   ```
   # clush -g ice-compute mount filesystem
   ```

   For *filesystem*, specify the name of the file system.

Autoinstalling an operating system image on compute nodes

The autoinstall process deploys an operating system image using a pre-scripted configuration file on a functioning compute node or service node. The operating system vendor defines the configuration file format. For example, **kickstart** is the RHEL method, and **autoyast** is the SLES method. Each method provides a defined syntax for scripting an unattended installation of their operating system. For information about the defined syntax, see the vendor documentation.

The target nodes must be nodes with disks that are attached directly to the admin node. For example, these nodes include service nodes and the compute nodes of a flat cluster system.

The following is an overview of how to integrate the autoinstall process with the cluster manager image deployment process:

1. Install an image on a functioning diskful node.

   You can autoinstall any operating system distribution image that is already under cluster manager control. Enter the following command to display a list of operating system distribution images:

   ```
   # cm repo show
   ```

   The cluster manager provides autoinstall template files for various operating system distros. HPE engineers use these template files during internal testing. You can copy and modify the template files to meet the autoinstallation needs of your site. These template files contain keywords that get replaced by appropriate values when autoinstalling a given operating system distribution to a specific node.

   The autoinstallation process includes the cluster manager software automatically. This process ensures that the freshly autoinstalled node is automatically integrated with the cluster manager.

2. Test the installed image to make sure it runs as expected.
3. Use the `cm image capture` command to write a copy of the image back to the admin node. At this point, the image is a standard cluster manager image.
4. From the admin node, run the `cm image provision` command to deploy the image to the compute nodes.

   Do not use autoinstall to provision the following types of nodes:

   - Any diskless compute nodes. For example, do not use autoinstall to provision ICE compute nodes.
   - Scalable unit (SU) leader nodes.
   - Compute nodes configured to run under an SU leader node.
   - ICE leader nodes.

NOTE: The autoinstallation templates include disk partition information that conflicts with the slot management feature. The autoinstall creates its own partition layout as defined in the cluster definition file. When you autoinstall a node, the node loses all data that was in any of the slots on the target disk (or disks) in that node.

For information about how to reimage an autoinstalled compute node with a standard cluster manager image, see the following:

Reimaging an autoinstalled compute node with a standard cluster manager image

---

You can use either the GUI or the command line interface to autoinstall an operating system image. The procedure is as follows:

Procedure

1. **Preparing to autoinstall an operating system image**
2. Complete one of the following procedures:

    - **Autoinstalling with the GUI**

      Or

    - **Autoinstalling with the CLI**

3. **Completing the autoinstall**

Preparing to autoinstall an operating system image

Procedure

1. Verify that the operating system distribution ISO resides in a directory on the admin node.

   If the ISO is already configured in the cluster manager, enter the following command on the admin node:

   ```
   # cm repo show
   ```

   Note the output from the `cm repo` command. If the chosen distribution does not appear in the `cm repo` output, use the `cm repo` command in the following format to copy the ISO to the admin node:

   ```
   cm repo add path_to_ISO_image
   ```

   For *path_to_ISO_image*, specify the full path to where the ISO image resides.

2. Examine the list of autoinstall template files and determine the file you want to use.

   On the admin node, the following directory contains the default autoinstall template files:

   ```
   /opt/clmgr/templates/autoinstall
   ```

   The following notes pertain to the default template files:

    - UEFI-enabled nodes require a separate FAT partition (`/boot/efi`) to boot into the operating system. To autoinstall these nodes, plan to use one of the UEFI-specific autoinstall template files.
    - You can create your own autoinstall template files. Use one of the default files as the base for your own, custom file. Make sure of the following when you create a custom file:
        - The file is compatible with the software release you want to autoinstall.
        - The NFS server and repository information are configured correctly.
    - The autoinstall template files for RHEL are RHEL kickstart files.

      The autoinstall template files for SLES are AutoYaST XML files.
    - The templates support specific operating systems and are as follows:

Table 2: Autoinstall template files

Autoinstall template name	Appropriateness
`autoinst_rh8.templ`	RHEL 9.X, RHEL 8.X on legacy PXE nodes
`autoinst_rh8_uefi.templ`	RHEL 9.X, RHEL 8.X on UEFI-enabled nodes
`autoinst_rh6_rh7.templ`	RHEL 7.X on legacy PXE nodes
`autoinst_rh6_rh7_uefi.templ`	RHEL 7.X on UEFI-enabled nodes
`autoinst_rh6_rh7_moonshot_m710.templ`	RHEL 7.X on ProLiant m710 Moonshot server cartridges
`autoinst_sles15sp2.templ`	SLES 15 SP2 and later on legacy PXE nodes
`autoinst_sles15sp2_uefi.templ`	SLES 15 SP2 and later on UEFI-enabled nodes
`autoinst_sles15.templ`	SLES 15 SP1 and SLES 15 on legacy PXE nodes
`autoinst_sles15_uefi.templ`	SLES 15 SP1 and SLES 15 on UEFI-enabled nodes
`autoinst_sles12.templ`	SLES 12 SPX on legacy PXE nodes
`autoinst_sles12_uefi.templ`	SLES 12 SPX on UEFI-enabled nodes

The autoinstall engine performs keyword substitutions. All keywords begin with `CMU_`. The autoinstall section of the following file describes many of the keywords:

```
/opt/clmgr/etc/cmuserver.conf
```

You can review the autoinstall templates in the following directory:

```
/opt/clmgr/templates/autoinstall
```

The templates show how the cluster manager uses the keywords to create image-specific and node-specific unattended operating system distribution configuration files.

3. (Optional) Modify the autoinstall variables.

   The `/opt/clmgr/etc/cmuserver.conf` configuration file includes an autoinstall section with the following two sets of variables:

   - Variables that affect the autoinstall process behavior:
     - `CMU_AUTOINST_INSTALL_TIMEOUT`

       You can increase this value if the autoinstall times out due to a long disk formatting time.
     - `CMU_AUTOINST_PIPELINE_SIZE`

In many cases, you autoinstall an operating system on only one compute node, and this node becomes the golden node. However, by default, you can autoinstall 16 nodes or fewer simultaneously. To increase this number, edit the `CMU_AUTOINST_PIPELINE_SIZE` variable.

- Variables for keyword substitution into autoinstall templates:
  - `CMU_CN_OS_LANG`

    For example, when you set `CMU_CN_OS_LANG=en_US` in one of the template files, `lang CMU_CN_OS_LANG` becomes `lang en_US`.
  - `CMU_CN_OS_TIMEZONE`
  - `CMU_CN_OS_CRYPT_PWD`
  - `CMU_CN_DEFAULT_GW`

    For example, you can use this variable to specify a per-node default gateway value during autoinstall or cloning. The following settings are available:

    - `default`, which configures the admin node default gateway IP address as the gateway for the autoinstalled node.
    - `cmumgt`, which configures `cmumgt` to set the admin node admin IP address as the gateway for the autoinstalled node.
    - The actual IP address of the gateway

    You can also use the GUI to change the default gateway IP address of a node. For more information, see the following:

    Changing the default gateway IP address of a node from the GUI

4. (Optional) Disable predictable NIC device names.

   Complete this step if you want to install a RHEL image and if your image requires legacy names.

   RHEL-specific predictable NIC device names (`eno1` or `ens1p1`) can be disabled during the autoinstall. Legacy names (`eth0` or `eth1`) can be used.

   Complete the following steps:

   a. Append the kernel command line parameter `net.ifnames=0` to the `CMU_KS_KERNEL_PARMS` in the `/opt/clmgr/etc/cmuserver.conf` file.

   For example:

   ```
   CMU_KS_KERNEL_PARMS="lang=CMU_CN_OS_LANG devfs=nomount ramdisk_size=10240
   console=CMU_CN_SERIAL_PORT ksdevice=CMU_CN_MAC_COLON initrd=autoinst-
   initrd-CMU_IMAGE_NAME net.ifnames=0"
   ```

NOTE: Modifications made to `/opt/clmgr/etc/cmuserver.conf` apply to all the RHEL autoinstall image groups created from that point forward.

To disable predictable NIC names for a specific RHEL autoinstall group, edit the `pcmlinux_template` file in the following directory:

`/opt/clmgr/image/`*`group_name`*

Add `net.ifnames=0` to the existing kernel parameters list.

b. Modify the autoinstall template to ensure that the `net.ifnames=0` parameter is persistent during the subsequent disk boots.

To the `bootloader` line, append `--append='net.ifnames=0'`.

For example:

```
#System bootloader configuration bootloader --location=mbr --
append='net.ifnames=0'
```

5. (Conditional) Restrict RHEL operating system partitions to one disk or LUN.

Complete this step if you want to autoinstall RHEL on a node that contains multiple disks or LUNs.

By default, the RHEL installer spreads the operating system installation across multiple disks or LUNs. As a result, certain operating system partitions, such as `/boot` and `/`, can spread across multiple disks or LUNs.

To confine the operating system installation to a single disk or LUN, edit the following file:

`/opt/clmgr/image/`*`group_name`*`/autoinst.tmpl-orig`

In the file, pass the `--ondisk=`*`disk_id`* option to the partitioning commands.

For example:

```
#Disk partitioning information
#USE THE APPROPRIATE DISK NAME
part /boot --fstype ext4 --size 1000  --asprimary --ondisk=sda
part swap --size 4096 --asprimary  --ondisk=sda
part / --fstype ext4 --size 1 --grow --asprimary --ondisk=sda
```

NOTE: The alternative autoinstall command `ignoredisk --only-use=sda` can be used instead of specifying the `--on-disk=sda` option for every partition command.

6. Ensure that the target node bootloader is set to `grub2`.

The autoinstallation process requires `grub2` to begin the autoinstallation process. Enter the `cm node set` command in the following format to ensure that the target node is configured for `grub2`:

```
cm node set -n node --diskbootloader grub2
```

For *node*, specify one or more node hostnames.

Autoinstalling with the GUI

Prerequisites

Preparing to autoinstall an operating system image

Procedure

1. Click **Options** > **Enter Admin mode** and enter the cluster credentials.
2. Click **Cluster Administration** > **Image Group Management**.
3. On the **New Image Group** popup window, complete the fields as follows and click **OK**:

   - In the **Group name** field, enter a name.

     This name becomes a directory in `/opt/clmgr/image`.

     For example, assume that you specify **`rh7u4_autoinstall`** in the **Group name** field. The cluster manager creates the following files in the `/opt/clmgr/image/rh7u4_autoinstall` directory:

     - `autoinst.tmpl.orig` - An exact copy of the autoinstall file
     - `repository` - A logical link to the autoinstall repository
     - README - More information about node-specific customization of autoinstall and PXE boot files

   - In the **Group type** field, select **autoinstall**.
   - In the **Repository path** field, enter the path to the location of the operating system distribution files or click the folder icon to browse.

     To retrieve the path to the operating system on the admin node, enter the following command:

     ```
     # cm repo show
     ```

   - In the **Autoinstall template** field, enter the path to the autoinstall template you want to use or click the folder icon to browse.
   - (Conditional) In the **Installer ISO** field, specify the full path to the SLES 15 SP1 or the SLES 15 `*Installer*.iso` file.

     Complete this step if installing SLES 15 SP1 or SLES 15.

     This file contains the appropriate AutoYaST tools need to autoinstall the SLES 15 SP1 or SLES 15 distribution.

4. In the left pane, left-click an image group and then right-click to select **Autoinstall (Kickstart | AutoYaST | Preseed | Unattended)**.

   When the autoinstall finishes, new files are added to the following directory:

   `/opt/clmgr/image/`*`group_name`*

   The new files are as follows:

   - `autoinst.tmpl-cmu` - A copy of your autoinstall file with additional required directives.
   - `autoinst-`*`compute_node_hostname`* - The autoinstall template with hard-coded node-specific information.
   - `pcmlinux_template` - The PXELINUX boot parameter file template for this image group.
   - `pcmlinux-`*`compute_node_hostname`* - The PXELINUX boot parameter file for a specific node.

   The autoinstall process boots the selected compute nodes over the network and initiates a typical RHEL Kickstart or SLES AutoYaST installation. During the operation, the autoinstall log displays on the terminal.

Autoinstalling with the CLI

Prerequisites

Preparing to autoinstall an operating system image

Procedure

1. Log into the admin node as the root user.

2. Enter the `cm repo show` command to determine the path to the operating system repository.

   For example:

   ```
   # cm repo show
   * Cluster-Manager-1.8-rhel8X : /opt/clmgr/repos/cm/Cluster-Manager-1.8-rhel8X
   * Red-Hat-Enterprise-Linux-8.X : /opt/clmgr/repos/distro/rhel8.X
     SLE-15-SPX-Full-x86_64 : /opt/clmgr/repos/distro/sles15spX
   ```

   The directory path output from this step is an input parameter for Step **4** in this procedure.

3. (Conditional) Add the operating system image to the admin node.

   Complete this step if the operating system ISO you want to use does not appear in the `cm repo` output.

   Use the `cm repo` command to add the ISO.

   For example:

   ```
   # cm repo add SLE-15-SPX-Full-x86_64-GM-Media1.iso
   Mounting ISO file loopback...
     Running: cp -a /tmp/ZSwuU1XTAO /opt/clmgr/repos/distro/sles15spX
   Exporting repository for use with yume....
   Exporting /opt/clmgr/repos/distro/sles15spX through httpd, http://node64/repo/opt/clmgr/repos/distro/sles15spX
   Updating default rpm lists...
   Updating: /opt/clmgr/image/rpmlists/generated/generated-rhel8.X.rpmlist
   Updating: /opt/clmgr/image/rpmlists/generated/generated-ice-rhel8.X.rpmlist
   Updating: /opt/clmgr/image/rpmlists/generated/generated-lead-rhel8.X.rpmlist
   Updating: /opt/clmgr/image/rpmlists/generated/generated-admin-rhel8.X.rpmlist
   ```

4. Use the `cm image create` command at the admin node prompt to create the image.

   Use the directory path output from the `cm repo show` command in Step **2** in this procedure as input to the command you enter in this step.

   ```
   # cm image create -i rh8uX_autoinst \
   -r /opt/clmgr/repos/distro/rhel8.X-x86_64 \
   -a /opt/clmgr/templates/autoinstall/autoinst_rh8_uefi.templ
   ```

   NOTE: To install the SLES 15 SP1 or SLES 15 distribution, also specify the `–I` option on the `cm image create` command line. This option specifies the path to the installation `.iso` file. For example:

   ```
   # cm image create -i sles15sp1_autoinst \
   -r /opt/clmgr/repos/distro/sles15sp1-x86_64 \
   -a /opt/clmgr/templates/autoinstall/autoinst_sles15_uefi.templ \
   -I /root/SLE-15-SP1-Installer-DVD-x86_64-GM-DVD.iso
   ```

5. Log into the admin node as the root user.

6. (Optional) Use a text editor to create a file called `nodes.txt`.

   If you have many nodes, consider creating a text file that contains a list of nodes.

   Within the file, include one node hostname per line.

If you do not create a text file, you need to specify the node hostnames in a comma-separated list.

7. Run the `cm node provision` command at the admin node prompt.

   If the node hosts an autoinstalled image at this time, see the following for information about additional parameters to include on the command line:

   Reimaging an autoinstalled compute node with a standard cluster manager image

   Example 1:

   ```
   # cm node provision -f nodes.txt -i rh7u7_autoinst
   ```

   Example 2:

   ```
   # cm node provision -n n0,n1,n2 -i rh7u7_autoinst
   ```

Completing the autoinstall

When the autoinstall process is completes successfully, the following events occur:

- The node or nodes reboot into the installed operating system.
- The public `ssh` key is installed on the target node(s). This enables the root user to `ssh` to each node.

If the autoinstall process does not complete successfully, and you suspect that there is an error in the autoinstall file, create one or more node-specific custom prefixes or custom autoinstall files in the autoinstall group directory. For example, to create a PXELINUX file specific to node `n1`, create the following file, and customize the file according to the guidelines in the `/opt/clmgr/image/`*`group_name`*`/README` file:

```
/opt/clmgr/image/group_name/pcmlinux-n1.custom
```

Complete the procedure in this topic to install the rest of the cluster manager credentials and other certificates.

Prerequisites

Autoinstalling with the GUI

Or

Autoinstalling with the CLI

Procedure

1. Log into the admin node as the root user.
2. Enter the following command:

   ```
   /opt/clmgr/tools/hpcm_push_secrets -t compute -n node
   ```

   For *node*, specify one or more autoinstalled node hostnames.

   The `hpcm_push_secrets` command might prompt you to enter the node password several times. After the `hpcm_push_secrets` command finishes, passwordless `ssh` is implemented on the node.
3. Use the following command to reboot the node:

   ```
   cm power reboot -n node
   ```
4. (Conditional) Repeat the preceding steps for other autoinstalled nodes.

Reimaging an autoinstalled compute node with a standard cluster manager image

If an autoinstalled image resides on a compute node, and you want to reimage the node with a standard cluster manager node image, include the `--force-disk` and `--wipe-disk` parameters on the `cm node provision` command.

For example:

```
# cm node provision -i IMAGE1 -n NODE5 --force-disk=/dev/sda --wipe-disk
```

Because the autoinstall process installs to a disk, it might set boot to disk first in the boot order. This practice conflicts with cluster manager boot practices, which sets nodes to PXE boot first, thereby providing more control over the boot process. When you deploy an image to a previously autoinstalled node, monitor the console and change the boot order so that the node PXE boots first. If the node's boot mode is UEFI, log into the autoinstalled node and use the efibootmgr command to change the boot order to request that the node PXE boot first.

Fabric management for HPE Slingshot deployments

The connective software that supports the cluster topology is a fabric network. The fabric network includes internal switches specific to one type of fabric network. These switches are located in the chassis. Each system is configured with one or two separate **subnets**. For HPE Slingshot switches, the documentation refers to these subnets as `hsn`*`num`*. For example, `hsn0`, `hsn1`, and so on.

For more information about the HPE Slingshot interconnect, see the HPE Slingshot documentation.

Fabric management for InfiniBand deployments and Omni-Path Express deployments

Each cluster implements one of the following network topologies:

- Hypercube
- Enhanced Hypercube
- All-to-All
- Fat Tree

The connective software that supports the cluster topology is a fabric network. The fabric network includes internal switches specific to one type of fabric network. These switches are located in the chassis. Each system is configured with one or two separate **fabrics** or **subnets**. For InfiniBand switches and Omni-Path Express switches, the documentation refers to these subnets as `ib0` and `ib1`.

On compute nodes, there might be several interfaces called `ib0`, `ib1`, and so on. Each interface might be connected to the same subnet.

The fabric network is one of the following:

- Omni-Path Express fabric network

  The cluster manager supports the Omni-Path Express fabric network and includes tools to manage the Omni-Path Express fabric.

  For more information about the Omni-Path Express fabric network, see the documentation from Cornelis Networks.

- InfiniBand Fabric Network

  The InfiniBand technology facilitates fast communication between the compute nodes within a rack and the compute nodes in separate racks. The InfiniBand network uses Open Fabrics Enterprise Distribution (OFED) software to monitor and control the InfiniBand fabric. For information about OFED, see the following link:

  https://www.openfabrics.org .

  The cluster system uses a distributed memory scheme. Parallel processes in an application pass messages, and each process has its own dedicated processor and address space. By default, the HPE Message Passing Interface (MPI) software uses only the `ib0` subnet. Typically, storage uses the `ib1` subnet. Other InfiniBand configurations are possible and can lead to better performance with specific workloads. For example, you can configure the MPI from HPE library, the Message Passing Toolkit (MPT), to use one or two InfiniBand subnets to optimize application performance.

Installing the InfiniBand software stack (OFED) from RHEL on cluster systems

Additional software is required to use any of the following utilities:

- InfiniBand fabric network utilities
- OpenSM subnet manager and administrator
- Omni-Path Express fabric network

The following procedure explains how to install the additional RHEL packages.

Procedure

1. Log into the admin node as the root user.
2. Use the following `cm image` command to install the RHEL InfiniBand software packages:

   ```
   cm image dnf -i image.rdma group install infiniband
   ```

   For *image*, specify the correct RHEL version.

   For example:

   ```
   # cm image dnf -i rhel8.5.rdma  group install infiniband
   ```

3. Use the following `cm image` command to install the RHEL InfiniBand software packages, along with the OpenSM package, on leader nodes and (optionally) service nodes:

   ```
   cm image dnf -i image.rdma group install --with-optional infiniband
   ```

   For *image*, specify the correct RHEL version.

   For example:

   ```
   # cm image dnf -i rhel8.5.rdma  group install --with-optional infiniband
   ```

4. Use the following `cm image` command to install the `opensm-multifabric` package for dual-pane fabrics on leader nodes and (optionally) service nodes:

   ```
   cm image dnf -i image.rdma install opensm-multifabric
   ```

   For *image*, specify the correct RHEL version.

   For example:

   ```
   # cm image dnf -i rhel8.5.rdma install opensm-multifabric
   ```

Installing the InfiniBand software stack (OFED) from SLES on cluster systems

Additional software is required to use any of the following utilities:

- InfiniBand fabric network utilities
- OpenSM subnet manager and administrator
- Omni-Path Express fabric network

The following procedure explains how to install the additional SLES packages.

Procedure

1. Log into the admin node as the root user.
2. Use the following `cm image` command to install the SLES OFED software packages:

   ```
   cm image zypper -i image.rdma install patterns-ofed-ofed
   ```

   For *image*, specify the correct SLES version.

For example:

```
# cm image zypper -i sles15sp3.rdma install patterns-ofed-ofed
```

3. Use the following `cm image` command to install the OpenSM package for the leader nodes and (optionally) service nodes:

   ```
   cm image zypper -i image.rdma install opensm
   ```

   For *image*, specify the correct SLES version.

   For example:

   ```
   # cm image dnf -i sles15sp3.rdma install opensm
   ```

4. Use the following `cm image` command to install the `opensm-multifabric` package for dual-pane fabrics for leader nodes and (optionally) service nodes:

   ```
   cm image zypper -i image.rdma install opensm-multifabric
   ```

   For *image*, specify the correct SLES version.

   For example:

   ```
   # cm image zypper -i sles15sp3.rdma install opensm-multifabric
   ```

Installing the InfiniBand software stack from NVIDIA Mellanox on RHEL cluster systems

The procedure in this topic explains how to install the MLNX OFED software that enables the InfiniBand fabric.

Procedure

1. Log into the admin node as the root user.

2. Verify that the MLNX-OFED `tar` file resides on the admin node and is unpacked.

3. Use the `mkdir` command to create a software repository directory for the InfiniBand software stack.

   For example:

   ```
   # mkdir -p /opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64
   ```

4. Copy the MLNX-OFED RPMs to the repository directory.

   The RPM files are typically found in the RPMS directory in the unpacked MLNX OFED `.tgz` file.

   Foe example:

   ```
   # cp *.rpm /opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64/
   ```

5. Initiate a `chroot` environment and prepare the target image for the installation of the MLNX-OFED packages.

   ```
   chroot target-image rpm -e --nodeps --noscripts libibverbs fio librdmacm
   ```

   For *target-image*, specify `/opt/clmgr/image/images/image_name`, which is the directory where the cluster manager image resides.

6. Use the `cm repo add` command, with the `--custom` parameter, to create a repository for the NVIDIA Mellanox software.

For example:

```
# cm repo add --custom mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64 \
/opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64
```

7. (Optional) Use the `cm repo show` command to display the repositories.

8. Use the `cm repo group add` command to create a repository group for the new NVIDIA Mellanox MLNX-OFED repository.

   Create a repository group with the minimum of software needed. For example, include the cluster manager, RHEL, and HPE MPI. Avoid adding update repositories to this repository group.

   ```
   # cm repo group add pcm180-rhel86-x86_64-mlnx-ofed-5.6-2.0.9.0-rhel8.6 \
   --repos Cluster-Manager-1.8-rhel86-x86_64 HPE-MPI-1.9.2-rhel86-x86_64 mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64 \
   Red-Hat-Enterprise-Linux-8.6.0-x86_64
   ```

9. Install the NVIDIA Mellanox MLNX-OFED packages into the image.

   For example:

   ```
   # cm image refresh -i rhel8.6.ofed --repo-group pcm180-rhel86-x86_64-mlnx-ofed-5.6-2.0.9.0-rhel8.6 \
   --rpmlist ./sort-mlnx-ofed-5.6-2.0.9.0-rhel8.6-x86_64-service.rpmlist
   ```

Installing the InfiniBand software stack from NVIDIA Mellanox on SLES cluster systems

The procedure in this topic explains how to install the MLNX OFED software that enables the InfiniBand fabric.

Procedure

1. Log into the admin node as the root user.

2. Verify that the MLNX-OFED `tar` file resides on the admin node and is unpacked.

3. Use the `mkdir` command to create a software repository directory for the InfiniBand software stack.

   For example:

   ```
   # mkdir -p /opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64
   ```

4. Copy the MLNX-OFED RPMs to the repository directory.

   The RPM files are typically found in the RPMS directory in the unpacked MLNX OFED `.tgz` file.

   For example:

   ```
   # cp RPMS/*.rpm /opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64/
   ```

5. Use the `cm repo add` command, with the `--custom` parameter, to create a repository for the NVIDIA Mellanox software.

   For example:

   ```
   # cm repo add --custom mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64 \
   /opt/clmgr/repos/other/mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64
   ```

6. (Optional) Use the `cm repo show` command to display the repositories.

7. Use the `cm repo group add` command to create a repository group for the new NVIDIA Mellanox MLNX-OFED repository.

   Create a repository group with the minimum of software needed. For example, include the cluster manager, SLES, and HPE MPI. Avoid adding update repositories to this repository group.

For example:

```
# cm repo group add pcm180-sles15sp4-x86_64-mlnx-ofed-5.6-2.0.9.0 \
Cluster-Manager-1.8-sles15sp4-x86_64 \
HPE-MPI-1.9.2-sles15sp4-x86_64 \
mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64 SLE-15-SP4-Full-x86_64
```

8. Install the NVIDIA Mellanox MLNX-OFED packages into the image.

   For example:

   ```
   # cm image refresh -i sles15sp4.ofed --repo-group mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64 \
   --rpmlist mlnx-ofed-5.6-2.0.9.0-sles15sp4-x86_64-leader.rpmlist
   ```

Initializing the Omni-Path Express fabric manager

Complete this procedure as follows:

- On clusters with ICE leader nodes, complete the procedure in this topic from each ICE leader node.
- On clusters without leader nodes, complete the procedure in this topic from the admin node.

The following procedure explains how to initialize the Omni-Path Express fabric.

Procedure

1. Log into one of the following nodes as the root user:

   - On clusters with ICE leader nodes, log into one of the leader nodes. For example, on an HPE SGI 8600 cluster, log into `r1lead`.
   - On clusters without ICE leader nodes, log into the admin node.

2. Edit the `opafm.xml` file to enable logging.

   Complete the following steps:

   - Open the following file within a text editor:

     ```
     /etc/opa-fm/opafm.xml
     ```

   - Within `opafm.xml`, search for the following line:

     ```
     <!-- <LogFile>/var/log/fm0_log/<LogFile> --> <!-- log for this instance -->
     ```

   - Remove the following character strings:

     - `<!--`
     - `--!>`

     The deletions leave the line as follows:

     ```
     <LogFile>/var/log/fm0_log/<LogFile> <!-- log for this instance -->
     ```

     The characters you remove are XML comment characters, and they are not needed.

     Keep the file open so you can add information.

3. In another window, use the `ibstat` command to display the port globally unique identifiers (GUIDs) for the node:

For example:

```
r1lead:~ # ibstat -p
0x001175010163ce82
0x001175010163cea3
```

The first line is the port GUID for `fm0`. The second line is the port GUID for `fm1`.

4. Add port GUIDs to the fabric manager instances in the `opafm.xml` file, as follows:

   - Within `opafm.xml`, search for the following:

     ```
     <Name>fm0</Name>
     ```

   - Within the block of text that describes `fm0`, locate the following string:

     ```
     <PortGUID>0x0000000000000000</PortGUID>
     ```

   - Replace `0x0000000000000000` with the GUID string for `fm0`, which is the first line of `ibstat` command output.
   - Within `opafm.xml`, search for the following:

     ```
     <Name>fm1</Name>
     ```

   - Within the block of text that describes `fm1`, locate the following string:

     ```
     <PortGUID>0x0000000000000000</PortGUID>
     ```

   - Replace `0x0000000000000000` with the GUID string for `fm1`, which is the second line of `ibstat` command output.
   - Remain within the block of text that describes `fm1`, and make the following additional changes:
     - Change `<Start>0</Start>` to `<Start>1</Start>`
     - Change `<Hfi>1</Hfi>` to `<Hfi>2</Hfi>`
     - Change `<Port>2</Port>` to `<Port>1</Port>`
     - Change `<SubnetPrefix>0xfe80000000001001</SubnetPrefix>` to `<SubnetPrefix>0xfe80000000000001</SubnetPrefix>`.

NOTE: It is possible that some of the file already contains some of the correct values. In this case, verify that the correct values are present.

5. Configure the fabric topology.

   Choose the appropriate routing engine for the fabric topology.

   The default for the `opafm.xml` file is as follows:

   ```
   <RoutingAlgorithm>shortestpath</RoutingAlgorithm>
   ```

   For a hypercube or enhanced hypercube topology, use the following line:

   ```
   <RoutingAlgorithm>hypercube</RoutingAlgorithm>
   ```

   For a Fat Tree topology, use the following line:

   ```
   <RoutingAlgorithm>fattree</RoutingAlgorithm>
   ```

6. Save and close file /etc/opa-fm/opafm.xml.

7. Enter the following command to start the fabric manager:

```
# systemctl start opafm
```

Configuring the InfiniBand fabrics on clusters with ICE leader nodes

Each cluster has two InfiniBand fabric network cards, ib0 and ib1. Each subnet has a subnet manager. Some InfiniBand switches are preconfigured for an InfiniBand subnet. If a cluster uses Mellanox OFED, you can use the configure-cluster menu system to configure the subnet manager.

Complete the procedure in this topic in the following situations:

- If the cluster switch is not preconfigured for InfiniBand.
- If you want to move the InfiniBand subnet to a different node.

  The subnet manager can run on an ICE leader node or on a compute node that is deployed as a service node.

The following procedure explains how to configure the master and the standby components and how to verify the configuration.

Procedure

1. Through an ssh connection, log into the admin node as the root user.

2. Enter the following command to disable InfiniBand switch monitoring:

   ```
   # cattr set disableIbSwitchMonitoring true
   ```

   The system sometimes issues InfiniBand switch monitoring errors before the InfiniBand network has been fully configured. The preceding command disables InfiniBand switch monitoring.

3. Use one of the following methods to access the InfiniBand network configuration tool:

   - Enter the following command to start the cluster configuration tool:

     ```
     # configure-cluster
     ```

     After the cluster configuration tool starts, select **F Configure InfiniBand Fabric**, and select **OK**.

   - Enter the following command to start the InfiniBand management tool:

     ```
     # tempo-configure-fabric
     ```

   Both of the preceding methods lead you to the same InfiniBand configuration page. On the InfiniBand configuration pages, **Quit** takes you to the previous screen.

4. Select **A Configure InfiniBand ib0**, and select **Select**.

5. On the **Configure InfiniBand** screen, select **A Configure Topology**, and select **Select**.

6. On the **Topology** screen, select the topology your system uses, and select **Select**.

   The menu selections are as follows:

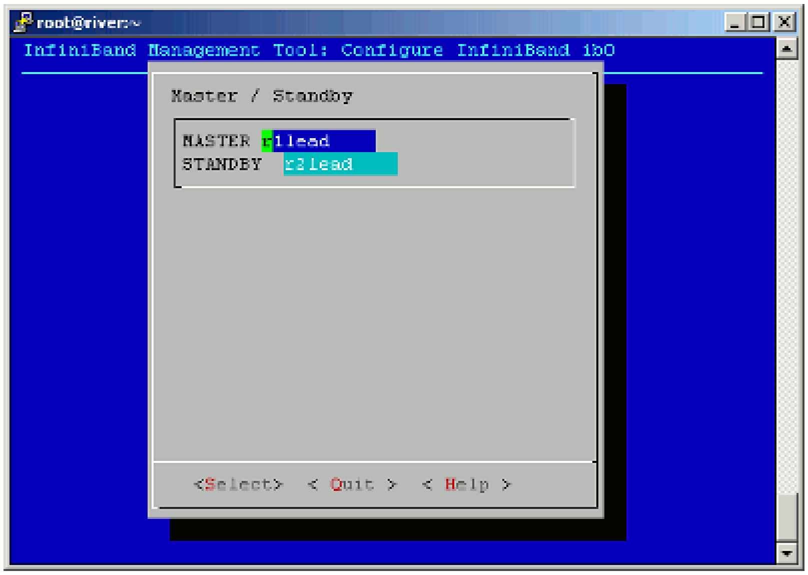
- **H HYPERCUBE**
- **E EHYPERCUBE** (Enhanced Hypercube)
- **F FAT TREE**
- **G BFTREE** (Balanced Fat Tree)

7. On the **Configure InfiniBand** screen, select **B Master / Standby**, and select **Select**.
8. On the **Master / Standby** screen, enter the component identifiers for the master (primary) and the standby (backup, secondary) subnet, and select **Select**.

   The nodes you select for the **MASTER** and **STANDBY** must have InfiniBand cards.

   For example, if you have only one ICE leader node, type the leader node hostname in the **MASTER** field, and leave the **STANDBY** field blank. If you have more than one leader node, specify different leader nodes in the **MASTER** and **STANDBY** fields.

   The following shows a completed screen for a cluster with ICE leader nodes:



```
InfiniBand Management Tool: Configure InfiniBand ib0

Master / Standby

MASTER r1lead
STANDBY r2lead
```

Figure 13: Completed InfiniBand (ib0) Master / Standby Screen

9. On the **Configure InfiniBand** screen, select **Commit**.

   Wait for the confirmatory messages to appear in the window before you continue.

   The next few steps repeat the preceding steps, but this time you configure the `ib1` interface.
10. On the InfiniBand Management Tool main menu screen, select **B Configure InfiniBand ib1**, and select **Select**.
11. On the **Configure InfiniBand** screen, select **A Configure Topology**, and select **Select**.
12. On the **Topology** screen, select the topology your system uses, and select **Select**.

    Select the topology that exists on your system. The menu selections are as follows:

    - **H HYPERCUBE**
    - **E EHYPERCUBE** (Enhanced Hypercube)

- **F FAT TREE**
- **G BFTREE**

13. On the **Configure InfiniBand** screen, select **B Master / Standby**, and select **Select**.

14. On the **Master / Standby** screen, enter the component identifiers for the master (primary) and the standby (backup, secondary) subnet, and select **Select**.

    For example, if you have only one ICE leader node, type the leader node hostname in the **MASTER** field, and leave the **STANDBY** field blank. If you have two leader nodes, you can flip the specifications for `ib1`. For example, assume that for `ib0`, you specified **MASTER** as `r1lead` and **STANDBY** as `r2lead`. For `ib1`, you can specify **MASTER** as `r2lead` and **STANDBY** as `r1lead`. If you have three or more leader nodes, specify different leader nodes in the **MASTER** and **STANDBY** fields.

15. On the **Configure InfiniBand** screen, select **Commit**.

    Wait for the confirmatory messages to appear in the window before you continue.

16. On the InfiniBand Management Tool main menu screen, select **C Administer Infiniband ib0**, and select **Select**.

17. On the **Administer InfiniBand** screen, select **Start**, and select **Select**.

18. On the **Start SM master_ib0 on ib0 succeeded!** screen, select **OK**.

19. Select **Quit** to return to the InfiniBand Management Tool main menu screen.

    The next few steps repeat the preceding steps, but this time you start the `ib1` interface.

20. On the InfiniBand Management Tool main menu screen, select **D Administer Infiniband ib1**, and select **Select**.

21. On the **Administer InfiniBand** screen, select **Start**, and select **Select**.

22. On the **Start SM master_ib1 on ib1 succeeded!** screen, select **OK**.

23. On the **Administer InfiniBand** screen, select **Status**, and select **Select**.

    The **Status** option returns information similar to the following:

    ```
    Master SM
    Host = r1lead
    Guid = 0x0002c9030006938b
    Fabric = ib0
    Topology = hypercube
    Routing Engine = dor
    OpenSM = running
    ```

24. Wait for the status messages to stop, and press **Enter**.

25. Select **Quit** on the menus that follow to exit the configuration tool.

26. Use the `ssh` command to log into one of nodes you specified as **MASTER** or **STANDBY**.

27. Use the `ibstatus` command to retrieve the status information for this node.

    For example:

    ```
    # ibstatus
    Infiniband device 'mlx4_0' port 1 status:
        default gid:  fec0:0000:0000:0000:0002:c903:00f3:5311
        base lid:     0x1
        sm lid:       0x1
    ```

```
        state:         4: ACTIVE
        phys state:    5: LinkUp
        rate:          56 Gb/sec (4X FDR)
        link_layer:    InfiniBand

    Infiniband device 'mlx4_0' port 2 status:
        default gid:   fec0:0000:0000:0001:0002:c903:00f3:5312
        base lid:      0x2a
        sm lid:        0x2a
        state:         4: ACTIVE
        phys state:    5: LinkUp
        rate:          56 Gb/sec (4X FDR)
        link_layer:    InfiniBand
```

The output shows the status as `ACTIVE` on both ports, which is correct.

28. Log into one of the nodes that is linked to the fabric, and use the `ibhosts` command to display the nodes on the fabric.

Use this list to verify that the configured nodes are connected to the fabric. Run the `ibhosts` command from a node that is linked to the fabric.

For example:

```
n101:~ # ibhosts
Ca    : 0x0002c9030032bd50 ports 1 "r1i0n8 HCA-1"
Ca    : 0x0002c9030014a630 ports 1 "r1i0n7 HCA-1"
Ca    : 0x0002c9030014b140 ports 1 "r1i0n6 HCA-1"
Ca    : 0x0002c9030018cf00 ports 1 "r1i0n5 HCA-1"
Ca    : 0x0002c9030018cfa0 ports 1 "r1i0n13 HCA-1"
Ca    : 0x0002c9030018ce90 ports 1 "r1i0n15 HCA-1"
Ca    : 0x0002c9030014a610 ports 1 "r1i0n14 HCA-1"
Ca    : 0x0002c9030014b110 ports 1 "r1i0n16 HCA-1"
Ca    : 0x0002c9030018d170 ports 1 "r1i0n17 HCA-1"
```

The output shows each node connected, as expected.

29. (Conditional) Increase the sweep interval to 90 seconds or more.

Complete this step if the cluster has more than 256 nodes.

Use a command such as the following:

```
# sgifmcli --set --id master-ib0 --arglist sweep-interval=90
```

Configuring InfiniBand fabric software manually

Hewlett Packard Enterprise strongly recommends that you use the automated tools provided in the cluster manager to configure the InfiniBand fabric software.

The recommended configuration method uses the InfiniBand management tool, which uses a GUI. To start the tool, enter the following command from the admin node:

```
# tempo-configure-fabric
```

The following procedure explains how to configure the fabric without using the GUI.

Procedure

1. **Configuring a master fabric**
2. **Enabling the InfiniBand fabric failover mechanism**
3. **(Conditional) Configuring the InfiniBand fat-tree network topology**

Configuring a master fabric

When configuring the subnet manager master, the following rules apply:

- Log into the admin node to run the `sgifmcli` commands.
- Each InfiniBand fabric must have a subnet manager master.
- There can be at most one subnet manager master per InfiniBand fabric.
- Fabric configuration and administration can be done only through the subnet manager master.
- Fabric configuration becomes active after (re)starting the subnet manager master.
- If there is a standby, the action of deleting a subnet manager master automatically deletes the standby.

Procedure

1. Use the `sgifmcli` command to configure a subnet manager master.

   The format for the `sgifmcli` command to configure a subnet manager master is as follows:

   ```
   sgifmcli --mastersm --init --id identifier --hostname hostname --fabric fabric --topology topology
   ```

   The variables are as follows:

Variable	Specification
identifier	Any arbitrary string. The `--id` option creates a master with the name you supply.
hostname	The host from which you want the subnet manager master to launch.
fabric	Either `ib0` or `ib1`.
topology	One of the following: • `hypercube` • `enhanced-hypercube` • `ftree` • `balanced-ftree`.

For example, on a cluster with ICE leader nodes, the following command configures a master for fabric ib0 on a hypercube cluster:

```
# sgifmcli --mastersm --init --id master_ib0 --hostname r1lead \
--fabric ib0 --topology hypercube
```

2. Repeat the preceding step for each fabric you want to create.

Enabling the InfiniBand fabric failover mechanism

Hewlett Packard Enterprise recommends that you configure a failover subnet manager. If the master subnet manager fails, the standby subnet manager takes over operation of the fabric. The opensm software performs this failover operation automatically.

For example, on a cluster with ICE leader nodes, typically, rack1 is the MASTER for the ib0 fabric and rack2 has the MASTER for the ib1 fabric.

When you enable the InfiniBand failover mechanism, observe the following rules:

- As an option, each InfiniBand fabric can have exactly one standby.
- If a master subnet manager exists, you can create a standby subnet manager.
- When adding a standby after a master has already been defined and started, stop the master and then use the --init option to define the standby. After you define the standby, restart the master.
- A subnet manager master and subnet manager standby for a particular fabric cannot coexist on the same node.

The following procedure describes how to set up the failover mechanism.

Procedure

1. Stop any subnet manager masters that are defined and running.

   For example, use the following command:

   ```
   # sgifmcli --stop --id master_ib0
   ```

2. Define the subnet manager standby.

   Example 1, for a cluster without leader nodes:

   ```
   # sgifmcli --standbysm --init --id standby_ib0 \
   --hostname service1 --fabric ib0
   ```

   Example 2, for a cluster with ICE leader nodes:

   ```
   # sgifmcli --standbysm --init --id standby_ib0 \
   --hostname r2lead --fabric ib0
   ```

   Example 3, for a cluster with scalable unit (SU) leader nodes:

   ```
   # sgifmcli --standbysm --init --id standby_ib0 \
   --hostname leader2 --fabric ib0
   ```

3. Start the subnet manager master.

   For example:

   ```
   # sgifmcli --start --id master_ib0
   ```

   This command automatically starts the subnet manager master and the subnet manager standby for ib0.

4. Check the status of the subnet manager.

Example 1. The following example checks the status of `ib0` on a cluster without leader nodes:

```
# sgifmcli --status --id master_ib0

Master SM
Host = service0
Guid = 0x0008f10403987da9
Fabric = ib0
Toplogy = hypercube
Routing Engine = dor
OpenSM = running
Standby SM
Host = service1
Guid = 0x0008f10403987d25
Fabric = ib0
OpenSM = running
```

Example 2. The following example checks the status of `ib0` on a cluster with ICE leader nodes:

```
# sgifmcli --status --id master_ib0

Master SM
Host = r1lead
Guid = 0x0008f10403987da9
Fabric = ib0
Toplogy = hypercube
Routing Engine = dor
OpenSM = running
Standby SM
Host = r2lead
Guid = 0x0008f10403987d25
Fabric = ib0
OpenSM = running
```

Example 3. The following example checks the status of `ib0` on a cluster with scalable unit (SU) leader nodes:

```
# sgifmcli --status --id master_ib0

Master SM
Host = leader1
Guid = 0x0008f10403987da9
Fabric = ib0
Toplogy = hypercube
Routing Engine = dor
OpenSM = running
Standby SM
Host = leader2
Guid = 0x0008f10403987d25
Fabric = ib0
OpenSM = running
```

(Conditional) Configuring the InfiniBand fat-tree network topology

Complete the procedure in this topic if your cluster has an InfiniBand fat-tree network topology. After the cluster is provisioned, if you add an external switch to the cluster with fat-tree topology, perform this procedure to configure the external InfiniBand switch.

The node discovery commands configure external InfiniBand switches. After you run a node discovery command, you can use the `sgifmcli` command to add and initialize an external switch on the InfiniBand system.

The fat-tree topology involves external InfiniBand switches. For the list of supported external switches, see the `sgifmcli` manpage.

InfiniBand switches are of two types: leaf switches and spine switches. Leaf switches can connect to any kind of cluster node. Spine switches connect leaf switches together. The integrated InfiniBand switches in cluster systems are considered to be leaf switches. The external InfiniBand switches used to connect the leaf switches together in a fat-tree topology are considered to be spine switches.

The `sgifmcli` command lets you specify the following keywords for fat-tree topologies: `ftree` and `balanced-ftree`. The `balanced-ftree` keyword configures balanced fat-tree. If the fat-tree topology is not balanced, choose `ftree`. If the fat-tree topology is balanced, choose `bftree`.

The `discover --switch` command is equivalent to `sgifmcli --init` and `sgifmcli --add` when adding an external switch. If the external switch is configured not as an external switch, but as a general node, run the `sgifmcli --init` and `sgifmcli --add` commands.

Procedure

1. Verify the following:

   - The switch has been configured into the cluster database. The switch has an IP address on the management network.
   - The switch is properly connected to the InfiniBand network.
   - The admin port of the switch is properly connected to the Ethernet network.

2. Power on the switch.

   For more information, see your switch documentation.

3. From the admin node, use the `sgifmcli` command to initialize the switch.

   The syntax is as follows:

   ```
   sgifmcli --init --ibswitch --model modelname  --id mgmtsw_id --switchtype [leaf | spine]
   ```

   For *mgmtsw_id*, specify the switch hostname.

   For example:

   ```
   # sgifmcli --init --ibswitch --model voltaire-isr-2004  --id isr2004 \
   --switchtype spine
   ```

   The preceding example command configures a Voltaire ISR2004 switch, with the hostname of `isr2004`, as a spine switch. `isr2004` refers to the admin port of the switch. The switch is now initialized and the root globally unique identifier (GUID) from the spine switches has been downloaded.

4. From the admin node, use the `sgifmcli` command to add the switch to the fabric.

   The syntax is as follows:

   ```
   sgifmcli --add --id fabric --switch mgmtsw_id
   ```

   For example, the following command connects `isr2004` is connected to the `ib0` fabric:

   ```
   # sgifmcli --add --id ib0 --switch isr2004
   ```

5. (Conditional) Stop and restart the subnet manager master.

   Complete this step if the subnet master manager was running when you added the switch.

For example:

```
# sgifmcli --stop --id master_ib0
# sgifmcli --start --id master_ib0
```

6. Restart the subnet manager master and the optional subnet manager standby.

   For example:

   ```
   # sgifmcli --start --id master_ib0
   ```

   If you define a standby, the standby assumes control over the switch if the subnet manager master fails.

Omni-Path Express fabric management

The following topics explain how to manage the Omni-Path Express fabric management software:

- **Starting and stopping the Omni-Path Express fabric managers**
- **Managing Omni-Path Express fabric software**

Starting and stopping the Omni-Path Express fabric managers

- The following command starts the Omni-Path Express fabric managers:

  ```
  # systemctl start opafm
  ```

- Each fabric manager instance is a separate process. There is no interdependency among fabric manager instances. To start or stop an individual instance, enter one of the following commands:

  - `/usr/lib/opa-fm/bin/opafmctrl start -i fabric_manager_instance`
  - `/usr/lib/opa-fm/bin/opafmctrl stop -i fabric_manager_instance`

    For *fabric_manager_instance*, specify `0` for the first plane or `1` for the second plane.

- The following command stops all the running Omni-Path Express fabric managers:

  ```
  # systemctl stop opafm
  ```

Managing Omni-Path Express fabric software

The following examples show the commands that you can use to manage Omni-Path Express fabric software. These commands assume that the Omni-Path Express fabric software is installed and running as expected. If these commands fail, install the Omni-Path Express software on the servers.

Example 1. The following command displays the Omni-Path Express software version:

```
[root@n0 ~]# opaconfig -V
10.6.0.0.134
```

Example 2. The `ibstatus` command checks the state of a node and tests to see if an HFI is installed on the admin node. You can run this command from any node.

The following command shows that the link is up and that there is a single host fabric interface (HFI) adapter on the node named `n0`:

```
[root@n0 ~]# ibstatus
  Infiniband device 'hfi1_0' port 1 status:
          default gid:   fe80:0000:0000:0000:0011:7501:0167:1ecd
```

```
          base lid:      0xa
          sm lid:        0x1
          state:         4: ACTIVE
          phys state:    5: LinkUp
          rate:          100 Gb/sec (4X EDR)
          link_layer:    InfiniBand
```

The following command shows two adapters installed on the same leader node:

```
r1lead:~ # ibstatus
Infiniband device 'hfi1_0' port 1 status:
    default gid:     fe80:0000:0000:0000:0011:7501:0163:ce82
    base lid:     0xa
    sm lid:         0xa
    state:         4: ACTIVE
    phys state:     5: LinkUp
    rate:         100 Gb/sec (4X EDR)
    link_layer:     InfiniBand

Infiniband device 'hfi1_1' port 1 status:
    default gid:     fe80:0000:0000:0001:0011:7501:0163:cea3
    base lid:     0x11a
    sm lid:         0x11a
    state:         4: ACTIVE
    phys state:     5: LinkUp
    rate:         100 Gb/sec (4X EDR)
    link_layer:     InfiniBand
```

Example 3. You can check the state of the fabric manager running on the leader or admin node. If there are two separate fabrics, then `SM 0` and `SM 1` display as running. The command is as follows:

```
r1lead:~ # /usr/lib/opa-fm/bin/opafmctrl status
Checking IFS Fabric Manager
Checking SM 0: fm0_sm: Running
Checking FE 0: fm0_fe: Disabled
Checking SM 1: fm1_sm: Running
Checking FE 1: fm1_fe: Disabled
Checking SM 2: fm2_sm: Disabled
Checking FE 2: fm2_fe: Disabled
Checking SM 3: fm3_sm: Disabled
Checking FE 3: fm3_fe: Disabled
```

Example 4. You can use `systemd` commands, such as the following, to check the state of the fabric:

```
# systemctl status opafm
 opafm.service - OPA Fabric Manager
   Loaded: loaded (/usr/lib/systemd/system/opafm.service; disabled; vendor preset: disabled)
   Active: active (running) since Thu 20XX-04-26 14:52:29 CDT; 7min ago
  Process: 5186 ExecStart=/usr/lib/opa-fm/bin/opafmd -D (code=exited, status=0/SUCCESS)
 Main PID: 5190 (opafmd)
    Tasks: 17 (limit: 512)
   CGroup: /system.slice/opafm.service
           ├─5190 /usr/lib/opa-fm/bin/opafmd -D
           └─5191 /usr/lib/opa-fm/runtime/sm -e sm_0
```

As the following command shows, for two fabrics, the `systemctl` command hints at two subnet managers running:

```
# systemctl status opafm
 opafm.service - OPA Fabric Manager
   Loaded: loaded (/usr/lib/systemd/system/opafm.service; enabled; vendor preset: disabled)
   Active: active (running) since Fri 20XX-04-20 11:29:17 CDT; 6 days ago
 Main PID: 1747 (opafmd)
   CGroup: /system.slice/opafm.service
```

```
          ├─1747 /usr/lib/opa-fm/bin/opafmd -D
          ├─1788 /usr/lib/opa-fm/runtime/sm -e sm_0
          └─1789 /usr/lib/opa-fm/runtime/sm -e sm_1
```

Example 5. The following command checks the count of the switches on the fabric:

```
r1lead:~ # opareport -o lids -q -Q -F nodetype:SW
LID Summary
283 LID(s) in Fabric:
   LID(Range) NodeGUID          Port Type Name
0x0002        0x00117501026776dd   0 SW SGI Switch Node
0x0003        0x00117501026775f3   0 SW SGI Switch Node
0x0004        0x0011750102677602   0 SW SGI Switch Node
0x0005        0x0011750102677702   0 SW SGI Switch Node
0x0006        0x00117501026776d9   0 SW SGI Switch Node
0x0007        0x001175010267770a   0 SW SGI Switch Node
0x0008        0x00117501026776fc   0 SW SGI Switch Node
0x0009        0x00117501026776f1   0 SW SGI Switch Node
0x000b        0x00117501026776e0   0 SW SGI Switch Node
0x000c        0x00117501026776f6   0 SW SGI Switch Node
0x000d        0x00117501026776d4   0 SW SGI Switch Node
0x000e        0x0011750102677723   0 SW SGI Switch Node
0x000f        0x00117501026776eb   0 SW SGI Switch Node
0x0010        0x00117501026776f4   0 SW SGI Switch Node
0x0011        0x00117501026776f0   0 SW SGI Switch Node
0x0012        0x00117501026776f3   0 SW SGI Switch Node
16 Reported LID(s)
```

The following command checks the count of HFIs on the fabric:

```
r1lead:~ # opareport -o lids -q -Q -F nodetype:FI
LID Summary
283 LID(s) in Fabric:
   LID(Range) NodeGUID          Port Type Name
0x0001        0x001175010167006e   1 FI n1 hfi1_0
0x000a        0x001175010163ce82   1 FI r1lead hfi1_0
0x0013        0x001175010179e6d5   1 FI r1i3n0 hfi1_1
0x0014        0x001175010179010d   1 FI r1i3n1 hfi1_1
0x0015        0x001175010179e244   1 FI r1i3n2 hfi1_1
0x0016        0x001175010179e246   1 FI r1i3n3 hfi1_1
0x0017        0x001175010179e02f   1 FI r1i3n4 hfi1_1
0x0018        0x001175010179e24c   1 FI r1i3n5 hfi1_1
...
0x0119        0x001175810179e01b   1 FI r1i4n32 hfi1_0
0x011a        0x001175810179e242   1 FI r1i4n33 hfi1_0
0x011b        0x001175810179e27c   1 FI r1i4n34 hfi1_0
0x011c        0x001175810179d92c   1 FI r1i4n35 hfi1_0
267 Reported LID(s)
```

The values reported from the HFI and switch report must equal the number of devices on the fabric. In this case, 16+267=283.

InfiniBand fabric management

The following topics explain how to manage the InfiniBand fabric management software:

- **InfiniBand fabric operations on clusters with leader nodes**
- **Using the InfiniBand management tool GUI**

- **InfiniBand fabric management commands**
- **Automatic InfiniBand fabric management**

InfiniBand fabric operations on clusters with leader nodes

The cluster manager supports the OFED OpenSM software package and the `sgifmcli` tool for InfiniBand fabric management.

The InfiniBand fabric connects the leader nodes, the ICE compute nodes, and the compute nodes. It does not connect to the admin node or the chassis controller blades. Clusters with leader nodes usually have two separate InfiniBand fabrics, which are referred to as `ib0` and `ib1`.

Each InfiniBand fabric (also sometimes called an InfiniBand subnet) has its own subnet manager, which runs on a leader node. For a system with two or more racks, the subnet manager for each fabric is usually configured to run on different leader nodes. In a single rack system, both subnet managers run on the single leader node. Each subnet manager might also be paired with a standby subnet manager. If the primary subnet manager fails, the standby subnet manager takes over.

Leader nodes do not always have InfiniBand fabric host channel adapters (HCAs). In some cases, one or two leader nodes have HCAs to run the OFED subnet manager. In other cases, subnet management is done on separate fabric management nodes, so no leader nodes have InfiniBand HCAs.

Leader nodes associate a subnet manager instance with a particular port on the leader node. Usually, the following mapping exists:

- `ib0` is mapped to port 1 of the InfiniBand host channel adapter (HCA) on the subnet manager node.
- `ib1` is mapped to port 2 of the HCA on the subnet manager node.

To configure the subnet manager for `ib0` and `ib1`, use one of the following files:

- `/etc/ofa/opensm-ib0.conf`
- `/etc/ofa/opensm-ib1.conf`

NOTE: After a system reboots, the `opensm` daemons start running automatically.

For information about how to configure the InfiniBand fabric using the GUI, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

For information about `sgifmcli`, see the following:

InfiniBand fabric management commands

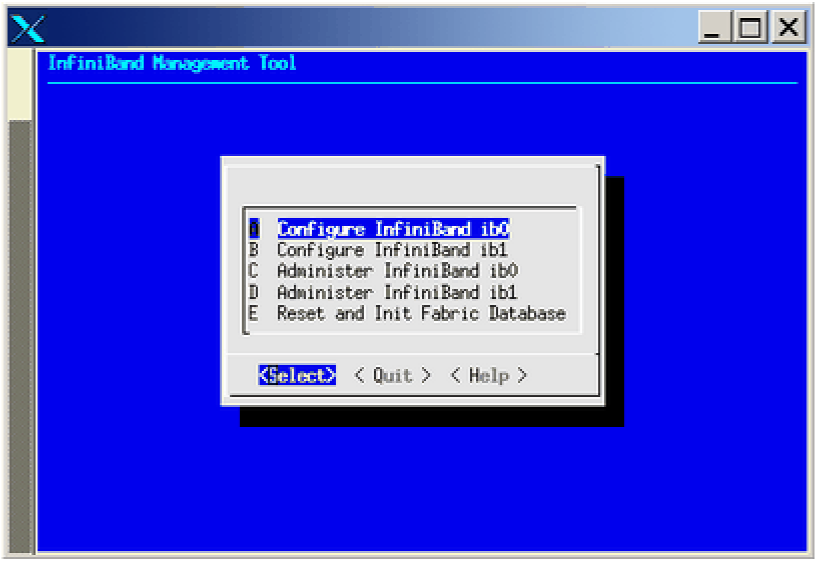
Using the InfiniBand management tool GUI

You can use the InfiniBand management GUI tool to configure, administer, or verify the InfiniBand fabric on the cluster.

To start the tool, log into the admin node and enter the following command:

```
admin:~ # tempo-configure-fabric
```

The following figure shows the **InfiniBand Management Tool** GUI:



InfiniBand Management Tool

```
A  Configure InfiniBand ib0
B  Configure InfiniBand ib1
C  Administer InfiniBand ib0
D  Administer InfiniBand ib1
E  Reset and Init Fabric Database
```

Figure 14: InfiniBand Management Tool screen

To highlight and select the action you want, use the following:

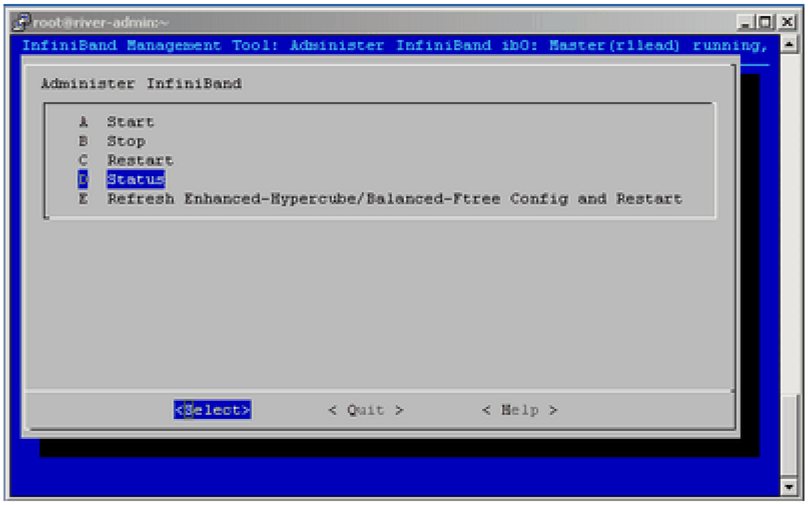
- The mouse
- The keyboard arrow keys
- The **Enter** key
- The **Tab** key

After you highlight menu choice, the following actions are possible:

- **Select** selects an action and displays to a submenu.
- **Quit** returns to the previous screen.
- **Help** displays online help for each of the GUI actions.

If the `tempo-configure-fabric` command fails in a configuration or administrative operation, use the `sgifmcli` command to debug the problem.

After configuring and bringing up the InfiniBand network, select the **Administer InfiniBand ib0** option or the **Administer InfiniBand ib1** option. You can use this screen to start, stop, restart, or refresh a fabric. You can verify the status through the **Status** option. The following is an example of the GUI after you select **Administer InfiniBand ib0** or **Administer InfiniBand ib1**:



```
InfiniBand Management Tool: Administer InfiniBand ib0: Master(r1lead) running,
```

Figure 15: Administer InfiniBand Status option

The **Status** option returns information similar to the following:

```
Master SM
Host = r1lead
Guid = 0x0002c9030006938b
Fabric = ib0
Topology = hypercube
Routing Engine = dor
OpenSM = running
```

To return to the `configure-cluster` GUI, press the **Enter** key.

The **Refresh Enhanced Hypercube Config and Restart** option applies only to the Enhanced Hypercube topology. You are required to refresh the fabric configuration when you either add, remove, or move one or more ICE compute nodes. The refresh action updates the `guid` routing order file that balances InfiniBand traffic for the Enhanced Hypercube topology. In addition, this action also automatically restarts the master subnet manager and the optional standby subnet manager for the specified fabric. Ideally, perform a refresh action for a fabric when there are no jobs running in the system. Restarting the subnet manager can have an adverse impact on the running jobs in the system.

For information about `sgifmcli`, see the following:

InfiniBand fabric management commands

InfiniBand fabric management commands

Hewlett Packard Enterprise recommends that you use the InfiniBand Management tool GUI for most fabric management operations. To start the GUI, enter the following command at the admin node system prompt:

```
# tempo-configure-fabric
```

For more information about the GUI, see the following:

Using the InfiniBand management tool GUI

The following topics explain how to use the fabric management commands:

- **`sgifmcli` InfiniBand fabric management command**
- **`sgifmdb` InfiniBand fabric management database command**

`sgifmcli` InfiniBand fabric management command

For advanced fabric management, use the `sgifmcli` command. For example, use `sgifmcli` for the following actions:

- Initializing and configuring external InfiniBand switches

  To configure an external InfiniBand switch, clusterwide InfiniBand connectivity is not required. The only requirements are as follows:

  - The supplied switch host name is resolvable
  - A working networking connection to the external InfiniBand switch exists

- Verifying the integrity of the InfiniBand fabric. This activity requires that the fabric is configured properly.

See the `sgifmcli`(8) manpage for the following information:

- Command syntax and examples.
- A list of switches that Hewlett Packard Enterprise supports on clusters.
- Information about adding external InfiniBand switches to your cluster fabric.
- Information about fabric verification operation.

The following are additional command examples.

Example 1. The syntax to start a subnet manager master is as follows:

```
sgifmcli --start --id identifier
```

For example, to start the `master_ib0` subnet manager master, enter the following:

```
# sgifmcli --start --id master_ib0
```

At this point, a master for the fabric `ib0` is running on the `r1lead`. The fabric `ib0` is available for compute jobs. If a standby is defined, the command launches both the standby and the master.

Example 2. The syntax to stop a subnet manager master is as follows:

```
sgifmcli --stop --id identifier
```

The following command stops the `master_ib0` subnet manager master running on host `r1lead`:

```
# sgifmcli --stop --id master_ib0
```

If a standby is defined, the standby also stops.

Example 3. The command syntax that checks the status of a subnet manager master is as follows:

```
sgifmcli --status --id identifier
```

The following command displays the status of the `master_ib0` subnet manager master:

```
# sgifmcli --status --id master_ib0
Master SM
Host = rlead
Guid = 0x0002c902002838f5
Fabric = ib0
Topology = hypercube
Routing Engine = dor
OpenSM = running
```

The command reports the status of the master subnet manager master `master_ib0` running on host `r1lead`. If a standby is defined, the command reports the status of both the standby and the master.

Example 4. The syntax to remove a subnet manager master is as follows:

```
sgifmcli --remove --id identifier
```

To remove the master_ib0 subnet manager master, first stop it and then perform the **-remove** option, as follows:

```
# sgifmcli --stop --id master_ib0

# sgifmcli --remove --id master_ib0
```

The subnet manager master is removed from the entity list. If a standby is defined, the command removes both the standby and the master.

Example 5. To remove the standby_ib0 subnet manager standby, first stop its master. Then, use the --remove option, as follows:

```
# sgifmcli --stop --id master_ib0
# sgifmcli --remove --id standby_ib0
```

The subnet manager standby is removed from the entity list. If a standby has been defined, the command removes both the standby and the master.

Example 6. To find the ID of the master subnet manager in the database, enter the following:

```
# sgifmcli --dump --id ib0 | grep MASTER
```

Example 7. To print the fabric configuration, enter the following:

```
# sgifmcli --showconfig

--------------
NAME = ib1
TYPE = ibfabric
MASTER =
STANDBY =
SWITCH_LIST =
--------------
NAME = ib0
TYPE = ibfabric
MASTER =
STANDBY =
SWITCH_LIST =
```

Example 8. To list the switches related to a particular fabric, enter the following command:

```
# sgifmcli --switchlist --id fabric
```

sgifmdb InfiniBand fabric management database command

The fabric component maintains a database of managed objects. The database version is automatically set during cluster install. You do not need to set it. Most likely, this database will change over time. To manage multiple database versions, use the sgifmdb command, which reports the managed objects database version.

For information about the sgifmdb command, enter the following from the admin node:

```
admin:~ # sgifmdb -h
SGI Fabric Component DB tool
Usage: db_version [--get|-g] [--dump|-d] [-v|--version] [-r|--reset] [--help|-h]

        -g, --get       Read DB version object from DB
        -d, --dump      Dump the DB
        -v, --version   Print version
        -r, --reset     Reset the database and start clean
        -h, --help      Show this text
```

Automatic InfiniBand fabric management

By default, each subnet manager performs a light sweep of the fabric it is managing every 10 seconds. The time is set on the admin node in the sweep_interval variable in the following file:

```
/opt/sgi/var/sgifmcli/opensm-ibx.conf.templ
```

If a subnet manager detects a change in the fabric during a light sweep, it performs a **heavy** sweep. Examples of changes are updates such as the addition or deletion of a node. The heavy sweep changes the fabric configuration to reflect the current state of the system.

There is one opensm instance for each fabric. Each instance associates itself with a particular globally unique identifier (GUID) for a port on the node upon which opensm runs. This association is configured with the guid entry in the corresponding opensm-ib*x*.conf file.

NOTE: If your cluster has more than 256 nodes, increase the sweep_interval variable to 90 seconds or more.

To reset the sweep interval from the GUI, do the following:

- Edit the sweep_interval variable in the /opt/sgi/var/sgifmcli/opensm-ib0.conf.templ file.
- To launch the GUI, run the tempo-configure-fabric command.
- Do a **Commit** operation in the GUI.

Alternatively, use the sgifmcli command --arglist parameter. This parameter sets various subnet manager configuration parameters including the sweep interval.

For example:

```
# sgifmcli --set --id master_ib0 --arglist sweep_interval=90
```

For information about fabric sweeping, see the opensm(8) manpage on the leader node.

InfiniBand network topology

Cluster systems with a hypercube topology use the dimension order routing (DOR) algorithm. The DOR algorithm is based on the min-hop algorithm. The algorithm uses the shortest paths. The goal of using the shortest path is in contrast to other goals. A different goal might spread out traffic across different paths with the same shortest distance. When choosing the shortest path, it chooses among the available shortest paths based on an ordering of dimensions.

Cluster systems with a fat-tree topology use the Unicast routing algorithm (UPDN) as the default routing algorithm. UPDN is also based on the minimum hops to each node, but it is constrained to ranking rules.

There are two opensm daemons, one for each fabric, opensmd-ib0 and opensmd-ib1. The init.d scripts control the opensm daemons. Each init.d script has a separate configuration file for each fabric, opensm-ib0 and opensm-ib1.

The sminfo command shows the GUID of the subnet manager master.

For more information on routing variables, see the opensm(8) manpage.

Utilities and diagnostics for Omni-Path Express fabrics and InfiniBand fabrics

The diagnostics package on your cluster contains tools and diagnostic software for the Open Fabrics Enterprise Distribution (OFED) software. On clusters with leader nodes, these tools reside on the leader nodes in the /usr/sbin directory. In addition, the opensm(8) manpage describes options that control logging and debugging.

To run the commands in the following topics, log into a node that is attached to the fabrics. The following topics include information about fabric diagnostic software:

- **`ibstat` and `ibstatus` commands**
- **`perfquery` command for InfiniBand fabric**
- **`ibnetdiscover` command for InfiniBand fabric**
- **`ibdiagnet` command for InfiniBand fabric**
- **Logging and debugging options for Omni-Path Express subnets and InfiniBand subnets**

`ibstat` and `ibstatus` commands

The `ibstat` command displays the status of the host channel adapters (HCAs) in your Omni-Path Express fabric or InfiniBand fabric. The status includes the HCAs on the leader nodes.

Example 1. The following shows `ibstat` output.

```
r1lead:/usr/bin # ibstat
CA 'mlx4_0'
        CA type: MT4099
        Number of ports: 2
        Firmware version: 2.40.5030
        Hardware version: 0
        Node GUID: 0xe41d2d03006f51e0
        System image GUID: 0xe41d2d03006f51e3
        Port 1:
                State: Active
                Physical state: LinkUp
                Rate: 56
                Base lid: 1
                LMC: 0
                SM lid: 1
                Capability mask: 0x0251486a
                Port GUID: 0xe41d2d03006f51e1
                Link layer: InfiniBand
        Port 2:
                State: Active
                Physical state: LinkUp
                Rate: 56
                Base lid: 1
                LMC: 0
                SM lid: 1
                Capability mask: 0x0251486a
                Port GUID: 0xe41d2d03006f51e2
                Link layer: InfiniBand
```

Example 2. The following `ibstatus` command shows the link rate. The `ibstatus` command is more terse than the `ibstat` command.

```
r1lead:/usr/bin # ibstatus
Infiniband device 'mlx5_0' port 1 status:
        default gid:     fec0:0000:0000:0000:e41d:2d03:006f:51e1
        base lid:        0x1
        sm lid:          0x1
        state:           4: ACTIVE
        phys state:      5: LinkUp
        rate:               56 Gb/sec (4X FDR)
        link_layer:         InfiniBand

Infiniband device 'mlx5_0' port 2 status:
        default gid:        fec0:0000:0000:0001:e41d:2d03:006f:51e2
        base lid:           0x1
        sm lid:             0x1
        state:              4: ACTIVE
        phys state:         5: LinkUp
        rate:               56 Gb/sec (4X FDR)
        link_layer:         InfiniBand
```

`perfquery` command for InfiniBand fabric

The `perfquery` command finds errors on one or more host channel adapters (HCAs) and errors on switch ports. You can also use `perfquery` to reset HCA and switch port counters.

For example output, enter one or more of the following commands on a node that is attached to the InfiniBand fabric. Typically, a leader node or a compute node is attached to the InfiniBand fabric. Example commands are as follows:

- To display command options, enter the following:

  ```
  # perfquery --help
  ```

- To display a list of counters, enter the following:

  ```
  # perfquery
  ```

`ibnetdiscover` command for InfiniBand fabric

The `ibnetdiscover` command configures the InfiniBand fabric.

For example output, enter one or more of the following commands on a node that is attached to the InfiniBand fabric. Typically, a leader node or a compute node is attached to the InfiniBand fabric. The commands are as follows:

- To display command options, enter the following:

  ```
  # ibnetdiscover --help
  ```

- To display status information, enter the following:

  ```
  # ibnetdiscover
  ```

`opareport` command for Omni-Path Express fabrics

The `opareport` command provides Omni-Path Express fabric analysis and reports. Run this command on a node that is connected to the Omni-Path Express Fabric with the Omni-Path Express Fabric Suite FastFabric tool set installed. Typically, this node is a leader node or a compute node.

`ibdiagnet` command for InfiniBand fabric

The `ibdiagnet` command scans the fabric and extracts information about connectivity and devices.

For example output, type one or more of the following commands on a node that is attached to the InfiniBand fabric. Typically, a leader node or a compute node is attached to the InfiniBand fabric. The commands are as follows:

- To display command options, enter the following:

  ```
  # ibdiagnet --help
  ```

- To display a summary report, enter the following:

  ```
  # ibdiagnet
  ```

The following example shows how to use `ibdiagnet` to load the fabric for testing.

```
r1lead:/opt/sgi/sbin # ibdiagnet -c 5000
Loading IBDIAGNET from: /usr/lib64/ibdiagnet1.2
Loading IBDM from: /usr/lib64/ibdm1.2
-W- Topology file is not specified.
    Reports regarding cluster links will use direct routes.
-W- A few ports of local device are up.
    Since port-num was not specified (-p option), port 1 of device 1 will be
    used as the local port.
-I- Discovering the subnet ... 10 nodes (2 Switches & 8 CA-s) discovered.


-I---------------------------------------------------
-I- Bad Guids Info
-I---------------------------------------------------
-I- No bad Guids were found

-I---------------------------------------------------
-I- Links With Logical State = INIT
-I---------------------------------------------------
-I- No bad Links (with logical state = INIT) were found

-I---------------------------------------------------
-I- PM Counters Info
-I---------------------------------------------------
-I- No illegal PM counters values were found

-I---------------------------------------------------
-I- Bad Links Info
-I---------------------------------------------------
-I- No bad link were found

-I- Done. Run time was 8 seconds.
```

Logging and debugging options for Omni-Path Express subnets and InfiniBand subnets

The following information pertains to logging and debugging:

- The Omni-Path Express subnet manager initializes the fabric and facilitates fabric topology management. You can use the `opafmcmd` command for debugging Omni-Path Express problems.
- The InfiniBand subnet manager is called OpenSM. The `opensm`(8) manpage describes the ranges for the debugging and logging options. When you start a troubleshooting session, Hewlett Packard Enterprise recommends that you set the following parameters:
  - `-D 0x7`, which sets a reasonable log verbosity level.
  - `-d 2`, which clears the logs immediately after each log message.

For more information about the OpenSM utility, log into one of the leader nodes and see the `opensm`(8) manpage.

System maintenance and troubleshooting

Troubleshooting tools

The following topics describe some troubleshooting tools.

cm-info-gather command

The cm-info-gather command collects support information about the cluster, the admin node, and the leader nodes. To collect information about a node, use the -w option.

For more information, enter the following:

```
# cm-info-gather -h
```

cminfo command

Many cluster scripts use the cminfo command internally. In a troubleshooting situation, you can use the command to gather information about your system.

Example 1. To display the node controller IP address of a leader node on an HPE SGI 8600 cluster, enter the following command:

```
r1lead:~ # cminfo --bmc_base_ip
192.168.160.0
```

Example 2. To display the leader node DNS domain on an HPE SGI 8600 cluster, enter the following command:

```
r1lead:~ # cminfo --dns_domain
cm.clusterdomain.com
```

Example 3. To see the IP address of the ib1 InfiniBand fabric on an HPE SGI 8600 cluster, enter the following command:

```
r1lead:~ # cminfo --ib_1_base_ip
10.149.0.0
```

Obtaining a system dump from a node with an iLO device

On clusters with nodes that include iLO devices, such as HPE Apollo clusters, complete the procedure in this topic to obtain a system dump. An iLO device is a type of node controller.

Procedure

1. Log into the admin node as the root user.
2. Use the cadmin command to obtain the username and password for the node controller that resides inside the node for which you want to obtain a system dump.

   The format for this command is as follows:

   ```
   cadmin --get-bmc-password --node node[,node]
   ```

   For *node*, specify one or more node hostnames.

   For example:

   ```
   # cadmin --get-bmc-password --node fmn
   admin
   admin
   ```

3. Use the ssh command to connect to the iLO node controller in the node.

Specify the username and password displayed by the previous step.

4. Enter the following command to generate a system dump:

```
</>hpiLO-> NMI
```

`kdump` utility

The `kdump` utility is a `kexec`-based crash dumping mechanism for the Linux operating system. By default, the `kdump` utility is enabled on all nodes except the admin node. Also by default, on a cluster with leader nodes, the `kdump` crash dump capability is enabled after installation.

The cluster manager works with the `kdump` utility in the following way:

- If a `crashkernel` attribute exists, but is empty, the cluster manager does not generate a crash dump.
- If a `crashkernel` attribute exists and has a value, the cluster manager uses that value.
- If a `crashkernel` attribute does not exist, the cluster manager uses the default value.

Obtaining a traceback or system dump on an HPE SGI 8600 cluster

On the admin node, system dump information resides in the following locations:

- Traceback information is in the following file:

  ```
  /var/log/consoles/node_hostname
  ```

- System dump information is in the following directory:

  ```
  /var/crash/sgi_kdump/IP-date
  ```

  For example:

  ```
  root@r1lead ~]# cd /var/crash/sgi_kdump/10.159.0.6-20XX-04-21-14\:54\:20/
  [root@r1lead 10.159.0.6-20XX-04-21-14:54:20]# ls -l
  total 128488
  -rw-r--r-- 1 sgi_kdump sgi_kdump     69408 Apr 21 14:54 vmcore-dmesg.txt
  -rw-r--r-- 1 sgi_kdump sgi_kdump 131499356 Apr 21 14:54 vmcore.flat
  ```

On an ICE leader node, the system dump information is in the following directory:

```
/var/crash/sgi_kdump/IP-date
```

For an ICE compute node, leader node, or compute node, access system dump information as follows:

1. Log into the admin node.
2. Bring up a console to the node in question.

   For example:

   ```
   # cm node console -n n14
   ```

3. Enter the following to obtain the crash dump for the selected node:

   ```
   ^e c l 1 8
   ^e c l 1 t       #traceback
   ^e c l 1 c       #dump
   ```

For example, to obtain information from a leader node, enter the following:

```
# cm node console -n r1i0n0
^e c l 1 8
^e c l 1 t       #traceback
^e c l 1 c       #dump
```

NOTE: This example shows the letter “c”, a lowercase L “l”, and the number one “1” in all three lines.

kdump examples

The kdump directory listings include the node hostname and the node IP address.

Example 1. On RHEL platforms, the kdump crash directories are located in /var/crash/sgi_kdump. The following example shows a listing of crash files from diskless nodes on a cluster without leader nodes:

```
admin_node: # cd /var/crash/sgi_kdump
admin_node:/var/crash/sgi_kdump # ls -la
lrwxrwxrwx  1 sgi_kdump sgi_kdump   30 Mar  1 17:24 n0-20XX-03-01-17:19:24 -> 172.23.0.4-20XX-03-01-17:19:24
lrwxrwxrwx  1 sgi_kdump sgi_kdump   30 Mar  1 17:31 n0-20XX-03-01-17:26:35 -> 172.23.0.4-20XX-03-01-17:26:35
lrwxrwxrwx  1 sgi_kdump sgi_kdump   30 Mar  1 19:07 n0-20XX-03-01-19:07:21 -> 172.23.0.4-20XX-03-01-19:07:21
lrwxrwxrwx  1 sgi_kdump sgi_kdump   30 Mar  1 19:49 n0-20XX-03-01-19:48:35 -> 172.23.0.4-20XX-03-01-19:48:35
```

When you change to the n0-20XX-03-01-17:19:24 directory and list the files again, the system displays the following files:

- vmcore-dmesg.txt
- vmcore.flat

Example 2. The following example is from a SLES cluster. The example shows the ICE compute node crash files. These files are on the leader node.

```
r1lead:/var/crash/sgi_kdump # ls -la
drwxr-xr-x 3 sgi_kdump users   30 Mar 28 15:00 10.159.0.2
lrwxrwxrwx 1 sgi_kdump users   10 Mar 28 15:00 r1i0n0 -> 10.159.0.2
r1lead:/var/crash/sgi_kdump # cd r1i0n0
r1lead:/var/crash/sgi_kdump/r1i0n0 # ls -F
20XX-03-28-22:00/
r1lead:/var/crash/sgi_kdump/r1i0n0 # cd 20XX-03-28-22:00
r1lead:/var/crash/sgi_kdump/r1i0n0/20XX-03-28-22:00 # ls
dmesg.txt  makedumpfile-R.pl  README.txt  rearrange.sh  vmcore          # CRASH FILES ARE HERE
```

Retrieving the current kdump memory allocation setting

The following procedure explains how to retrieve the current kdump setting for a specific system image.

Procedure

1. Log into the admin node as the root user.
2. Use the cm image show command to retrieve the current kdump memory allocation.:

   For example:

   ```
   # cm image show -s -i rhel8.X
   custom-partitions    = Undefined
   crashkernel          = crashkernel=320M
   hard-quota           = Undefined
   kernel-extra-params  =
   kernel-distro-params = ro root=dhcp intel_idle.max_cstate=1
   ```

```
processor.max_cstate=1 selinux=0 net.ifnames=0 biosdevname=0
numa_balancing=disable
kernel-leader-params =
nfsroot-extra-params = Undefined
quota-timer          = Undefined
repo-group           = Undefined
soft-quota           = Undefined
```

Disabling `kdump`

When `kdump` is enabled, the system reserves some memory for crash dumps. To make this memory available to user programs, you can disable the `kdump` facility. You can also reduce the size of the memory used for the `kdump` facility.

The following procedure explains how to disable `kdump`.

Procedure

1. Log into the admin node as the root user.
2. Use the following command to disable the `kdump` facility:

   ```
   cm image set -i image --crashkernel 'crashkernel='
   ```

   For *image*, specify the name of one of the leader node, ICE compute node, or compute node operating system images.

   For example:

   ```
   # cm image set -i ice-sles15spX --crashkernel 'crashkernel='
   ```
3. Push the changes to the desired nodes.

   For information about how to push changes, see the following:

   Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Disabling `kdump initrd` generation on NFS compute nodes

You can disable the `kdump initrd` generation at the time of image activation. The cluster manager supports this capability for nodes with NFS file systems on clusters with scalable unit (SU) leader nodes and on clusters without leader nodes.

When you disable `kdump initrd` generation, you also disable `kdump` capabilities on the node.

Typically, when you activate a read-only NFS file system image for a compute node, `kdump initrd` is generated at the time of activation. This process maintains the availability of the NFS file system while the nodes are booting. The `kdump initrd` generates a lot of small file I/O, which can degrade NFS file system performance. To suppress `kdump initrd` generation, use the `--disable-kdump` parameter as shown in this topic.

Procedure

1. Log into the admin node as the root user.
2. Enter the `cm image activate` command with the `--disable-kdump` parameter, as follows:

   ```
   cm image activate --disable-kdump image_name
   ```

Setting a site-specific `crashkernel` value

The following procedure explains how to specify the amount of memory you want to devote to `kdump`.

Procedure

1. Log into the admin node as the root user.
2. Use the following command to specify the amount of memory to use for the `kdump` facility:

   ```
   cm image set -i image --crashkernel 'crashkernel=mem_size'
   ```

   The variables are as follows:

Variable	Specification
image	The name of one of the leader node, ICE compute node, or compute node operating system images.
mem_size	The amount of memory to allocate to `kdump`.

   For example:

   ```
   # cm image set -i sles15spX --crashkernel 'crashkernel=512M'
   ```

   For more information, see the `kdump`(7) manpage.
3. Push the changes to the desired nodes.

   For information about how to push changes, see the following:

   Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Resetting the `crashkernel` value to the system default

The procedure in this topic explains how to reset the `kdump` value to the system default value.

Complete this procedure to revert to the default settings after either of the following:

- After disabling `kdump`
- After resetting the amount of memory for `crashkernel`

Procedure

1. Log into the admin node as the root user.
2. (Optional) Display a list of images and display the default `crashkernel` value.

   Use the following commands:

   ```
   cm image show
   cm image show -s -i image
   ```

   For *image*, specify one of the images that the `cm image show` command displayed.

   For example:

   ```
   # cm image show
   ice-sles15spX
   lead-sles15spX
   sles15spX
   rhel8.X
   ```

```
# cm image show -s -i rhel8.X
custom-partitions    = Undefined
crashkernel          = crashkernel=320M
hard-quota           = Undefined
kernel-extra-params  =
kernel-distro-params = ro root=dhcp intel_idle.max_cstate=1
processor.max_cstate=1 selinux=0 net.ifnames=0 biosdevname=0
numa_balancing=disable
kernel-leader-params =
nfsroot-extra-params = Undefined
quota-timer          = Undefined
repo-group           = Undefined
soft-quota           = Undefined
```

3. Use the following command to specify the amount of memory to use for the `kdump` facility:

```
cm image set -i image --crashkernel 'crashkernel=value'
```

The variables are as follows:

Variable	Specification
image	The name of an operating system image for one of the leader nodes, ICE compute nodes, or compute nodes.
value	Specify a size.

For example:

```
# cm image set -i rhel8.X --crashkernel 'crashkernel=256M'
```

4. Push the changes to the nodes.

For information about how to push changes, see the following:

Provisioning compute nodes on HPE Cray EX clusters and HPE Apollo clusters

Hardware maintenance procedures

Displaying network interface card (NIC) information for a node

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to display NIC characteristics:

```
cm node nic show -n node
```

For *node*, specify one node hostname.

The preceding command shows only the required parameter. The command accepts additional parameters.

For example:

```
# cm node nic show -n x9000c1s0b0n1
ID  NAME     IP          MAC                IPV6  BOND_MASTER  BOND_MODE      INTERFACE_NAME      MANAGED  TYPE  NETWORK_NAME
75  enp65s0  10.168.0.2  00:40:a6:83:05:2e  None  bond0        active-backup  x9000c1s0b0n1       True     mgmt  hostmgmt2001
```

```
110 hsn1     10.150.0.7  None              None  None         None          x9000c1s0b0n1-hsn1  True     data  hsn
112 hsn0     10.150.0.9  None              None  None         None          x9000c1s0b0n1-hsn0  True     data  hsn
```

Adding network interface card (NIC) information to a node

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to add NIC information:

   ```
   cm node nic add -n nic_name -w net_name -n node
   ```

   This command adds the specified NIC to the node on the specified network.

   The preceding command shows only the required parameters. The command accepts additional parameters. The variables are as follows:

Variable	Specification
nic_name	The NIC name.
net_name	The network hostname. For `cm node nic add`, make sure that the network to which you wish to attach the node has already been created so that it can be specified using the required `−w` option
node	One node hostname.

3. Reboot the node.

Deleting network interface card (NIC) information for a node

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node nic show` command, in the following format, to display information about the NICs in the node:

   ```
   cm node nic show -n node
   ```

   For *node*, specify one node hostname.

   In the output, determine whether the NIC you want to delete is named in a unique way on this node.

3. Enter one of the following commands:

   Enter the following command if the *nic_name* is unique on the node:

   ```
   cm node nic delete -N nic_name -n node
   ```

   Or

   Enter the following command if the *nic_name* is not unique on the node:

   ```
   cm node nic delete -I ID -n node
   ```

These commands delete NIC information for the specified node.

The preceding commands show only the required parameters. The commands accept additional parameters. The variables are as follows:

Variable	Specification
nic_name	The NIC name.
node	The node hostname.
ID	The numeric identifier for the NIC.

4. Reboot the node.

Modifying network interface card (NIC) information for a node

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node nic show` command, in the following format, to display information about the NICs in the node:

   ```
   cm node nic show -n node
   ```

   For *node*, specify one node hostname.

   In the output, determine whether the NIC you want to modify is named in a unique way on this node.
3. Enter one of the following command to modify NIC information:

   Enter the following command if the *nic_name* is unique on the node:

   ```
   cm node nic set -N nic_name setting -n node
   ```

   Or

   Enter the following command if the *nic_name* is not unique on the node:

   ```
   cm node nic set -I ID setting -n node
   ```

   These commands set or clear NIC characteristics.

   The preceding commands show only the required parameters. The commands accept additional parameters. The variables are as follows:

Variable	Specification
nic_name	The NIC name.
node	The node hostname.

Table Continued

Variable	Specification
ID	The numeric identifier for the NIC.
setting	One or more characteristics to set or clear. For example, you can set or reset the NIC bonding mode, give the NIC a new IP address, or specify other NIC characteristics. Enter the following to display the list of possible characteristics: `# cm node nic set -h`

4. Reboot the node.

Taking one node offline for maintenance temporarily

The following procedure explains how to temporarily take a leader node, an ICE compute node, or a compute node offline for maintenance.

Procedure

1. Disable the node in the batch scheduler.

   See your batch scheduler documentation for this procedure.

2. Power off the node.

   For example:

   ```
   # cm power off -n r1i0n0
   ```

3. Mark the node offline.

   For example:

   ```
   # cm node set --administrative-state offline -n r1i0n0
   ```

4. Perform maintenance on the blade.

5. Mark the node online, as follows:

   For example:

   ```
   # cm node set --administrative-state online -n r1i0n0
   ```

6. Power up the node.

   For example:

   ```
   # cm power on -n r1i0n0
   ```

7. Enable the node in the batch scheduler.

   See your batch scheduler documentation for this procedure.

Taking one ICE leader node in a highly available (HA) leader node configuration offline for maintenance temporarily (HPE SGI 8600)

This topic explains how to shut down one of the two ICE leader nodes. The procedure shuts down the node in an orderly way. The procedure also avoids the unexpected results that can occur when the cluster manager perceives one of the nodes being down as a failure. This procedure works for all nodes, regardless of their status in the cluster as a master node or a slave node. This procedure is best run on two windows.

Procedure

1. As the root user, log into the leader node that you do not want to take down.

   In other words, log into the leader node that you want to remain operational during the maintenance procedure.

2. In one window, enter the following command so you can monitor the leader node transition to the standby node:

   ```
   # crm_mon
   ```

3. In a second window, enter the following command to set the other node (the target node) to standby status:

   ```
   # crm_standby --update=on --node
   ```

4. In the first window, monitor the progress of the target node as it transitions to `standby` mode.

   Make sure that the node is not listed as `Online`.

5. (Optional) Enter the following command to safely power off the target node at this time:

   ```
   # cm power shutdown -t leader r1lead
   ```

6. Perform the maintenance activity on the target node.

7. Enter the following command to clear the `standby` status on the node:

   ```
   # crm_standby --update=off --node
   ```

   Monitor the first window to make sure that the status changes from `standby` to `OFFLINE`.

8. (Conditional) Power on the target node.

   Complete this step if the node is not powered on.

   Enter the following command:

   ```
   # cm power on -t leader r1lead
   ```

Replacing a failed blade on an HPE SGI 8600 cluster

On some platforms, beware that some blades include multiple nodes.

NOTE: See your technical support representative for the physical removal and replacement of ICE compute nodes (blades).

The following procedure explains how to permanently replace a failed blade.

Procedure

1. (Optional) Disable the node in the batch scheduler.

   See your batch scheduler documentation for this procedure.

2. Power off the node.

   For example:

   ```
   # cm power off -t node r1i0n0
   ```

3. Mark the node offline.

For example:

```
# cm node set --administrative-state offline -n r1i0n0
```

4. Physically remove and replace the failed blade.

   It is not necessary to run `discover-rack` when you replace a blade. The `blademond` daemon performs that task.

5. Power on the node.

   For example:

   ```
   # cm power on -t node r1i0n0
   ```

6. Set the node to boot the required compute image.

   For example:

   ```
   # cimage --set mycomputeimage mykernel r1i0n0
   ```

   For information about this step, see the following:

   - Run the `cimage --show-images` command, and observe the output.
   - **Managing ICE compute node images**

7. (Optional) Enable the node in the batch scheduler.

   See your batch scheduler documentation for this procedure.

Replacing a management switch

The procedure in this topic explains how to back up a management switch and replace the switch with a new switch. Use the information in this topic when it is possible to back up the existing switch or if you have already backed up the existing switch.

To replace a management switch for which you have none of the original configuration information, treat the switch as if you were configuring a new switch into the cluster. In this case, use the information in the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

The following topics explain how to replace a management switch:

- **Backing up the current management switch configuration file**
- **Configuring the new management switch**
- **Management switch replacement example**

Backing up the current management switch configuration file

The following procedure explains how to back up a management switch configuration file.

Procedure

1. Save the running configuration file as the startup configuration file.

   Use the `switchconfig` command in the following format:

   ```
   switchconfig config --switches hostname --save
   ```

   For *hostname*, specify the switch hostname.

For example:

```
# switchconfig config --switches mgmtsw0 --save
```

2. Save the startup configuration file locally.

   Use the `switchconfig` command in the following format:

   ```
   switchconfig config --switches hostname --pull
   ```

   For *hostname*, specify the switch hostname.

   The preceding command saves the startup configuration file to the following default location on the admin node:

   ```
   /opt/clmgr/tftpboot/mgmtsw_config_files/hostname/config_file
   ```

   For example, the following command writes the switch configuration file to `/opt/clmgr/tftpboot/mgmtsw_config_files/mgmtsw0/primary.cfg`:

   ```
   # switchconfig config --switches mgmtsw0 --pull
   ```

3. Proceed to the following:

   Configuring the new management switch

Configuring the new management switch

The following procedure copies the saved configuration file from the admin node to the new switch.

Procedure

1. Use the documentation from the switch manufacturer to physically replace the old switch with the new switch.

   Make sure that the cabling is identical to the way the old switch cabling was configured.

2. Log into the admin node as the root user.

3. Use the `cm node set` command to update the cluster database with the MAC address of the new switch.

   Use the following format:

   ```
   cm node set -n switch --mac-address mac_address [--dev device]
   ```

   The variables are as follows:

Variable	Specification
switch	The hostname of the switch that is being replaced.
device	The hostname of the network interface card. For example, `eth1`. By default, the command operates on the primary interface.
mac_address	The MAC address, on the switch, that the switch can use to obtain an IP address.

   For example:

   ```
   # cm node set -n mgmtsw0 --mac-address 02:04:96:98:3c:91 --dev eth0
   ```

4. Use the `switchconfig` command to push the configuration file to the new switch.

   The format is as follows:

   ```
   switchconfig config --switches switch --push --local-file config_file
   ```

For *config_file*, specify the file that you want the new switch to load when you boot the switch. To find the name, look in the following directory:

```
/opt/clmgr/tftpboot/mgmtsw_config_files/switch/startup_config_file
```

For example:

```
# switchconfig config --switches mgmtsw0 --push --local-file primary.cfg
```

5. Use the `switchconfig` command to boot the new management switch.

   The format is as follows:

   ```
   # switchconfig restart --switches switch
   ```

   For example:

   ```
   # switchconfig restart --switches mgmtsw0
   ```

Management switch replacement example

The following example shows the process for replacing a management switch. Assume that the switches are as follows:

- The old switch has MAC address `02:04:96:99:88:77`.
- The replacement switch has MAC address `02:04:96:98:3c:91`.
- The switch hostname is `mgmtsw1`.
- The cabling for the two switches is identical.
- DCHP is enabled on the replacement switch.

Example:

```
# cm node show -M -n mgmtsw1                                             # Verify old switch's MAC address
NODE       NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS          MGMTSERVERIP    DEFAULTGATEWAY
mgmtsw1    None            172.23.255.254  255.255.0.0     40:b9:3c:a0:68:47   default         default
# switchconfig config --switches mgmtsw1 --save                          # Save and back up mgmtsw1's config file
# switchconfig config --switches mgmtsw1 --pull
# ls /opt/clmgr/repos/mgmtsw_config_files/mgmtsw1                        # Find name of old switch config file
primary.cfg
# cm node set --mac-address 02:04:96:98:3c:91 -n mgmtsw1 --dev eth0      # Add new switch's MAC address
# cm node show -M -n mgmtsw1                                             # Verify new switch's MAC address
NODE       NETWORK.NAME    IPADDRESS       SUBNETMASK      MACADDRESS          MGMTSERVERIP    DEFAULTGATEWAY
mgmtsw1    None            172.23.255.254  255.255.0.0     02:04:96:98:3c:91   default         default
# switchconfig config --switches mgmtsw1 --push --local-file primary.cfg # Push config file to new switch
# switchconfig restart --switches mgmtsw1                                # Boot the new switch
```

Querying field replaceable units (FRUs)

The `cm inventory` command displays FRUs, firmware information system nodes, and integrated InfiniBand switches.

The following example shows the default output format:

```
# cm inventory
leader3.fru.SKU=867959-B21
leader3.fru.Model=ProLiant DL360 Gen10
leader3.fru.SerialNumber=MXQ831048C
leader3.fru.Manufacturer=HPE
leader3.fw.FirmwareVersion=iLO 5 v2.14
leader1.fru.SKU=867959-B21
leader1.fru.Model=ProLiant DL360 Gen10
leader1.fru.SerialNumber=MXQ8310482
leader1.fru.Manufacturer=HPE
leader1.fw.FirmwareVersion=iLO 5 v2.33
r1c1.fru.SKU=000000-001
```

```
r1c1.fru.Model=HPE Chassis Management Controller
r1c1.fru.SerialNumber=PWWGRX3LMCI01K
r1c1.fru.Manufacturer=Hewlett Packard Enterprise
r1c1.fw.FirmwareVersion=Version 0.73 Oct 23 2020
leader2.fru.SKU=867959-B21
leader2.fru.Model=ProLiant DL360 Gen10
leader2.fru.SerialNumber=MXQ8310480
leader2.fru.Manufacturer=HPE
leader2.fw.FirmwareVersion=iLO 5 v2.33
```

The command can display different types of information. For example:

- To display output in `json` format, specify the `-j` parameter on the `cm inventory` command.
- Various parameters direct the command to include more information.

Enter one of the following commands to obtain more examples or information about filtering options:

- `man cm inventory`
- `cm inventory -h`

Replacing an HPE Apollo Moonshot cartridge

The following procedure explains how to replace an HPE Apollo Moonshot cartridge.

Procedure

1. Follow your hardware documentation instructions to replace the failed cartridge.
2. Use the following command to scan the chassis and update the database:

   ```
   cm_scan_moonshot -L ilocm_ip(s)
   ```

   The command also updates the cluster database with cartridge and node location information. This command is essential for proper power operations.

Adding a new compute node or device to the cluster when the MAC address is known

The procedure in this topic explains how to add one or more compute nodes to the cluster without editing the cluster definition file. You can use this procedure if you have to add nodes quickly or temporarily.

Procedure

1. Log into the admin node as the root user.
2. Physically attach the node to the cluster.
3. (Optional) Enter the following command to display help text for the `cm node add` command:

   ```
   # cm node add -h
   ```

4. Use the `cm node add` command in the following format to add the node to the cluster:

   ```
   cm node add --node-def 'description1 [description2]'
   ```

For *description*, specify a string of configuration attributes to describe the node. Model the string of configuration attributes after those that appear in cluster definition files for nodes. For information about configuration attributes, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

The following example adds a node with internal name `service1` and hostname `node42` to the cluster:

```
# cm node add --node-def 'internal_name=service1,hostname1=node42,'\
'mgmt_net_macs=00:9C:02:99:1B:3C,mgmt_net_interfaces="eno1",'\
'mgmt_bmc_net_macs=00:9c:02:99:1b:3e,mgmt_bmc_net_ip=10.117.30.142,'\
'bmc_username=admin,bmc_password=adminadmin'
```

5. Manage the node addition as appropriate.

   For example, to make the node part of the cluster definition file, enter the following command:

   ```
   # discover --show-configfile
   ```

6. (Optional) Complete one of the following procedures to configure `syslog` forwarding:

   - **Enabling `syslog` forwarding for node, switch, and chassis controllers**
   - **Enabling `syslog` forwarding for unmanaged devices and nodes**

Adding a new compute node or device to the cluster when the MAC address is unknown

This procedure assumes that you need to add compute nodes or other devices to the cluster, but you do not know the MAC addresses of any of these new components. The commands in this procedure help you to complete the following steps:

1. Configure a temporary pool of IP addresses in DHCP. The DHCP server can assign these IP addresses to devices as the devices request them.
2. Collect the MAC addresses information from the devices that received IP addresses.
3. Probe the devices to see if they have PXE booted into the miniroot kernel. Each device that has booted is now known to the cluster manager.
4. Write the new devices to a cluster definition file. You can edit the file and then use the `cm node discover add` command to complete the configuration process for these nodes.

For an overview of this procedure, enter the following command:

```
cm node discover help
```

For more information about the commands in this procedure, enter the following:

```
cm node discover command --help
```

Prerequisites

This procedure assumes the following:

- The cluster has no leader nodes.
- At least one node is configured into the cluster.
- You want to add one or more additional nodes or devices to the cluster, but you do not know the MAC addresses of these new devices.

NOTE: The following are other ways to add nodes to a cluster:

- To add nodes or devices by using the GUI, see one of the following:
    - **Adding nodes**
    - **Importing nodes**
- If you know the MAC addresses of the new nodes or devices you want to add, update the cluster definition file and then use the `cm node add` command as described in the installation guide for your platform.

  For links to the installation guides, see the following:

  Cluster manager documentation

Procedure

1. Install the new compute node in the computer center, and connect the network cables for the node to the cluster.

   Use your hardware documentation for this step.

2. Use the `cm node discover enable` command to start the configuration process.

   This command sets up a pool of IP addresses in the DHCP server and activates the discovery feature. The discovery feature creates a DHCP range of IP addresses on the `head` and `head-bmc` subnetworks. A later step uses the `cm node discover disable` command, which stops the discovery feature.

   The format is as follows:

   ```
   cm node discover enable [-n network] [-i image]
   ```

Variable	Specification
network	If you specify `-n network`, the command creates the DHCP range of IP addresses on the network you specify. By default, the cluster manager creates the IP addresses on the `head` network. If the node controllers are not on the same physical network as the `head` network, specify `-n 'head,head-bmc'`. To determine whether the `head` network and the `head-bmc` network are on the same physical network, check the network cabling.
image	The name of the image to be used to PXE boot the configured nodes. This image must exist on the cluster. To display existing images, use the `cm image show` command. If the cluster nodes include servers with different types of architectures, specify an architecture-appropriate `image_name`. By default, the system selects an image randomly.

For example, the following command starts the process with the following assumptions:

- The `head` and `head-bmc` subnetworks are on the same physical network
- You select the `rhel8.X` image to be used to configure the nodes.

```
# cm node discover enable -i rhel8.X
```

3. Apply power to the node controller in the new server node, or power-on the new server node.
4. Enter the `cm node discover status` command to confirm that a new server node was detected.

   You might have to enter the `cm node discover status` command several times to confirm that the new server node has been detected and has reached the PXE boot stage.

   The command displays status messages that show the IP addresses and MAC addresses for each node. Nodes that are already configured receive predefined IP addresses. The new node receives an IP address from the pool. The new node boots a miniroot kernel from the image specified on the `cm node discover enable` command.

   For example:

   ```
   [root@node64 ~]# cm node discover status
   ===== Leased IPs:

   192.168.64.244 = 04:09:73:b0:c6:80 (3)
   192.168.64.245 = 00:11:0a:51:c5:00 (3)
   192.168.64.246 = 00:9c:02:99:20:ae (3), uid = "\001\000\234\002\231", client-hostname = "ILOUSE233C1JD"
   192.168.64.247 = 00:9c:02:99:20:ac (9)
   ===== Server checks:

   checking if 192.168.64.244 is a server... no.
   checking if 192.168.64.245 is a server... no.
   checking if 192.168.64.246 is a server... no.
   checking if 192.168.64.247 is a server... yes.        # The command can ssh into the miniroot envinronment on this node
   ===== Detected server MAC info:
   IP 192.168.64.247 BMCIP 192.168.64.246
     NIC_MAC 00:9c:02:99:20:ac BMC_MAC 00:9c:02:99:20:ae NIC_IF eno1
   ```

   The last line of output shows two MAC addresses and two IP addresses for the new node. They represent the following:

   - **`00:9c:02:99:20:ac`** is the MAC address of a server. This server received IP address **`192.168.64.247`**.
   - **`00:9c:02:99:20:ae`** is the MAC address of the server node controller, which is an iLO device. The iLO device received IP address **`192.168.64.246`**.

5. Use the `cm node discover mkconfig` command to create a cluster definition file that includes the newly configured servers displayed in the output from the `cm node discover status` command.

   The format is as follows:

   cm node discover mkconfig [-n *hostname*] [-x *num*] [-o *config_attrs*] *config_file*

   The variables are as follows:

Variable	Specification

hostname	Optional. You can use the `-n` option alone or in conjunction with the `-x` parameter to specify node names for the new nodes.
	The format of the `-n` parameter includes the special keywords `%i` and `%n` after the *hostname* variable, as follows:
	• `-n` *`hostname`*`%i`
	• `-n` *`hostname`*`%n`
	These keywords are replaced by a series of numbers in sequence. The `%i` and the `%n` keywords determine the initial number for the sequence, as follows:
	• The `%i` keyword indicates that you want to number the nodes starting with 1 more than the highest-numbered internal hostname in the cluster database.
	• The `%n` keyword indicates that you want to number the nodes starting with a number of your choosing. The chosen number is specified by the -x parameter. If the `-x` parameter is not specified, the default starting number is 1.
	The default behavior is `-n  'node%i'`. This means that if the highest numbered internal hostname is 10, then the hostnames of the new nodes in the new cluster definition file are `node11`, `node12`, and so on.
	Example 1. Assume that you want the default numbering sequence, and your existing nodes are named `worker`*`X`*. Specify the following parameter:
	`-n 'worker%i'.`
	Example 2. Assume that you want to add a new series of three login nodes named with a new base string of `login`. You want the numbering to start at 1. The resulting nodes name are to be `login1`, `login2`, and `login3`. Specify the following parameter:
	`-n 'login%n'`
	Example 3. Assume that you have five login nodes. You want to add a new series of three login nodes named with the existing base string of `login`. You want the numbering to start at `6`. The resulting node names are to be `login6`, `login7`, and `login8`. Specify the following parameters:
	`-n 'login%n' -x 6`

num	Optional. Specify this parameter if the following are both true: • You specified `-n '`*`hostname`*`%n'`. • You want the numeric part of the node names to start with a number other then 1.

Table Continued

Variable	Specification
	For example, to request that the command assign hostnames to three new nodes, and you want the first node to be `fee10`, specify both of the following parameters: `-n 'fee%n' -x 10`

config_attrs	Optional. One or more configuration attributes to assign to the node or device. For a list, see the installation guide for your platform.
	Example 1. Specify the following to add BMC credentials to each new server:
	`-o "bmc_username=admin, bmc_password=password"`
	Example 2. Specify the following to configure consoles:
	`-o "console_device=ttyS1, baud_rate=115200"`

config_file	The name of the file to which you want the cluster manager to write the cluster definition file.

6. After the cluster definition file in Step **5** is created, examine the file and add additional information as needed.

For example, if you did not specify configuration attributes on the command line, add the following information in the cluster definition file:

- The node controller credentials.

  These are required.

- Console settings.

  Hewlett Packard Enterprise recommends that you add the appropriate console settings for the new server. These settings are recommended in order for console logging and console access to work correctly.

7. Use the `cm node discover add` command to configure the new servers into the cluster.

The format is as follows:

```
cm node discover add [-i image] [-d disk] [-s] config_file
```

The parameters and variables are as follows:

Parameter or Variable	Specification
image	Optional. The name of the image to deploy on the new servers. This image must exist on the cluster. To display existing images, use the `cm image show` command. If you do not specify an *image*, the cluster managers powers down the new servers.
disk	Optional. The disk within the new servers upon which to install the image. By default, this is `/dev/sda`.

Table Continued

Parameter or Variable	Specification
-s	Optional. Directs the cluster manager to set credentials on the node controllers of the new servers.
config_file	The name of the cluster definition file.

NOTE: Use the `cm node discover add` command, as shown in this step, to add the new nodes. In this step, do not substitute the `cm node add` command for the `cm node discover add` command.

The `cm node discover add` command adds the new nodes to the cluster database and resets the node controller. In this way, the node controller picks up the newly assigned IP address from the cluster manager.

8. Disable the node configuration process:

```
# cm node discover disable
```

9. (Conditional) Provision the node.

   Complete this step if you did not specify an image in Step **7**.

   Use the `cm node provision` command to provision the node with an image the next time the node boots.

   For example:

```
# cm node provision -n node55 -i rhel8.X -s
```

Enabling `syslog` forwarding for unmanaged devices and nodes

The cluster manager configures the admin node, the ICE leader nodes, and the scalable unit (SU) leader nodes as `syslog` servers. The `syslog` servers listen on all management IP addresses on `tcp` and `udp` port 514.

All compute nodes and leader nodes forward their `syslog` data to the admin node or their leader nodes, if applicable, over the `tcp` `syslog` port.

The procedure in this topic explains how to forward `syslog` data from power distribution units (PDUs), cooling distribution units (CDUs), external nodes, and other unmanaged components.

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to display the management IP address of the admin node or leader node:

```
# cminfo --head_ip
```

3. Refer to the vendor's device-specific documentation for the procedure to configure `syslog` on the device or node.

   For more information, see the following:

   Enabling `syslog` forwarding for node, switch, and chassis controllers

Enabling `syslog` forwarding for node, switch, and chassis controllers

By default, all nodes, switches, and chassis controllers that the `cmcinventory` service auto-discovered are configured to send `syslog` data to the admin node or its assigned scalable unit (SU) leader node.

If the cluster includes nodes, switches, or chassis controllers that the `cmcinventory` service did not auto-discover, use the `cbios` command as described in the following procedure to enable `syslog` manually.

Procedure

1. Log into the admin node as the root user.
2. Use the `cm controller show` command to verify whether the controller for the component has its credentials set in the cluster database.

   For example:

   ```
   # cm controller show --access -c x9000c1s0b0
   NAME          USERNAME  PASSWORD
   x9000c1s0b0   root      U2FsdGVkX1+M+cp8Q/bfUZG/kqv50Wqmu3KnzFiHqro=
   ```
3. (Conditional) Use the `cm controller set` command to configure a username and password for the controller.

   Complete this step if the controller does not have a username or password.

   The format for this command is as follows:

   ```
   cm controller set -c controller --username username --password password
   ```

   The variables are as follows:

Variable	Specification
controller	The controller hostname.
username	The username set in the controller firmware.
password	The password set in the controller firmware.

4. Use the `cbios` command to enable `syslog`.

   For example, the following command enables `syslog` forwarding from rack `x3000c1r1b0` to the admin node:

   ```
   # cbios --SysLogConfAdd -n x3000c1r1b0
   ```
5. Use the `cbios` command in the following format to verify the `syslog` setting:

   ```
   cbios --Show -n controller | grep -A 5 "Syslog"
   ```

   For *controller*, specify the controller hostname.

   Example 1. The following command verifies the `syslog` setting:

   ```
   # cbios --Show -n x9000c1r1b0 | grep -A 5 "Syslog"
     "Syslog": {
       "Port": 514,
       "ProtocolEnabled": true,
       "SyslogServers": [
         "172.23.0.1"
       ],
       "Transport": "udp"
   ```

```
    }
}
```

Example 2. The `cbios` command can accept a different IP address. For this example, assume that node controller `x3000c1s0b0` has nodes that are assigned to an SU leader IP alias. To verify whether a node controller has a node assigned to an SU leader IP alias, run the following command to view the list of nodes under the node controller:

```
# cm controller show -c x3000c1s0b0 --nodes
x3000c1s0b0
   x3000c1s0b0n0
   x3000c1s0b0n1
```

Then use the `cm node show` command to verify whether an SU Leader IP alias is assigned:

```
# cm node show --su-leader -n x3000c1s0b0n0,x3000c1s0b0n1
node: x3000c1s0b0n0: 172.23.255.241
node: x3000c1s0b0n1: 172.23.255.241
```

In this case, the IP alias assigned to the nodes is `172.23.255.241`. Use this IP address to forward `syslog` data. The command is as follows:

```
# cbios --SysLogConfAdd -n x3000c1s0b0 —ip 172.23.255.241
```

Shutting down and restarting a system admin controller with high availability (SAC HA)

Procedure

1. Log into the virtual machine (VM) admin node as the root user.
2. Power-off the cluster:

   ```
   # cm power off -t system
   ```

3. Log into one of the physical admin nodes as the root user.
4. Stop the `virt` resource.

   This step differs depending on the admin node operating system, as follows:

   - On a RHEL admin node, enter the following commands:

     ```
     # pcs resource disable virt
     # pcs resource unmanage virt
     ```

   - On a SLES admin node, enter the following commands:

     ```
     # crm resource stop virt
     # crm resource unmanage virt
     ```

5. Stop the admin node:

   ```
   # shutdown -h now
   ```

6. Log into the other physical admin node as the root user.
7. Stop the admin node:

   ```
   # shutdown -h now
   ```

8. Complete the tasks you needed to complete while the admin node is shut down.

9. Use the `ipmitool` command to power on each physical admin node.

   The format is as follows:

   ```
   ipmitool -I lanplus -H admin_fqdn -U username -P password chassis power on
   ```

   The variables are as follows:

Variable	Specification
admin_fqdn	The fully qualified domain name for the first physical admin node.
username	The username for the root user.
password	The password for the root user.

10. After the login prompt appears on each physical admin node, wait one minute before proceeding to the next step.

11. Verify that the `/images` directory is mounted on each physical admin node.

    Enter the following command on each node:

    ```
    # df -h | grep "/images"
    ```

12. Restart the `virt` resource.

    This step differs depending on the admin node operating system, as follows:

    - On a RHEL admin node, enter the following commands:

      ```
      # pcs resource manage virt
      # pcs resource enable virt
      ```

    - On a SLES admin node, enter the following commands:

      ```
      # crm resource manage virt
      # crm resource start virt
      ```

13. Enter the following command to connect to the virtual admin node via the console:

    ```
    # virsh console sac
    ```

Troubleshooting chassis power up and automatic power down problems on an HPE SGI 8600 cluster

The powering-on process on an HPE SGI 8600 cluster

When you issue a power on, the first chassis controller in a power domain performs the power-on operation. When you power off, the last chassis controller in the power domain performs the power-off operation.

The power-on and the power-off processes occur in phases. When you use the `cm power` command to power up and power down, the command handles the process for you.

When you enter the `cm power on -n 'r*i*n*'` command on an admin node to power up the ICE compute nodes, the following occurs:

1. Each chassis controller turns on the power supplies.

   At this point the node controllers on the compute blades have power, are booted, and are running.

2. The chassis controller enables the fans and waits until it determines that air is moving through the chassis.

3. The chassis controller sends an IPMI command to the node controllers. This command that tells the node controllers to enable power to the compute blades.

4. The node controllers enable power to the compute blades.

Chassis controller monitoring on an HPE SGI 8600 cluster

During typical operation, the chassis controller monitors several aspects of the power supply.

Some cluster systems with ICE leader nodes have M-Racks. These systems use external cooling. On these systems, the chassis controller verifies the following:

- That communication between the chassis controller and its associated CRC and CDU is open. The chassis controller monitors the CRC and CDU for error conditions and, if needed, can power off the chassis. The rack number of the chassis controller determines the CRC and CDU that the chassis controller monitors.
- That the correct number of power shelves can be detected.

Some cluster systems with ICE leader nodes have D-Racks. On these systems, the chassis controller verifies the following:

- That a certain number of fans are present and spinning. Environmental software on the chassis controller controls the fan speed and reports failures.
- That the correct number of power shelves can be detected.

Power cycling the chassis on an HPE SGI 8600 cluster

The chassis controllers enable power to the chassis. If a cluster with ICE leader nodes loses power abruptly, power cycle the chassis. If power supplies are turned off, the LEDs on the power supplies flash green. If one of the power supplies has a fault, the LED is solid amber. Depending on how the software in the system detected the power off, log entries can provide more information. For information about the log entries, see the following:

Power consumption log files on an HPE SGI 8600 cluster

In most cases, if you can power cycle the chassis, the chassis controllers can restore power.

The following procedure explains how to power cycle the chassis.

Procedure

1. Log into the admin node as the root user.

2. (Conditional) Use the `cm node show` command to retrieve the list of ICE leader nodes and chassis controllers.

   Perform this step if you are unsure of the system ID for the affected ICE leader node and chassis controller.

   Enter the following commands:

   ```
   # cm node show -t system leader,cmc
   r1i0c
   r2i1c
   r1lead
   r2lead
   ```

   The preceding commands show the IDs for the leader nodes and chassis controllers on a two-rack system.

3. Use the `cm power` command to retrieve information about the chassis controllers in the rack.

For example:

```
# cm power status -t chassis 'r1i*'
xxxxx
xxxxx
.
.
.
r1i0c: power is On
r2i1c: power is On
```

4. Power off the chassis controllers in the rack.

   For example:

   ```
   # cm power off -t chassis 'r1i*'
   ```

5. Power on the chassis controllers in the rack.

   For example:

   ```
   # cm power on -t chassis 'r1i*'
   ```

Power supplies and the watchdog timer on an HPE SGI 8600 cluster

If all the chassis controllers in the power domain detect a fault condition, the following occur:

- The watchdog timer expires
- The system powers off all the power supplies

When the system is operating as expected, the chassis controllers detect no faults. Under these conditions, the chassis controllers send a watchdog reset every 10 seconds.

The power shelves must receive a watchdog reset once every 45 seconds from each chassis controller in the power domain. If the watchdog timer expires, the power shelf controller disables the power supplies on that shelf and sets the `WDOG` status bit.

The following conditions can prevent the chassis controller from sending the watchdog reset to the power shelves:

- The chassis controller cannot confirm that a minimum number of fans are spinning. This condition pertains to clusters with D-Racks.
- The chassis controller cannot communicate with the external CRC and/or CDU. The chassis controller detects a fault condition reported from the CRC and/or the CDU. These conditions pertain to clusters with M-Racks.

The output from the chassis controller `pfctl status` command shows the status of the `WDOG` status bit. You can enter the `pfctl status` from any chassis controller in the power domain. The command reports power shelf and supply status and reports fan or CRC/CDU status, depending on rack type.

Interpreting the power supply LEDs on an HPE SGI 8600 cluster

The following table explains how to read the status indicators on the power supply LEDs:

If the light is ...	Meaning
Solid green	Power supply is on and OK.
Blinking green	Power supply has AC, but it is not on. If all the power supply LEDs are blinking green, the power was turned off. This situation could result from one of the following: • The watchdog timer firing. • A power down issued by all chassis controllers because of a cooling problem. The cooling problem could be due to one of the following: ◦ The fan controller on a D-Rack system ◦ The CRC/CDU unit on an M-Rack system
Solid amber	The power supply has failed.
Blinking amber	This signal is a power supply warning. The supply is still operating. There is no AC to the power supply, but the power supply is plugged into the system. There is no AC input (under voltage).

Troubleshooting the devices on the CAN bus interface on an HPE SGI 8600 cluster

The CAN bus is the interface that connects all the chassis controllers, power shelves, and D-Rack fan controllers. You can use the `pfctl ping` command to retrieve the status of each device and then take corrective action. To use this command, log into the chassis controller and enter the command at the system prompt.

If the `pfctl ping` command reports missing power shelves, see the following:

Troubleshooting a missing power shelf on an HPE SGI 8600 cluster

Example 1. The following output was obtained for an M-Rack, with two chassis in a power domain.

```
> pfctl ping
PWR-UPPER-CMC1: r1i5c
PWR-UPPER-CMC0: r1i1c
PWR_SHELF3:     -
PWR_SHELF2:     PRESENT
PWR_SHELF1:     PRESENT
PWR_SHELF0:     PRESENT
PWR-LOWER-CMC1: r1i4c
PWR-LOWER-CMC0: r1i0c

EXTERNAL FANS
```

Example 2. The following output was obtained on an M-Rack, with one chassis in a power domain and twin node blades.

```
> pfctl ping
PWR-UPPER-CMC1: -
PWR-UPPER-CMC0: r1i4c
PWR_SHELF3:     -
PWR_SHELF2:     PRESENT
PWR_SHELF1:     PRESENT
```

```
PWR_SHELF0:      PRESENT
PWR-LOWER-CMC1: -
PWR-LOWER-CMC0: r1i0c
EXTERNAL FANS
```

Example 3. The output in this example is from a D-Rack. The fan controller hosts two programmable systems on a chip (PSOC) units. The fan controller controls 12 fans. The following output shows the fan controllers that appear in the `FAN–CONTROL` lines as `PRESENT`. This output is typical for a system that is operating properly.

```
> pfctl ping
PWR-UPPER-CMC1: -
PWR-UPPER-CMC0: r1i1c
PWR_SHELF3:      -
PWR_SHELF2:      -
PWR_SHELF1:      PRESENT
PWR_SHELF0:      PRESENT
PWR-LOWER-CMC1: -
PWR-LOWER-CMC0: r1i0c

FAN-UPPER-CMC1: -
FAN-UPPER-CMC0: r1i1c
FAN-CONTROL1     PRESENT
FAN-CONTROL0     PRESENT
FAN-LOWER-CMC1: -
FAN-LOWER-CMC0: r1i0c
```

Troubleshooting a missing power shelf on an HPE SGI 8600 cluster

It is possible for a power shelf to be physically present but not appear in `pfctl ping` command output. In this case, use the information in the following topics to troubleshoot:

- **Booting a power shelf manually on an HPE SGI 8600 cluster**
- **Fixing problems related to a newly installed power shelf on an HPE SGI 8600 cluster**

Booting a power shelf manually on an HPE SGI 8600 cluster

Occasionally, when the AC power breakers are enabled, the power shelf controller or the fan controller might not boot properly. The power shelf is said to be **wedged** in this situation. In this case, complete the following procedure to power cycle the AC power to all the power supplies in the power domain or cooling domain.

Procedure

1. Power cycle the system again.

   Use the following procedure:

   Power cycling the chassis on an HPE SGI 8600 cluster

2. Manually flip the power breakers on the power distribution unit (PDU) at the top of the rack.

3. Enter the following command from the chassis controller:

   ```
   > pfctl ping
   ```

4. Examine the output.

   The output shows the correct number of power shelves as present. The following examples show correct output for their specific systems.

Example 1. The following output was obtained for the D-Rack of a cluster:

```
> pfctl ping
PWR-UPPER-CMC1: -
PWR-UPPER-CMC0: r1i1c
PWR_SHELF3:     -
PWR_SHELF2:     -
PWR_SHELF1:     PRESENT
PWR_SHELF0:     PRESENT
PWR-LOWER-CMC1: -
PWR-LOWER-CMC0: r1i0c

FAN-UPPER-CMC1: -
FAN-UPPER-CMC0: r1i1c
FAN-CONTROL1    PRESENT
FAN-CONTROL0    PRESENT
FAN-LOWER-CMC1: -
FAN-LOWER-CMC0: r1i0c
```

Example 2. The following output was obtained for the M-Rack of a cluster, with one chassis in a power domain and single-node blades:

```
> pfctl ping
PWR-UPPER-CMC1: -
PWR-UPPER-CMC0: -
PWR_SHELF3:     -
PWR_SHELF2:     PRESENT
PWR_SHELF1:     PRESENT
PWR_SHELF0:     PRESENT
PWR-LOWER-CMC1: -
PWR-LOWER-CMC0: r1i0c

EXTERNAL FANS
```

Fixing problems related to a newly installed power shelf on an HPE SGI 8600 cluster

If you recently added or replaced a power shelf, the following problems might exist:

- The firmware on the new power shelf was not flashed
- A problem might exist with the CAN bus connection

The following procedure explains how to troubleshoot a new power shelf that is not integrated properly.

Procedure

1. Log into the chassis controller.

   For systems with M-Racks, log into the lower chassis controller in the chassis.

   For systems with D-Racks, log into the lower chassis controller in the pair.

   For information about how to log into the chassis controller, see the following:

   Power cycling the chassis on an HPE SGI 8600 cluster

2. Enter the following command to change to the directory that contains the firmware images:

   ```
   > cd /usr/local/firmware/psoc
   ```

3. Use the following command to flash the firmware:

   ```
   flashcan -f image -p controller -r
   ```

   The variables are as follows:

Variable	Specification
image	The name of one of the firmware images from the following directory: `/usr/local/firmware/psoc`
controller	One of the following: • `0`, which specifies power shelf 0. • `1`, which specifies power shelf 1. • `2`, which specifies power shelf 2. • `3`, which specifies power shelf 3. • `4`, which specifies fan controller 0. • `5`, which specifies fan controller 1.

4. Enter the `pfctl ping` command to retrieve the status of the power shelf.

   If the status returns `PRESENT` for the problem power shelf, you are finished.

   If the command does not return `PRESENT`, continue with this procedure to troubleshoot other causes of the problem.

5. Perform one or more of the following remedies:

- Reseat the power shelf.
- Visually inspect the connector on the shelf and make sure that it is correct.

  Inspecting the blind connector at the rear of the power shelf slot can be difficult to do.
- Visually inspect the LED lights.

  If there were power supplies in the shelf with the fail LED light, the failed supply might have damaged the power shelf.

  If a power supply turns on immediately when the AC power is applied, the shelf itself might be damaged.

  If the following are both true, the power shelf might be bad:

  - If the power supplies in all the other shelves are off (flashing green)
  - If the power supplies in the missing shelf turn solid green as soon as the breaker is enabled

  Find the sticker on the power shelf, and see if it is discolored.
- Inspect the CAN bus cable in the back on the rack.

Power consumption log files on an HPE SGI 8600 cluster

On a chassis controller, the following log files contain information that can help you troubleshoot a power problem:

- The /tmp/pfctld.log file contains entries from the power and fan control daemon, pfctld. When the pfctld daemon powers down a chassis, it records a log entry in the log file. The entry includes the reason for the power down.
- The /tmp/eric.log file contains output from an environmental software monitoring application, called ERIC. ERIC runs on the chassis controller. The actions of ERIC actions are written to this log file. ERIC monitors blade temperatures and adjusts fans speeds appropriately. ERIC also monitors the chassis controller inlet air temperature and powers down the chassis when appropriate. That is, ERIC powers down the blades associated with that chassis controller.

  If the cluster has D-Racks, and the following conditions are all present, the problem might be related to the chassis controller air inlet temperature:

  - The blades of only one chassis are powered down
  - The blades in the other chassis are still on
  - Power supply LEDs are solid green

  In an M-Rack configuration, there could be a problem with the chassis controller air inlet temperature. In this case, ERIC could power off only the upper or lower board in the blade.

Retrieving information about the power supplies on an HPE SGI 8600 cluster

After you log into the chassis controller, you can use the pmbus_drack and pmbus_mrack scripts to retrieve information about the power supplies. These scripts dump some of the PMBus data that is available.

The following example shows output from the pmbus_mrack script:

```
> pmbus_mrack
Shelf0 PS0 Vout: 12          Iout: 3.5        Temp: 29        Status: 0x0000
Shelf0 PS1 Vout: 11.6875     Iout: 0          Temp: 29.5      Status: 0x0000
Shelf0 PS2 Vout: 12.0312     Iout: 1          Temp: 28.5      Status: 0x0000
Shelf1 PS0 Vout: 11.625      Iout: 0          Temp: 29        Status: 0x0000
Shelf1 PS1 Vout: 11.6562     Iout: 0          Temp: 29.5      Status: 0x0000
Shelf1 PS2 Vout: 11.6875     Iout: 0          Temp: 28.5      Status: 0x0000
Shelf2 PS0 Vout: 11.6875     Iout: 0          Temp: 28.5      Status: 0x0000
Shelf2 PS1 Vout: 11.625      Iout: 0          Temp: 28        Status: 0x0000
Shelf2 PS2 Vout: 11.6562     Iout: 0          Temp: 29        Status: 0x0000
Shelf3 PS0 Vout:             Iout:            Temp:           Status: not available
Shelf3 PS1 Vout:             Iout:            Temp:           Status: not available
Shelf3 PS2 Vout:             Iout:            Temp:           Status: not available
```

The output shows the output voltage (Vout), the output current (Iout), and temperature from each of the power supplies.

You can use the Iout values to determine whether power supply load sharing is working correctly within the power domain. All power supply readings ideally report within +/- 10% of the average. The status bits record warnings and faults when they occur. All bits are decoded if present.

The following status messages can appear in the pmbus_mrack output:

VOUT

Output voltage warning or fault.

IOUT

Output current warning or fault.

INPUT

Input fault.

MFR

Manufacturer fault. Related to the 3.3v auxiliary supply used to power the power shelf controllers, fan shelf controllers, and the chassis controllers.

PWRGOOD

Power output is good (active low).

FANS

Internal fan failure.

OTHER

Another warning or fault not indicated by other status flags.

UNKNOWN

An internal power supply controller condition was detected.

OFF

Power supply is off.

VOUT_OV

Output voltage over limit.

IOUT_OC

Output current over limit.

VIN_UC

Input voltage under limit.

TEMP

Temperature warning or fault.

CML

Communication error. Can be ignored.

NOTA

None of the above.

Power supplies shut down on any fault condition and remain off unless the fault is a temperature fault. After the power supply has cooled, it re-enables itself. Generally, if you cycle the AC power to the faulted power supply, it resets all status flags. Hard failures reoccur. If the system is under heavy load and a power supply fails, the other supplies pick up the load. If yet another supply fails, this failure can cause an overcurrent across all supplies. An overcurrent can power down all compute blades in the power domain.

Retrieving information about the PMBus registers on an HPE SGI 8600 cluster

After you log into the chassis controller, you can use the `pfctl pmbus dump` command to retrieve information about the PMBus registers. This command queries all power supplies in the power domain. In the command output, look for nonzero readings to locate possible problems.

The following example shows output from the `pfctl pmbus dump` command:

```
> pfctl pmbus dump
PWR s0s0               VIN:   213.00
PWR s0s0               IIN:     0.78
PWR s0s0              VOUT:    11.97
PWR s0s0              IOUT:     9.00
PWR s0s0          3.3 VOUT:     3.34
PWR s0s0          3.3 IOUT:     1.34
PWR s0s0              TEMP:    23.00
PWR s0s0          FAN1 RPM: 6320
PWR s0s0          FAN2 RPM: 6320
PWR s0s0       STATUS_BYTE: 00
PWR s0s0       STATUS_WORD: 0000
PWR s0s0       STATUS_VOUT: 00
PWR s0s0       STATUS_IOUT: 00
PWR s0s0      STATUS_INPUT: 00
PWR s0s0 STATUS_TEMPERATURE: 00
PWR s0s0        STATUS_CML: 00
PWR s0s0        STATUS_3V3: 00
PWR s0s0   STATUS_FANS_1_2: 00
PWR s0s0        SMB ALERT : 00 NO
PWR s0s0 SOFTWARE REVISION : pri 167 app 169 boot 2
PWR s0s0            PMBUS: I 1 II 1
PWR s0s0               ID: DELTA
PWR s0s0            MODEL: AHF-2DC-2837W-12V-240V
PWR s0s0         REVISION: 1 6 167 169
PWR s0s0         LOCATION: DES
PWR s0s0             DATE: 23/10 (13:58:00 06/08/10)
PWR s0s0           SERIAL: A000379
```

Booting a leader node or a compute node

The information in this topic might be useful to troubleshoot boot problems. The default boot process is identical for leader nodes and compute nodes and leader nodes on an installed cluster. The process assumes the following environment for a node:

- The node is configured in the cluster database.
- The node is registered in clusterwide DHCP server database.
- At least one node image is available for this node in the admin node image repository. An image is assigned to the node.
- The GRUB2 network configuration files are available.
- The node has been installed.
- The BIOS hardware on the node is configured for network booting. This state is the factory-defined state.
- You have not configured the node to boot from a local disk.

The following topics describe the boot process:

1. **Phase 1 - Initiating the boot**
2. **Phase 2 - Loading the kernel for the node**
3. **Phase 3 - Loading the miniroot**
4. **Phase 4 - Starting the operating system on the node**

For information about the boot-from-disk feature, see the following:

Booting leader nodes or compute nodes from a local disk

NOTE: The node boot process differs for configured clusters versus unconfigured clusters.

Phase 1 - Initiating the boot

This topic explains the first set of events that occur when you boot a leader node or a compute node. For general information and a process summary, see the following:

Booting a leader node or a compute node

The following events occur when you initiate a boot:

1. From the admin node, a user enters the following command:

   ```
   cm power on -t leader node
   ```

   For example, on a scalable unit (SU) leader node, enter the following:

   ```
   # cm power on -t leader leader1
   ```

2. The node turns on.
3. The factory-defined boot process starts. The boot proceeds as follows:
   - The BIOS boots from the network interface card (NIC).
   - The ROM on the NIC activates the PXE protocol to boot over the network.
   - The ROM sends a DHCP request for node network information. The node IP address is a static address. In this case, the admin node receives the DHCP request. If this node is an ICE compute node that is managed by a leader node, the leader node receives the request. The DHCP configuration files reside as follows:
     - On RHEL systems:

       ```
       /etc/dhcp/dhcpd.conf.d/ice.conf
       ```

     - On SLES systems:

       ```
       /etc/dhcpd.conf.d/ice.conf
       ```

     Include files reside in the `dhcpd.conf.d` directory.
4. The admin node or leader node responds to the DHCP request. It sends DHCP packets that include the following:
   - The network configuration of the node.
   - The location of the GRUB2 boot loader.

     The loader resides on the admin node in one of the following locations:
     - On x86_64 legacy platforms, this location is as follows:

       ```
       /opt/clmgr/tftpboot/grub2/i386-pc/grub-cm-i386.0
       ```

     - On x86_64 UEFI platforms, this location is as follows:

       ```
       /opt/clmgr/tftpboot/grub2/x86_64-efi/grub-cm-x86_64.efi
       ```

5. The NIC loads GRUB2.

6. The admin node directs the node to boot GRUB2.
7. The NIC starts.

Phase 2 - Loading the kernel for the node

This topic explains the second set of events that occur when you boot a leader node or a compute node. For general information and a process summary, see the following:

Booting a leader node or a compute node

In Phase 2, GRUB2 starts up and loads the kernel onto the node. All requests and all transfers are done using TFTP.

The final steps in this subprocess are for GRUB2 to load the following:

- The kernel for the node
- The operating system initialization files

The cluster manager uses the kernels and initialization files for booting. These files reside on the admin node in a directory that match the image name. This directory is as follows:

`/opt/clmgr/tftpboot/images/`*`os_name_and_path`*

For example, for RHEL 8.X, the paths are as follows:

- The kernel path is as follows:

  `/opt/clmgr/tftpboot/images/rhel8.X/vmlinuz-3.10.0-957.el7.x86_64`

- The `initr` path is as follows:

  `/opt/clmgr/tftpboot/images/rhel8.X/initramfs-3.10.0-957.el7.x86_64`

The steps are as follows:

1. GRUB2 sends a TFTP request to the admin node for a configuration file.
2. The admin node responds to GRUB2 by sending the following file:

   `/opt/clmgr/tftpboot/grub2/grub.cfg`.

   The file includes instructions that explain how to load kernel files and `initrd` files.

3. The `grub.cfg` file requests the node-specific configuration file.

   The configuration file resides on the admin node in the following location, which includes the node IP address:

   `/opt/clmgr/tftpboot/grub2/cm/`

   For example, the configuration file might reside in the following file:

   `/opt/clmgr/tftpboot/grub2/cm/172.23.0.2.cfg`

   The node configuration file includes the following information:

   - The kernel and the kernel parameters that the node needs.
   - The `initrd` file that the node needs.
   - The image assigned to the node. The image itself resides in the following location:

     `/opt/clmgr/image/images`

- Instructions for booting the node. The file also includes instructions for installing the node.
- The miniroot images reside in the following directory:

  `/opt/clmgr/image/miniroot/squeezed`

  Depending on the transport method used, sometimes the miniroot is compressed.

4. GRUB2 saves the kernel and `initrd` over the network and into memory.
5. GRUB2 boots the kernel.
6. The kernel starts, and then the kernel loads `initrd` from memory.

Phase 3 - Loading the miniroot

This topic explains the third set of events that occur when you boot a leader node or a compute node. For general information and a process summary, see the following:

Booting a leader node or a compute node

In Phase 3, `initrd` loads the miniroot. The steps are as follows:

1. `initrd` completes the following actions:

   - The `initrd` daemon loads the miniroot using UDPcast or `rsync`.
   - It uses `udpcast` or `rsync` to obtain the miniroot.
   - It runs additional internal scripts associated with the node image.

2. UDPcast transfers the miniroot from the port number assigned to the miniroot to the node.
3. The `initrd` daemon unpacks the miniroot.
4. The miniroot begins operating on the node.

   It runs additional internal scripts associated with the node image.

Phase 4 - Starting the operating system on the node

This topic explains the fourth set of events that occur when you boot a leader node or a compute node. For general information and a process summary, see the following:

Booting a leader node or a compute node

In Phase 4, the miniroot starts running processes on the node, and the operating system takes over. The steps are as follows:

1. The miniroot finds the root and boot file systems.
2. The miniroot mounts the root and boot file systems.
3. The operating system distribution startup scripts start to run.

Overriding installation scripts

When creating or updating the miniroot, the cluster manager copies files from the `/opt/clmgr/lib/` directory on the admin node into the miniroot. These files are as follows:

```
/opt/clmgr/lib/miniroot-init
/opt/clmgr/lib/miniroot-functions
/opt/clmgr/lib/miniroot-node-install
/opt/clmgr/lib/miniroot-admin-install
```

The preceding scripts drive node installation and booting.

If a version of any of the following files exists in the `/opt/clmgr/lib` directory, it is used in place of the original:

```
/opt/clmgr/lib/miniroot-init-local
/opt/clmgr/lib/miniroot-functions-local
/opt/clmgr/lib/miniroot-node-install-local
/opt/clmgr/lib/miniroot-admin-install-local
```

For example, if `/opt/clmgr/lib/miniroot-node-install-local` exists, then the cluster manager copies `/opt/clmgr/lib/miniroot-node-install-local` into the miniroot as `/opt/sgi/lib/miniroot-node-install`.

When you create a `-local` file, you override the defaults. Rather than creating `-local` files, Hewlett Packard Enterprise recommends that you use the following features:

- Custom partitioning. For information about custom partitioning, see the following:
  - The installation guide for your platform. For links to the installation guides, see the following:

    Cluster manager documentation
  - **Creating custom partitions**
- Disk reservations. This feature enables you to reserve space at the end of the disk for a scratch space.
- System imager pre-installation and post-installation scripts.

CAUTION: Hewlett Packard Enterprise does not recommend that you override the default files. Hewlett Packard Enterprise updates these files with features and fixes in both patches and in releases. The updated content is not reflected in locally managed copies. If you use this feature, make sure to update the code in your version to match the versions from the cluster manager. To avoid customized versions becoming outdated or incompatible between releases, perform merges. Patches do not touch the customized files, and patches might fix important bugs.

The following operations update the miniroot, including copying the files:

```
cinstallman --update-miniroot
cinstallman --zypper-image     # as long as --duk not present
cinstallman --dnf-image        # as long as --duk not present
cinstallman --yum-image        # as long as --duk not present
cm image update -i image --miniroot
cm image yum
cm image dnf
cm image zypper
```

Retrieving cluster manager service status information

The `configure-cluster` command requires the `cmdb` service to be running. The following procedure explains how to obtain the status of the `cmdb` service.

Procedure

1. Log into the admin node as the root user.
2. Enter the following command:

```
# systemctl status cmdb.service
```

3. In the output, verify that the `cmdb` service is running.

Cluster manager command log files

Each cluster manager command logs information in a dedicated log file. All log files are available in `/opt/clmgr/log`.

Monitoring log files

The `/opt/clmgr/log/MainMonitoringDaemon_MGTXXX.log` file contains the output of the monitoring daemon running on the admin node.

The `/opt/clmgr/log/SmallMonitoringDaemon_NodeXXX.log` file contains the output of the monitoring daemon running on the compute node.

Designate one of the compute nodes as the secondary server for the network group. The `/opt/clmgr/log/SecondaryServerMonitoring_NodeXXX.log` file contains the output of the secondary server process.

`cmuserver` log files

When using the GUI, all actions are sent to the `cmuserver` daemon running on the admin node. The `/opt/clmgr/log/cmuserver-0.log` file on the admin node contains the current output of the `cmuserver` process.

Assume that you enter the following `systemctl` command:

```
systemctl action cmdb.service
```

For *action*, the command accepts one of the following:

- `restart`
- `start`
- `status`
- `stop`

The cluster manager logs the results of the `systemctl` command to `/var/log/cmuservice_`*`hostname`*`.log` .

In the log file name, the *hostname* component is the hostname of the admin node.

Preserving log files during an upgrade

The cluster manager preserves old versions of log files and rotates the log files. By default, the cluster manager retains many versions for 365 days.

To retain log files for longer than 365 days, edit the following file:

```
/etc/logrotate.d/conserver
```

The cluster manager has an additional log rotate service for managing consoles and hosts. To modify the default log rotation behavior, edit one of the following two files:

- `/etc/sysconfig/cm-logrotate-parallel-hosts`
- `/etc/sysconfig/cm-logrotate-parallel-consoles`

When you specify to retain a log file for more than 365 days, the cluster manager preserves the log file after you upgrade to a newer HPE Performance Cluster Manager release level.

For more information, see the `logrotate` man page.

Administrative commands fail to run

Procedure

1. Verify that `ssh` is configured on the nodes.
2. Verify that the `ssh` root is enabled with the node password to all nodes of the cluster.
3. Verify that the database contains the correct IP address and host name.

GUI problems

Cluster manager GUI cannot be launched from browser

Symptom

The GUI does not launch from the browser.

Solution 1

Action

1. Clear the browser cache.
2. Clear the Java cache using the Java Control Panel applet.

   Run the appropriate tool (for example, `jcontrol`) to access the Java Control Panel.

Solution 2

Cause

Browser proxy settings are blocking the GUI launch.

Action

If you receive a certificate validation error while launching the GUI, check the network settings in the Java Control Panel applet. If it is set to use browser settings, browser proxy settings might be blocking the GUI launch. Try using `Direct` `connection`" in the the Java Control Panel. Run the appropriate tool (for example, `jcontrol`) to access the Java Control Panel.

Solution 3

Action

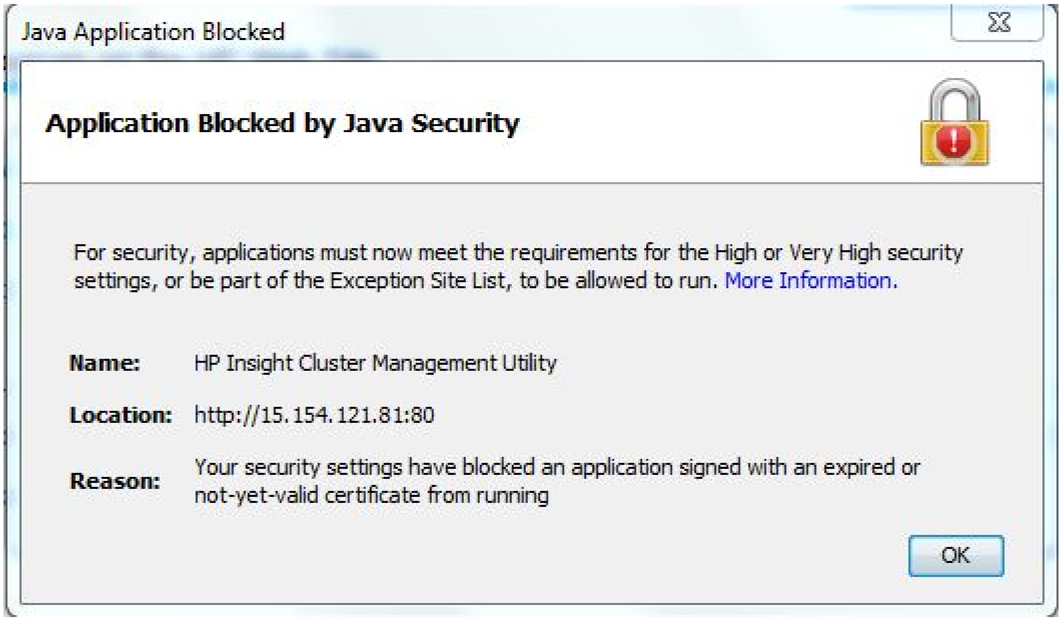
1. If you receive a certificate expiration error while launching the GUI, add the admin node IP address to the exception list in the Java Control Panel applet and launch the GUI again.

   a. **Control panel** > **java** > **Security** > **Exception Site list**

   b. In the **Location** field, enter the IP address of the admin node.

   c. Click **Add**.

      A security warning for HTTP location displays.



Java Application Blocked

Application Blocked by Java Security

For security, applications must now meet the requirements for the High or Very High security settings, or be part of the Exception Site List, to be allowed to run. More Information.

Name: HP Insight Cluster Management Utility

Location: http://15.154.121.81:80

Reason: Your security settings have blocked an application signed with an expired or not-yet-valid certificate from running

   d. Click **OK**.

      The site IP address is added to the exception list.

   e. Click **OK**.

GUI cannot contact the remote cluster manager service

Symptom

The GUI cannot contact the remote cluster manager service.

Action

1. Enter the following command to verify that the cluster manager service is running properly on the admin node:

```
# systemctl status cmdb.service
```

If the cluster manager service is not running properly, enter the following commands to stop and then start the service:

```
# systemctl stop cmdb.service
# systemctl start cmdb.service
```

2. Verify that the GUI on the client system is connected to the correct server.
3. Verify the `GENERAL_RMI_HOST` setting in the `cmu_gui_local_settings` file on the client system.
4. If `cmuserver` is running properly on the admin node, verify the following:

    - The firewall configuration on the admin node and the client system.
    - The RMI connections. Verify that RMI connections are allowed between the two hosts.

GUI is running, but the monitoring sensors are not updated

Symptom

The GUI is running, but the monitoring sensors are not been updated.

Action

1. Verify that the cluster manager service is running properly on the admin node.

    ```
    # systemctl status cmdb.service
    ```

    If the cluster manager service is not running properly, enter the following commands to stop and then start the cluster manager service:

    ```
    # systemctl stop config_manager.service
    # systemctl stop cmdb.service
    # systemctl start cmdb.service
    # systemctl start config_manager.service
    ```

2. Verify that the host file of the nodes is properly configured. Each node must have access to the IP address of all other nodes in the cluster.
3. Verify that `rsh` or `ssh` is enabled between all nodes of the cluster and the admin node. All nodes must be able to execute commands as `root` for any other node without needing a password.
4. Verify that the cluster manager RPM is properly installed on all nodes.

    The following commands return information that shows the RPM being properly installed:

    - On the admin node:

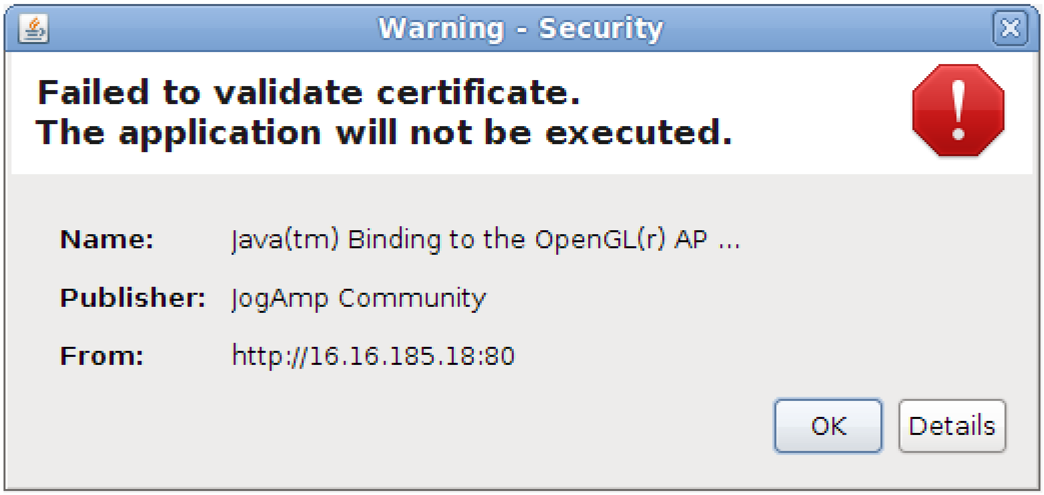
        ```
        admin:~ # rpm -q cmu
        cmu-X.X.xxx.release.xxx.x86_64
        ```

    - On nodes other than the admin node:

        ```
        n1:~ # rpm -q cmu_cn
        cmu_cn-X.X.xxx.release.xxx.x86_64
        ```

Failed to validate certificate error displays

If the GUI is unable to start, the following **Failed to validate certificate** message displays:



Failed to validate certificate.
The application will not be executed.

Name: Java(tm) Binding to the OpenGL(r) AP ...

Publisher: JogAmp Community

From: http://16.16.185.18:80

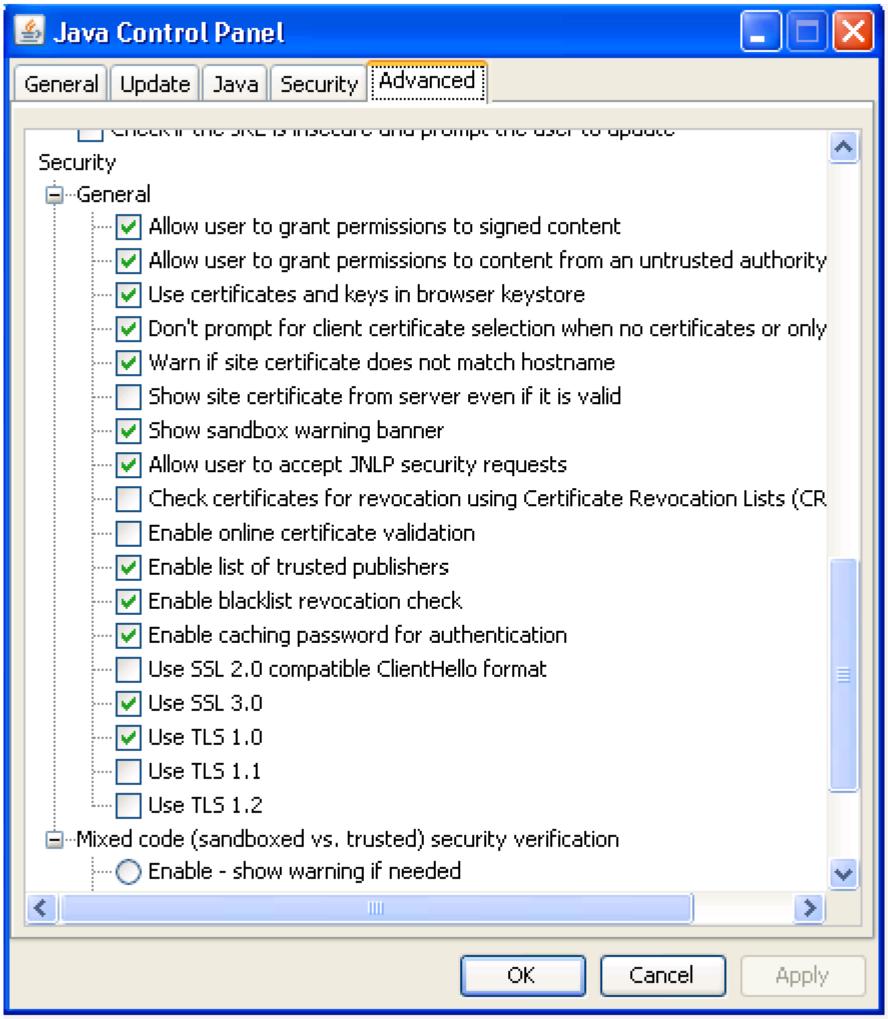
Figure 16: Certificate error

The detailed Java exception is as follows:

```
java.security.cert.CertPathValidatorException:
java.security.InvalidKeyException: Wrong key usage
```

Change the Java setting value. The default value changed between Java 1.6u31 and Java 1.7u12.

If you are connected to the Internet, activate **Enable online certificate validation**. If you are not connected to the Internet, deactivate **Enable online certificate validation**.



□ Check if the JRE is insecure and prompt the user to update

Security

- General
  - ☑ Allow user to grant permissions to signed content
  - ☑ Allow user to grant permissions to content from an untrusted authority
  - ☑ Use certificates and keys in browser keystore
  - ☑ Don't prompt for client certificate selection when no certificates or only
  - ☑ Warn if site certificate does not match hostname
  - □ Show site certificate from server even if it is valid
  - ☑ Show sandbox warning banner
  - ☑ Allow user to accept JNLP security requests
  - □ Check certificates for revocation using Certificate Revocation Lists (CR
  - □ Enable online certificate validation
  - ☑ Enable list of trusted publishers
  - ☑ Enable blacklist revocation check
  - ☑ Enable caching password for authentication
  - □ Use SSL 2.0 compatible ClientHello format
  - ☑ Use SSL 3.0
  - ☑ Use TLS 1.0
  - □ Use TLS 1.1
  - □ Use TLS 1.2
- Mixed code (sandboxed vs. trusted) security verification
  - ○ Enable - show warning if needed

Figure 17: Java control panel

On Windows GUI client nodes, go to **System Preferences** > **Other** > **Java** > **Advanced** > **Enable online certificate validation**.

On Linux, run `javaws -viewer` in a shell, click the **Advanced tab**, then **Enable online certificate validation**.

TIP: If you still encounter problems, try toggling the setting.

Image creation fails

Symptom

When the cluster manager fails to create an image, it notifies you in one of the following ways:

- In the following console message:

  ```
  yume operation exited with 106, command was: command
  ```

- In the following log message written to the `/var/log/cinstallman` file:

  ```
  Download (curl) error for 'http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.8-sles15spX-x86_64/media.1/media':
  Error code: Curl error 60
  Error message: SSL certificate problem: unable to get local issuer certificate
  ```

Cause

The image creation process uses `admin` as a hostname to reach the repositories stored on the admin node. The cluster manager uses the admin node web server to access the image repositories. Tools such as `zypper` try to go through the web proxy instead of the admin node web server, which causes image creation to fail. To retain the web proxy on the admin node, configure certain addresses not to use the proxy.

The following example shows the console messages that the cluster manager generates:

```
# chroot /opt/clmgr/image/images/sles15spX yume --rpmmgr zypper \
--repo +yast2+http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.8-sles15spX-x86_64 \
--repo +yast2+http://admin/repo/opt/clmgr/repos/distro/sles15spX-x86_64 --non-interactive update \
--auto-agree-with-licenses
yume operation exited with 106, command was: chroot /opt/clmgr/image/images/sles15spX yume --rpmmgr zypper \
--repo +yast2+http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.8-sles15spX-x86_64 \
--repo +yast2+http://admin/repo/opt/clmgr/repos/distro/sles15spX-x86_64 --non-interactive update \
--auto-agree-with-licenses
This error code is specific to zypper and indicates a failed package installation.
See man page for zypper for details about this code.
```

In the preceding example, the first line of output shows the message.

Action

1. Set the following environment variable on the admin node:

   `no_proxy="localhost, 127.0.0.1, admin"`

   You can also include this setting in an `/etc/profile.d` script, which sets the correct environment for all users when they log in. This setting ensures access to the admin node web server for image creation. This setting works even if the web proxy is set up for other domains or IP addresses.

2. Verify the proxy setting.

   The verification differs, depending on operating system, as follows:

   - On RHEL systems, the proxy information resides in the following file:

     `/etc/environment`

     The following shows how to retrieve this information from the `bash` shell on a RHEL system:

     ```
     node:~$ printenv |grep no_proxy
     no_proxy=localhost,127.0.0.1,admin,cm.clusterdomain.com
     ```

   - On SLES systems, examine the following file to verify the definition:

     `/etc/sysconfig/proxy`

     The following is an example entry:

     ```
     # Example: NO_PROXY="www.me.de, .do.main, localhost"
     #
     NO_PROXY="localhost,127.0.0.1,admin,cm.clusterdomain.com"
     ```

Node provisioning fails when the image contains an updated kernel

Symptom

During node provisioning, the cluster manager issues the following message:

```
cinstallman: ERROR: unable to find initrd in /opt/clmgr/tftpboot/images/image_name
```

Cause

The kernels and the initrd are not synchronized within the image. This situation can occur when someone has updated the image manually.

Action

Use the `cm image` command to update the image and include the `-k` or `--kernels` parameter on the command line.

When you specify the `-k` or `--kernels` parameter, the command updates the image kernels.

For example, use the command with the following parameters:

```
cm image update -k -i image_name
```

Cannot `ssh` from a scalable unit (SU) leader node to a compute node

Symptom

The system issues messages, and access from an SU leader node to a different, target node is not granted

Cause

The domain name service (DNS) search cannot be configured to allow `ssh` from an SU leader node to a different node by using only the target node hostname.

Action

Append the network to the hostname on the `ssh` command.

Example 1.

```
# ssh n11.head
```

Example 2. Assume that you are logged into a scalable unit (SU) leader node on an HPE Cray EX cluster. You want to reach an HPE Cray EX compute node over the management network. Enter the following command:

```
# ssh x1000c1s5b0n0.hostmgmt
```

If the SU leader node has an HPE Slingshot adapter installed, enter the following command to reach the HPE Cray EX compute node over the HPE Slingshot network:

```
# ssh x1000c1s5b0n0.hsn0
```

Preventing a node from booting after a thermal event

Symptom

After a thermal event occurs on a node, it is possible that the node might try to reboot. You can set the BIOS to prevent the node from booting after a thermal event.

Action

1. Use the `ssh` command to log into the node.

2. Enter the following command to reset the BIOS:

   ```
   # cm node bios set --critical-temp-remain-off -n "x*"
   Bio Change: {u'Oem': {u'Hpe': {u'CriticalTempRemainOff': True}}}
   ```

Managing firmware

The cluster manager supports firmware management. For example, you can view and compare BIOS settings and BIOS firmware versions across a set of chosen nodes. This functionality helps you confirm that your cluster is configured correctly and consistently. You can also run a firmware executable on a set of nodes to upgrade your firmware to the latest version available. The firmware executable is an online flash component.

NOTE: Do not use the information in this chapter to flash firmware on an HPE Cray EX supercomputer. For firmware flashing instructions, see the cluster manager release notes or the appropriate hardware documentation for the component. The following topic explains how to access the cluster manager release notes:

Cluster manager documentation

Firmware management on an HPE Apollo 9000 cluster

The following commands flash firmware on HPE Apollo 9000 clusters:

- The `cm node firmware` command. This command flashes the cluster firmware on a specific node.
- The `cm chassis cmc firmware` command. This command flashes the chassis controller firmware and the tray controller firmware.

NOTE: The preceding commands do not flash managed Ethernet top-of-rack switch firmware.

The following topics explain how to flash firmware:

- **Flashing the firmware on an HPE Apollo 9000 compute node**
- **Flashing the firmware on an HPE Apollo 9000 chassis controller**

Flashing the firmware on an HPE Apollo 9000 compute node

The `cm node firmware` command flashes the following devices within a compute node:

- The iLO device
- The BIOS
- The CPLD

If your goal is to flash the BIOS firmware, flash the iLO device before you flash the BIOS. The procedure in this topic shows this order.

For more information about the `cm node firmware` command, enter one or more of the following commands at the system prompt:

- `cm node firmware -h`
- `cm node firmware show -h`

  The `cm node firmware show` command accepts the `--cmdiff` parameter. The inclusion of `--cmdiff` lets you compare firmware on multiple nodes interactively.

- `cm node firmware status -h`
- `cm node firmware update -h`

Procedure

1. Contact your HPE representative to obtain the compute node firmware.
2. Log into the admin node as the root user.
3. Write the new firmware to the following location on the admin node:

   ```
   /opt/clmgr/tftpboot
   ```

4. Make sure that all the firmware files are world-readable.

   For example:

   ```
   # cd /opt/clmgr/tftpboot
   # chmod 744 *
   ```

5. Flash the iLO device.

   Always flash the iLO device firmware before you flash the BIOS firmware. Use the following command:

   ```
   cm node firmware update -n node iLO_firmware_file
   ```

   The variables are as follows:

Variable	Specification
node	One or more node hostnames.
firmware_file	The name of the iLO firmware file. This file is of the following form: `iLOidentifier.bin`

   For example, the following command flashes the iLO firmware for all nodes in rack 1:

   ```
   # cm node firmware update -n 'r1*' iLO5_FW_BIN_212-ilo5_212.bin
   ```

6. (Optional) Enter one or more of the following commands to monitor the flash:

   ```
   cm node firmware status -n node
   ```

   For *node*, specify the same node hostnames you specified in the previous step.

   This command shows the latest firmware update status, if any. Typically, the flash takes 2-3 minutes.

   After flashing, the cluster manager restarts the iLO.

7. (Conditional) Flash the BIOS firmware.

   Complete this step if you want to flash the BIOS firmware at this time.

   To flash the BIOS firmware, always flash the iLO firmware first, as shown in this procedure.

   Complete the following steps:

   a. (Recommended) Power down the node (or nodes) that contains the BIOS you want to flash.

Hewlett Packard Enterprise recommends that you power down the target nodes before you flash the BIOS firmware.

b. Use the following command to flash the BIOS firmware:

```
cm node firmware update -n node firmware_file
```

The variables are as follows:

Variable	Specification
node	Specify the same node hostnames you specified when you flashed the iLO firmware.
firmware_file	The file that contains the new BIOS firmware.

For example, the following command shows the format of a BIOS firmware file:

```
# cm node firmware update -n 'r1*' A45_1.20_01_30_20XX.signed.bin
```

c. (Conditional) Reboot the nodes upon which the BIOS firmware was flashed.

Complete this step if you flashed the BIOS firmware when the nodes were powered down.

8. (Conditional) Flash the CPLD firmware.

Complete this step if you want to flash the CPLD firmware at this time.

You can flash the node CPLD firmware at any time without flashing the iLO firmware first.

Enter the following command:

```
cm node firmware update -n node firmware_file
```

The variables are as follows:

Variable	Specification
node	The same node hostnames you specified when you flashed the iLO firmware.
firmware_file	The CPLD firmware file. This file is of the following form: `CPLD_identifier.bin`

Flashing the firmware on an HPE Apollo 9000 chassis controller

The `cm chassis cmc firmware` command flashes the chassis controller firmware and the firmware for the following devices that are associated with a chassis controller:

- The chassis controller itself. Each chassis controller contains the following devices that host firmware:
  - The P1022.
  - The complex programmable logic device (CPLD).
- The tray controllers. The chassis controller manages the following devices on the tray controller, all of which host firmware:

- The K64.
    - The CPLD.
    - The BCM52161.
    - The P1022.

For more information about the `cm chassis cmc firmware` command, enter one or more of the following commands at the system prompt:

- `cm chassis cmc firmware -h`
- `cm chassis cmc firmware show -h`.

  The `cm chassis cmc firmware show` command accepts the `--cmdiff` parameter. The inclusion of `--cmdiff` lets you compare firmware on multiple nodes. By default, the output appears on your screen. You can specify to send the output to a file.
- `cm chassis cmc firmware update -h`

Procedure

1. Contact your HPE representative to obtain the compute node firmware.
2. Log into the admin node as the root user.
3. Write the new firmware to the following location on the admin node:

   ```
   /opt/clmgr/tftpboot
   ```
4. Make sure that all the firmware files are world-readable.

   For example:

   ```
   # cd /opt/clmgr/tftpboot
   # chmod 744 *
   ```
5. Enter one or more commands, in the following form, to flash the firmware in the chassis controller:

   ```
   cm chassis cmc firmware update -n node firmware_file
   ```

   The variables are as follows:

Variable	Specification
node	One or more node hostnames. The command flashes targets in sets of up to 128 targets at a time.

firmware_file	The name of one of the firmware files. There are four firmware files for the chassis controllers. These files are as follows:
	• The firmware file that updates the chassis controller firmware itself, the tray controller, and the switch tray is of the following form:
	`CMC_`*`dates`*`.bin`
	When you flash the chassis controller, the chassis controller restarts.
	• The firmware file that updates the CPLD device in the chassis controller is of the following form:
	`CMC-cpld_`*`numbers`*`.bin`
	The flashing commands return after the command sends the flashing request successfully. Most firmware flashes take 2-3 minutes. The exception to this is the chassis controller CPLD target, which takes 15-20 minutes to complete after the command returns.
	• The firmware file that updates the CPLD device in the tray controller is of the following form:
	`CMC-Brz-cpld_`*`numbers`*`.bin`
	• The firmware file that updates the BCM device in the chassis controller is of the following form:
	`CMC_BCM`*`numbers`*`.bin`

Retrieving system firmware information

Your cluster system comes preinstalled with the appropriate firmware. For information about updates to the node controller, BIOS, and chassis controller firmware, see your HPE representative.

Use the following methods to retrieve firmware information:

- To identify the BIOS, you need both the version and the release date. You can get these using the `dmidecode` command. Log onto the node from which you want information, and enter the following command:

  ```
  # dmidecode -s bios-version; dmidecode -s bios-release-date
  ```

- To retrieve the node controller firmware revision, use the `ipmiwrapper` command. For example, from the admin node, the following command displays the node controller firmware revision for `r1i0n0`:

  ```
  # ipmiwrapper r1i0n0 bmc info | grep 'Firmware Revision'
  ```

- On HPE SGI 8600 systems and other systems that have chassis controllers, you can use the `version` command to retrieve the chassis controller firmware version. For example, from the `r1lead` leader node, the following command displays the chassis controller firmware version:

  ```
  # ssh root@r1i0-cmc version
  ```

- The `ibstat` command retrieves information for the fabric links and includes the firmware version. The following command displays the fabric firmware version:

  ```
  # ibstat | grep Firmware
  ```

- The `firmware_revs` script on the admin node displays firmware information for all nodes.

Flashing the firmware on a power shelf or fan controller on an HPE SGI 8600 cluster

In rare situations, the power shelf firmware or fan controller firmware can become corrupted. In this situation, the power shelf or the fan controller becomes broken or remains perpetually in bootloader mode. If in bootloader mode, the fan controller firmware can respond to the firmware flashing utility.

The following procedure explains how to flash the firmware.

Procedure

1. Log in to the power shelf or the fan controller.

   For a power shelf, log into the lowest CMC in the power domain.

   For information about how to log into a CMC, see the following:

   Power cycling the chassis on an HPE SGI 8600 cluster

2. Enter the following command to change to the directory that contains the firmware images:

   ```
   > cd /usr/local/firmware/psoc
   ```

3. Use the following command to flash the firmware:

   ```
   flashcan -f image -p target -r
   ```

   The variables are as follows:

Variable	Specification
image	The name of one of the firmware images from the following directory: `/usr/local/firmware/psoc`
target	One of the following: • `0`, which specifies power shelf 0. • `1`, which specifies power shelf 1. • `2`, which specifies power shelf 2. • `3`, which specifies power shelf 3. • `4`, which specifies fan controller 0. • `5`, which specifies fan controller 1.

Security features and credentials

The Linux distributions include security features. The cluster manager provides additional security features, and among those features are the following:

- Secure provisioning over the management network
- Restricted node-to-node login access
- No root user `ssh` files in default images
- Secure environment for the cluster manager database

To provide these features, the cluster manager creates encryption passwords, certificates, `ssh` keys, and other security constructs. The cluster manager documentation refers to these items collectively as **secrets**. The secrets that the cluster manager uses to build the node security infrastructure are called **bootstrap secrets**.

For information about how to create security certificates from a site-specific certificate authority (CA), see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Creating secrets

The cluster manager creates security secrets when the cluster manager is installed and configured. The installation step that creates security secrets is the `configure-cluster` step.

Using the newly created secrets, the installation process creates the necessary security framework on the admin node. The installer creates the complementary security infrastructure on the other nodes as they are installed.

The following topics explain how the cluster manager creates secrets:

- **Packaging and file residence**
- **Recreating secrets on ICE leader nodes and on compute nodes**

Packaging and file residence

The cluster manager manages the secrets files. You do not need to manually manage the files.

The secrets files reside in the following directory on the admin node:

```
/opt/sgi/secrets
```

The bootstrap secrets reside in compressed, encrypted files in the following directory:

```
/opt/sgi/secrets/bootstrap-secrets
```

The following command output shows the two secrets files:

```
[root@myadmin ~]# ls /opt/sgi/secrets/bootstrap-secrets
compute.tar.xz.aes  leader.tar.xz.aes
```

File `compute.tar.xz.aes` is the secrets file for ICE compute nodes and compute nodes. File `leader.tar.xz.aes` is the secrets file for leader nodes.

For added security, the encrypted bootstrap secrets can be transferred to a node only if the node is marked for installation.

Recreating secrets on ICE leader nodes and on compute nodes

For added security, you can recreate the set of secrets. Doing so, however, requires that you reinstall the cluster nodes afterwords, except for the admin node. If the cluster includes leader nodes, reinstall the leader nodes, too.

The following procedure explains how to recreate secrets.

Procedure

1. Run the following script:

   ```
   /opt/sgi/lib/create-secrets --force
   ```

2. Enter the following command to restart the `cmdb` service:

   ```
   # systemctl restart cmdb.service
   ```

3. Mark all ICE leader and compute nodes for reinstallation.

   Example 1. On a cluster without ICE leader nodes, run the following command:

   ```
   # cm node provision -n 'n*'
   ```

   Example 2. On an ICE cluster with leader nodes, run the following command:

   ```
   # cm node provision -n 'n*','r*lead'
   ```

   You do not need to mark all the ICE compute nodes for reinstallation. The cluster manager reinstalls them automatically after the leader nodes reboot.

Secure provisioning

By default, the cluster manager uses UDPcast to transfer images from the admin node to the other nodes. In some cases, factors such as node type and performance might cause you to choose another transport method. Secure provisioning is ensured regardless of the transfer method you choose.

The cluster manager uses secure provisioning for all image transfers, regardless of method. During provisioning, the cluster manager authenticates, encrypts, and decrypts the images at the appropriate transfer points. The bootstrap secrets provide the resources for authentication, encryption, and decryption.

To guard against unauthorized requests for images, the cluster manager does a series of checks. The following are some of the checks:

- The cluster manager verifies the provisioning request. If the node has a disk and was not included in a `cm node provision` command, the cluster manager does not send the image.
- The password that decrypts the bootstrap secrets is transferred. As part of this activity, the cluster manager sends the image only to a root-specific path on the node being installed.

For information about how to choose the appropriate transfer method for your site, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Restricted node-to-node login access

The action of logging into a node without a password is called a **passive login**.

The cluster manager imposes security constraints on the root user. The cluster manager constrains the root user because many cluster manager interfaces run under the auspices of the root user, passively logging into nodes.

The cluster manager allows the root user to passively log into any node from the admin node. It permits other passive logins. By default, the root user is allowed to passively log in as follows:

- From the admin node to a leader node
- From the admin node to an ICE compute node
- From the admin node to a compute node
- From a leader node to another leader node
- From a compute node to another compute node

The cluster manager prevents the root user from passively logging into some nodes from other nodes. By default, the root user is not allowed to passively log in as follows:

- From a compute node to a leader node
- From an ICE compute node to a leader node
- From a compute node to the admin node

The following figure Illustrates the login flows allowed for the root user.

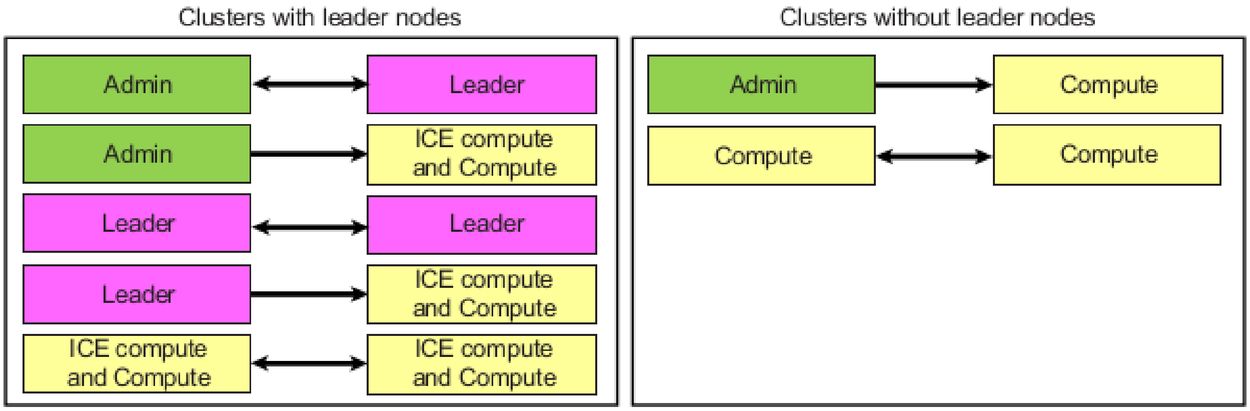


Figure 18: Root user passive logins

The following topics contain more information about login access:

- **Cluster manager `ssh` zones**
- **Default images and customizing your `ssh` configuration**
- **Security recommendations**

Cluster manager `ssh` zones

The cluster manager uses the `ssh` security scheme (RSA 2 private and public keys) to restrict the node-to-node login access for the root user.

The root-user-authorized key files on the admin and leader nodes are configured such that root users from compute nodes or ICE compute nodes cannot passively log in. The root-user-authorized key files for the compute nodes contain the public keys for the admin nodes, leader nodes, and other compute nodes. The root-user-authorized key files do not contain public keys for the root users from the compute nodes.

To document this login flow in the authorized key files, the cluster manager does the following:

- The cluster manager uses the notion of zones.
- The cluster manager labels each public key it places in the file according to the login flow it allows. The cluster manager places the zone label as a comment at the end of the key.

For example, the following command returns the labels and zones for the public keys:

```
r1lead:~/.ssh # cat authorized_keys
```

The following table summarizes the zones that the cluster manager places in the authorized key file of the root user:

Node type	Zones in authorized key file for the root user
Admin	Admin node, leader node
Leader	Admin node, leader node
ICE compute nodes and computes nodes	Admin node, leader node, ICE compute nodes, compute node

With such zones in the authorized key file, you can quickly ascertain the types of hosts that are allowed passive root access.

For information about how the cluster manager restricts node-to-node login access for the root user, see the following:

Restricted node-to-node login access

Default images and customizing your ssh configuration

For added security, the default images on the admin node do not contain ssh keys. When the cluster manager installs a node, it populates the ssh files for the root user of that node.

After a node is installed, you can add your own root user ssh keys and configuration to the image on the admin node. Then, when you install that image, the cluster manager uses your ssh keys and configuration instead of its own.

CAUTION: If you modify the ssh configuration for the root user on the admin or a leader node, do not disrupt the login flows for the root user on those nodes. For more information about logging into nodes, see the following:

Restricted node-to-node login access

The following topic explains how the cluster manager creates the ssh keys (RSA 2) and other bootstrap secrets:

Creating secrets

Security recommendations

To maintain a secure cluster, you must go beyond the protections provided by the ssh zoning feature for the root user. Your site must restrict the running of user code and user jobs on admin and leader nodes. The cluster manager does not prevent such jobs and processes from running. Your site must assess the risks and benefits.

Setting credentials for node controllers on HPE Cray systems

Use the procedure in this topic to set credentials for the following:

- Node controllers
- Switch controllers

Procedure

1. Log into the admin node as the root user.
2. Use the `cm controller set` command to update the username and password for a specific controller.

   The format is as follows:

   ```
   cm controller set -c node -u username -p password [--update-on-controller]
   ```

   The variables and parameters are as follows:

Variable or parameter	Specification
node	The hostname of the node controller.
username	The new username for the controller.
password	The new password for the controller.
`--update-on-controller`	Optional. If unspecified, the cluster manager updates the cluster database with the new username and password. The cluster manager does not update the controller itself. If specified, the cluster manager updates the following: • The cluster database with the new username and password • The controller with the new password It is assumed that the controller itself already has an account with the specified username.

   For example, the following command updates the cluster database and the node controller with a new password, and it updates the cluster database with a new username:

   ```
   # cm controller set -c x9000c1s0b0 -u newusername -p newpassword --update-on-controller
   ```

Setting credentials for iLO, iLOCM, IPMI, and baseboard management controller (BMC) devices attached to scalable unit (SU) leader nodes and compute nodes

This procedure explains how to set credentials for iLO, iLOCM, IPMI, and BMC devices These are the credentials that the cluster manager can use to accesses the SU leader nodes and compute nodes attached to these types of devices.

The cluster manager processes logs into individual nodes in the cluster when performing cluster operations. The processes uses the credentials you store in this procedure for cases in which the credentials of an individual node are not stored.

Use the procedures in this topic to change the following on iLO, iLOCM, IPMI, and BMC devices:

- The administrator username
- The administrator password

Notice that you cannot change just the administrator username.

When creating passwords or usernames, make sure that these new credentials do not include any of the following characters:

- Quotation mark (")
- Apostrophe (')
- Dollar sign ($)
- Ampersand (&)
- Equal sign (=)
- Pound sign (#)
- Tab character
- Space character

Repeat this procedure for each credential you want to change.

Procedure

1. Log into the admin node as the root user, and use the `cm node show` command to display information about the cluster.

   For example, see the **CARDTYPE** column in the following output:

   ```
   # cm node show -B
   NODE      CARDIPADDRESS   CARDMACADDRESS      CARDTYPE   PROTOCOL    USERNAME
   node42    10.117.30.142   00:9c:02:99:1b:3e   iLO        dcmi,ipmi   admin
   node43    10.117.30.143   00:9c:02:a5:29:08   IPMI       dcmi,ipmi   admin
   node44    10.117.30.144   00:9c:02:99:1b:b2   IPMI       dcmi,ipmi   admin
   node45    10.117.30.145   00:9c:02:99:2f:2a   iLO        dcmi,ipmi   admin
   ```

   The `cm node show` command does not return information for nodes that you did not configure into the cluster.

2. Use the `cmu_get_bmc_access` command to display the credentials for one or more nodes.

   The format is as follows:

   ```
   /opt/clmgr/tools/cmu_get_bmc_access -n node cardtype
   ```

   The variables are as follows:

Variable	**Specification**
node	One node hostname.
cardtype	Specify `ILO`, `IPMI`, or `ILOCM`.

For example:

```
# /opt/clmgr/tools/cmu_get_bmc_access -n n0 ILO
admin admin
```

3. Use the `cmu_set_bmc_access` command to initiate a dialog with the cluster regarding the credentials for a specific node having the given card type.

   The format of the command is as follows:

   ```
   /opt/clmgr/tools/cmu_set_bmc_access -n node cardtype
   ```

   For example:

   ```
   # /opt/clmgr/tools/cmu_set_bmc_access -n node1 ILO
   ```

4. Respond to the system prompts for the new login and new password.

   For example:

   ```
   # /opt/clmgr/tools/cmu_set_bmc_access -n node10 ILO
   ILO login> admin
   ILO password>
   Reenter ILO password>
   ```

   The cluster manager does not echo the password to the screen.

5. (Conditional) Repeat the preceding steps to set the credentials on other devices that exist in your cluster.

Setting credentials for an iLO, an iLOCM, or baseboard management controller (BMC)

The procedure in this topic lets you set or update the username, the password, or both the username and password for a BMC. You can use this procedure to change credentials on any node

Procedure

1. **Setting the credentials**
2. (Conditional) Propagate the new credentials.

   Complete one of the following procedures if you updated credentials on a high availability rack leader controller (HA RLC):

   - **Propagating the credentials on a RHEL high availability rack leader controller (HA RLC)**
   - **Propagating the credentials on a SLES high availability rack leader controller (HA RLC)**

Setting the credentials

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node set` command in the following format to set or to change the BMC username, the password, or both the username and password:

   ```
   cm node set credential  [--update-bmc] nodes
   ```

The variables are as follows:

Parameter or Variable	Specification
credential	The credential you want to change. Enter one or both of the following parameters: • `--bmc-password` *`new_password`* For *new_password*, specify the new password. • `--bmc-username` *`new_username`* For *new_username*, specify the new username.
`--update-bmc`	Optional. When specified, the command contacts the BMC and changes the username or password on the BMC. If you do not specify `--update-bmc`, it is your responsibility to change the username or password manually in the BMC.

nodes	The hostname of each node upon which you want to update the BMC credentials. Specify the nodes in one of the following ways:
	• `-n` *`node`*`,`*`node`*`,...`
	Specify one or more node hostnames. If specifying individual hostnames, specify them in a comma-separated list. You can use wildcard characters.
	For information about how to specify hostnames, see the following:
	Using the `cm` commands
	For example:
	`@gpu-nodes`
	`n0`
	`node?`
	`node[13]`
	`node[10-14]`
	`node[001-022]`
	`node[2-6,20-26,36]`
	`node52*`
	`admin`
	• `-f` *`file`*
	For *file*, specify the full path to a file that contains a list of node hostnames.
	• `-t {custom,image,network,system}` *`group_names`*
	Copies only the nodes in the `custom`, `image`, `network`, or `system` group. For *group_names*, specify one group name or a comma-separated list of group names.
	Common system group names include `compute`, `ice_compute`, and `leader`.
	If you specify `-t system ALL` , the command runs on all cluster nodes including the admin node.

Propagating the credentials on a RHEL high availability rack leader controller (HA RLC)

Prerequisites

Setting the credentials

Procedure

1. Use the `ssh` command to log into the HA RLC upon which you changed the credentials as the root user.
2. Enter the `pcs stonith show` command, and note the fence resource IDs in the output:

   ```
   # pcs stonith show
   Warning: This command is deprecated and will be removed. Please use 'pcs stonith status' instead.
     * p_ipmi_fencing_1     (stonith:fence_ipmilan):      Started hikari2-ptp
     * p_ipmi_fencing_2     (stonith:fence_ipmilan):      Started hikari-ptp
   ```

   NOTE: You can safely ignore the Warning message in the output.

3. Enter the following command:

   ```
   pcs stonith update fence_resource_id fence_ipmilan username=username password=password
   ```

   The variables are as follows:

Variable	Specification
fence_resource_id	Specify the fence resource ID for the HA RLC node that had its credentials changed.
username	The username used to access the BMC on the RLC.
password	The new password used to access the BMC on the RLC.

   For example:

   ```
   # pcs stonith update p_ipmi_fencing_1 fence_ipmilan username=admin password=newadmin
   ```

4. Repeat Step **3** for the other RLC.

   For example:

   ```
   # pcs stonith update p_ipmi_fencing_2 fence_ipmilan username=admin password=newadmin
   ```

Propagating the credentials on a SLES high availability rack leader controller (HA RLC)

Prerequisites

Setting the credentials

Procedure

1. Use the `ssh` command to log into the HA RLC upon which you changed the credentials as the root user.
2. Enter the `crm configure show` command, and note the fence resource IDs in the output.

For example:

```
# crm configure show
….
primitive p_ipmi_fencing_1 stonith:external/ipmi \
              meta target-role=started is-managed=true \
              operations $id=stonith-ipmi-1-operations \
              op monitor interval=15 timeout=20 on-fail=restart start-delay=15 \
              op start interval=0 timeout=15 on-fail=restart \
              params hostname=hikari ipaddr=150.166.33.247 userid=ADMIN passwd="******" interface=lanplus pcmk_delay_max=30
primitive p_ipmi_fencing_2 stonith:external/ipmi \
              meta target-role=started is-managed=true \
              operations $id=stonith-ipmi-2-operations \
              op monitor interval=15 timeout=20 on-fail=restart start-delay=15 \
              op start interval=0 timeout=15 on-fail=restart \
              params hostname=hikari2 ipaddr=150.166.33.248 userid=ADMIN passwd="******" interface=lanplus
```

An HA RLC configuration consists of two HA leader nodes. The output shows `p_ipmi_fencing_1` as the fence resource ID for `hikari` and `p_ipmi_fencing_2` as the fence resource ID for `hikari2`.

3. Write the HA configuration to a temporary file.

   For example:

   ```
   # crm config show > /tmp/stonith.config
   ```

4. Use a text editor to open the temporary file, search for the `passwd=` keyword in the file, edit the file to update the credentials, save the file, and close the file.

   For example:

   ```
   # vim /tmp/stonith.config
   ```

5. Enter the following command to update `stonith.config`:

   ```
   # crm config load update /tmp/stonith.config
   ```

Cluster manager database security

The primary data repository for the cluster manager resides in a relational database on the admin node. Many cluster management functions read from and write to this database.

For increased security, the cluster manager provides the following database safeguards:

- The cluster manager limits direct access to the root user only. The root user can read and write to the cluster database.
- The cluster manager encrypts the requests and data used by the cluster daemons to access the database.
- The cluster manager does not replicate the database to other servers.

  Secure access to the database is available through the `https` interface over the cluster management network.
- GUI uses Java remote method invocation (RMI). RMI communication is protected with transport layer security (TLS).
- The database is serverless. There are no database-external network interfaces except for `localhost`. By default, the RESTful API is accessible externally (from outside the cluster). The RESTful API that exposes the database content is password protected or certificate protected.
- A username and either a password or a certificate are used to identify the user. These credentials determine the permission level of the user (read-only, read/write, admin, and so on).

  For more information about permissions, see the following file:

  `/opt/clmgr/etc/admins`
- Communication is encrypted using HTTPS (TLS).

For more information, start the cluster manager GUI, and on the landing page, click the **REST API Documentation** link in the **Resources** pane.

Disabling deprecated versions of transport layer security (TLS)

The cluster manager enables the current version of TLS by default for the cluster database. The procedure in this topic explains how to disable older versions, for example of TLSv1.2 and TLSv1.1, on a cluster-wide basis.

Procedure

1. Log into the admin node as the root user.
2. Open the Java security file in a text editor.

   The security file is often in the following location:

   ```
   $JAVA_HOME/jre/lib/security/java.security
   ```
3. Search for the following line:

   ```
   jdk.tls.disabledAlgorithms
   ```
4. Edit the line to look as follows:

   ```
   jdk.tls.disabledAlgorithms=TLSv1.2,TLSv1.1
   ```
5. Save and close the file.
6. Enter the following command to restart the `cmdb` service:

   ```
   # systemctl restart cmdb.service
   ```

Using the cluster manager with products from independent software vendors (ISVs) and the open-source community

The cluster manager interoperates with products from several ISVs and with software from the open-source community. The following topics explain how to use these products with the cluster manager. For products that include a GUI interface, the topics in this chapter explain how to enable the product GUI from within the cluster manager GUI.

Retrieving information about third-party software products

The cluster manager includes third-party software products. The following procedure explains how to display information about these products.

Procedure

1. Log into the admin node as the root user.
2. (Conditional) Log into the node that hosts the software for which you need information.

   For example, log into one of the compute nodes or log into the node dedicated to this software.
3. (Optional) Enter the following command to list the packages installed on the node:

   ```
   # rpm -qa
   ```
4. Enter the `rpm` command in the following format:

   ```
   rpm -q package
   ```

   For *package*, specify the package name.

Enabling Altair PBS Works

The Altair PBS Works connector provides the following:

- New cluster manager menu options that are specific to Altair PBS Works.
- Monitoring metrics related to Altair PBS Works that are available to cluster manager administrators.

Procedure

1. Access the following link and use the included instructions to install the connector software on appropriate nodes:

   https://www.pbsworks.com/hpcm-pbspro-connector/
2. Download the connector software from the following location:

   https://github.com/PBSPro/hpcm-pbspro-connector/releases

Enabling Ansible

You can configure the cluster manager to supply dynamic inventory information to Ansible, an IT automation tool, in the form of an Ansible-format `hosts` file. As you add, delete, or modify nodes and groups, the cluster manager automatically updates this inventory information. No manual refreshing is necessary.

Procedure

1. Click **Options** > **Enter Admin mode**, and specify the cluster credentials.
2. Click **Custom Tools** > **Ansible** > **About Ansible Integration**
3. Complete the following steps to enable dynamic inventory file generation:

   a. Open the following file with a text editor:

   ```
   /opt/clmgr/etc/cmuserver.conf
   ```

   b. Search for the `CMU_ANSIBLE_DYNAMIC_INVENTORY` variable, and set it to true.

   For example:

   ```
   CMU_ANSIBLE_DYNAMIC_INVENTORY=true
   ```

   c. Save and close the file.

   d. Restart the `cmdb` service:

   ```
   # systemctl restart cmdb.service
   ```

4. Follow the instructions that appear in the **About Ansible Integration** window.

   The following directory contains sample Ansible playbooks:

   ```
   /opt/clmgr/contrib/ansible
   ```

   You can run the playbooks from the command line or with the sample entries supplied in the following directory:

   ```
   /opt/clmgr/etc/cmu_custom_menu
   ```

Enabling Mellanox Unified Fabric Manager (UFM)

The Mellanox UFM connector combines cluster manager admin node information with information about Mellanox UFM. This connector enables you to view, in one location, information about both the server and the network. This capability reduces operational efforts and reduces the time it takes to resolve problems.

Procedure

1. Select a node to host the connector software.

   You can add the connector software package to the admin node or to a service node.

2. Access the following link, download the connector software, and use the included instructions to install the connector software on the appropriate nodes.

   http://mellanox.com/content/pages.php?pg=support_index

   Only registered Mellanox customers can access the preceding link. The UFM connector to the cluster manager is packaged within UFM.

   After you install the connector, the cluster manager imports the fabric topology from Mellanox UFM.

3. On the top menu, click **Options** > **Enter Admin mode** mode and enter the cluster credentials.
4. In the left pane, expand **Custom Groups**, and examine the custom groups.

   The cluster manager represents fabric connectivity as one or more custom groups. Each group contains the nodes that are connected to a specific leaf switch in the fabric. You can examine the nodes included in each group to identify cabling errors, cluster manager configuration errors, or nonoptimized job distribution.
5. Right-click one of the custom groups, and observe the UFM actions available.

   Verify that these actions are the actions you configured.
6. On the top menu, select **Custom Tools**, and right-click **HPCM UFM Connector**.

   Observe the UFM actions available, and verify that these actions are the actions you configured.

Enabling Slurm

The cluster manager includes Slurm integration scripts that provide custom menu options for controlling the `slurm` daemon.

Procedure

1. Use the instructions in the following file to enable the Slurm connector:

   ```
   /opt/clmgr/contrib/cmu_slurm_connector/README
   ```
2. On the top menu, click **Options** > **Enter Admin mode** and enter the cluster credentials.
3. Use the instructions in the following topic to enable dynamic custom groups for this connector:

   Enabling dynamic custom groups

Enabling dynamic custom groups

A **dynamic custom group** is a named set of nodes that the cluster manager creates for you when a job is running under a job scheduler. The groups appear in the left menu just like a custom group that you define. The cluster manager creates dynamic custom groups when you use certain connectors, but you need to enable the capability first.

If a connector uses dynamic custom groups, the documentation for the connector refers you to this procedure.

Procedure

1. Log into the admin node as the root user.
2. Open the following file:

   ```
   /opt/clmgr/etc/cmuserver.conf
   ```
3. Use one of the following methods to enable dynamic custom groups for the new connector.

   Method 1. Uncomment the line for the connector you want to deploy. Use this method if this connector is the only connector installed on the cluster. Remove the pound character (#) from column 1.

   For example, after editing, the line for Slurm is as follows:

   ```
   CMU_DYNAMIC_UG_INPUT_SCRIPTS=/opt/clmgr/contrib/
   slurm_dynamic_custom_groups
   ```

Method 2. Add a colon character (:) and the new connector string to the existing list of connectors. Use this method to add an additional connector.

For example, if you had the Slurm connector and you want to add the Moab connector, the edited line is as follows:

```
CMU_DYNAMIC_UG_INPUT_SCRIPTS=/opt/clmgr/contrib/slurm_dynamic_custom_groups:/opt/clmgr/contrib/
pbs_dynamic_custom_groups
```

4. Edit the `CMU_JOBSCHEDULER_NODE` line to include the hostnames of the job scheduler nodes.

   Uncomment the line and add service node hostnames or other node hostnames as needed. To specify multiple schedulers, use a colon (:) as a separator. For example, after editing, the line might look like as follows:

   ```
   CMU_JOBSCHEDULER_NODE=localhost:n0:n1
   ```

5. Make sure that you configured passwordless `ssh` from the admin node to the node that hosts the job scheduler.

6. Verify that `CMU_DYNAMIC_UG_INPUT_SCRIPTS` is available on the node that hosts the job scheduler.

   Typically, this node is the admin node or a service node.

7. Uncomment the lines that include the following strings, and adjust the setting for each line as needed:

   ```
   CMU_DYNAMIC_UG_POLLING_INTERVAL ...
   CMU_DYNAMIC_UG_MIN_JOB_DURATION ...
   CMU_DYNAMIC_UG_JOB_PREFIX ...
   ```

   For example, if you decide to retain the default settings, after editing, the lines are as follows:

   ```
   CMU_DYNAMIC_UG_POLLING_INTERVAL=60
   CMU_DYNAMIC_UG_MIN_JOB_DURATION=120
   CMU_DYNAMIC_UG_JOB_PREFIX="job_"
   ```

8. Save and close the file.

9. Enter the following command to restart the `cmdb` service:

   ```
   admin # systemctl restart cmdb.service
   ```

10. Test the connector.

    For example, use the cluster manager GUI to run a job and perform Slurm functions from within the GUI:

    a. Use the Slurm command-line interface to start a job on the cluster, note the job number, and make sure that the job is running.

       For example:

       ```
       admin # sbatch submit.sh
       Submitted batch job 75
       admin # squeue
            JOBID PARTITION      NAME      USER ST      TIME  NODES
       NODELIST(REASON)
               75      HPCM      test      root R       0:19      2 n[0-1]
       ```

    b. In the cluster manager GUI, in the left pane, expand **Custom Groups**.

       Observe that the cluster manager has created a dynamic group. The dynamic group includes the nodes upon which job 75 is running. The dynamic group is named for the job owner and the job number. In this case, the dynamic group is named `job_root_75`.

    c. In the right pane, right-click anywhere in the **Time View** to expose the Slurm menu entry and the features you can access through the cluster manager GUI.

       The **Time View** lets you observe how the job is running on each node.

d. After the job finishes, in the left pane, expand **Archived Custom Groups**.

   After the job finishes, the cluster manager retains information about this job in the archived custom groups list. The job name in the **Archived Custom Group** list is the same as the job name when it was a dynamic group. For more information about archived custom groups, see the following:

   Limitations for displaying archived custom groups from the GUI

e. (Optional) Use the following command to retrieve job start and job ending times:

```
cmu_show_archived_custom_groups -H -c -f
```

Using Singularity containers

A Singularity container is a portable and self-contained computing environment.

The following topics explain how to create and run Singularity containers.

Procedure

1. **Installing the container software**
2. **Building a container**
3. **Running a container**

Installing the container software

Procedure

1. Log into the admin node as the root user.
2. Obtain the Singularity RPM.

   For example:

   - Navigate to the following website and browse to the required RPM:

     https://github.com/sylabs/singularity/

   - Download the source file and build your RPM. For example, enter the following command:

     ```
     # wget https://github.com/sylabs/singularity/releases/download/v3.10.2/singularity-ce-3.10.2-1.el8.x86_64.rpm
     ```

3. Create a directory to store the RPM in order to create a repository.

   For example, enter the following commands:

   ```
   admin:~# mkdir /opt/clmgr/repos/other/singularity-ce
   admin:~# mv singularity-ce-3.10.2-1.el8.x86_64.rpm /opt/clmgr/repos/other/singularity-ce
   ```

4. Create a custom repository.

   For example:

   ```
   admin:~# cm repo add /opt/clmgr/repos/other/singularity-ce --custom singularity-ce
   Creating rpm-md metadata for repo...
   Creating repodata cache for /opt/clmgr/repos/other/singularity-ce
   Directory walk started
   ```

```
Directory walk done - 1 packages
Temporary output repo path: /opt/clmgr/repos/other/singularity-ce/.repodata/
Preparing sqlite DBs
Pool started (with 5 workers)
Pool finished
Exporting repository for use with yume....
Exporting /opt/clmgr/repos/other/singularity-ce through httpd, http://admin/repo/opt/clmgr/repos/other/singularity-ce
Done adding custom repository.  Note: Remember to add packages from
this source to your custom rpm lists as needed.
```

5. Create a repository group.

   For example, the following command creates a repository group called `singularity`:

```
admin:~# cm repo group show singularity
Group: singularity
  Cluster-Manager-1.7-rhel86-x86_64
  Red-Hat-Enterprise-Linux-8.6.0-x86_64
  singularity-ce
```

6. Add the Singularity RPM to an image.

   For example:

```
admin:~# cm image yum -i singularity-image --repo-group singularity \
install singularity-ce
Repo group singularity specified, using repos: Cluster-Manager-1.7-rhel86-x86_64 singularity-ce Red-Hat-Enterpri
Starting yume, rpm manager yum,  within chroot for image singularity-image ...
Command:  chroot /opt/clmgr/image/images/singularity-image yume --rpmmgr yum
    --repo http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.7-rhel86-x86_64
    --repo http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.7-rhel86-x86_64/RPMS
    --repo http://admin/repo/opt/clmgr/repos/cm/Cluster-Manager-1.7-rhel86-x86_64/SRPMS
    --repo http://admin/repo/opt/clmgr/repos/other/singularity-ce
    --repo http://admin/repo/opt/clmgr/repos/distro/rhel8.6.0-x86_64/AppStream
    --repo http://admin/repo/opt/clmgr/repos/distro/rhel8.6.0-x86_64/BaseOS --noplugins -y install singularit
Cluster-Manager-1.7-rhel86-x86_64_1 package repository                                              2.9 MB/s | 3
RPMS_2 package repository                                                                           3.7 MB/s | 3
SRPMS_3 package repository                                                                          1.7 MB/s | 3
singularity-ce_4 package repository                                                                 2.7 MB/s | 3
AppStream_5 package repository                                                                      3.1 MB/s | 3
BaseOS_6 package repository                                                                         2.4 MB/s | 2
Dependencies resolved.
==================================================================================================================
 Package                  Architecture    Version                                   Repository
==================================================================================================================
Installing:
 singularity-ce           x86_64          3.10.2-1.el8                              other_singularity-ce_4
Installing dependencies:
 runc                     x86_64          1:1.0.3-2.module+el8.6.0+14673+621cb8be   rhel8.6.0-x86_64_AppStream_

Transaction Summary
==================================================================================================================
Install  2 Packages

Total download size: 40 M
Installed size: 122 M
Downloading Packages:
(1/2): runc-1.0.3-2.module+el8.6.0+14673+621cb8be.x86_64.rpm                                        21 MB/s | 3.
(2/2): singularity-ce-3.10.2-1.el8.x86_64.rpm                                                       56 MB/s |  3
------------------------------------------------------------------------------------------------------------------
Total                                                                                               60 MB/s |  4
Running transaction check
Transaction check succeeded.
Running transaction test
Transaction test succeeded.
Running transaction
  Preparing        :
  Installing       : runc-1:1.0.3-2.module+el8.6.0+14673+621cb8be.x86_64
  Installing       : singularity-ce-3.10.2-1.el8.x86_64
  Running scriptlet: singularity-ce-3.10.2-1.el8.x86_64
  Verifying        : singularity-ce-3.10.2-1.el8.x86_64
  Verifying        : runc-1:1.0.3-2.module+el8.6.0+14673+621cb8be.x86_64

Installed:
  runc-1:1.0.3-2.module+el8.6.0+14673+621cb8be.x86_64 singularity-ce-3.10.2-1.el8.x86_64

Complete!
```

NOTE: The preceding output was truncated from the right for inclusion in this documentation.

7. Assign the new image to a node.

   For example:

   ```
   admin:~# cm node provision -n cn01 -i singularity-image

   Assigning image "singularity-image" and kernel "4.18.0-372.9.1.el8.x86_64" to the nodes...

   Configuration manager initiating node configuration.
   1 of 1 nodes completed in 2.6 seconds, averaging 0.7s per node
   1 of 1 nodes completed in 2.6 seconds, averaging 0.7s per node
   Node configuration complete.

   Checking node power status...
   Halting the nodes that are not down...



   direct node cn01 has been issued a halt command

   Waiting 15 seconds for nodes to halt...

   Checking node power status...

   Setting non-autoinstall nodes to provision on their next boot...



   Checking node power status...
   Issuing node reset to non-autoinstall nodes that are "On"...

   direct node cn01 has been issued a reset command
   ```

Building a container

Prerequisites

Installing the container software

Procedure

1. Log into the node that hosts the image with the Singularity RPM.

   For example:

   ```
   admin:~# ssh cn01
   ```

2. Create a container recipe file.

   The following is an example container recipe file:

   ```
   [root@cn01 ~]# cat rhel.recipe
   Bootstrap: yum
   OSVersion: 8
   MirrorURL: http://admin/repo/opt/clmgr/repos/distro/rhel8.6.0-x86_64/BaseOS
   Include: yum
   %setup
   rm ${SINGULARITY_ROOTFS}/etc/yum.repos.d/*
   cp smc-yum.repo ${SINGULARITY_ROOTFS}/etc/yum.repos.d
   ```

```
%post
yum -y install vim gcc
%environment
LANG=en_US.UTF-8
```

The following information applies to the container recipe file in this step:

- This file is for the container recipe. For example, you can name this file `distroname.recipe`.

  You can copy the text from the example repository file in this step and modify it to reflect your environment. For example, change *MirrorURL* to your distribution name.

- In the `%setup` section, there are two lines. The first line deletes the repositories from the container. The second line copies the custom repository file.
- The `%post` section installs the `vim` command and the GNU Compiler Collection.
- The `%environment` section sets the system locale to the USA.

3. Create a repository file.

   The following is an example repository file:

   ```
   [root@cn01 ~]# cat rhel.repo
   [main]
   cachedir=/var/cache/yum/$basearch/$releasever
   keepcache=0
   logfile=/var/log/yum.log
   exactarch=1
   obsoletes=1
   gpgcheck=0
   plugins=0
   distroverpkg=redhat-release
   [Base]
   name=RHEL-Base
   baseurl=http://admin/repo/opt/clmgr/repos/distro/rhel8.6.0-x86_64/BaseOS
   enable=1
   gpgcheck=0
   [AppStream]
   name=RHEL-AppStream
   baseurl=http://admin/repo/opt/clmgr/repos/distro/rhel8.6.0-x86_64/AppStream
   enabled=1
   gpgcheck=0
   ```

4. Build the container.

   For example:

   ```
   [root@cn01 ~]# singularity build rhel-container rhel.recipe
   INFO:    Starting build...
   INFO:    Skipping GPG Key Import
   warning: Generating 18 missing index(es), please wait...
   INFO:    Adding owner write permission to build path: /tmp/build-temp-3909833592/rootfs
   INFO:    Running setup scriptlet
   + mkdir -p /tmp/build-temp-3909833592/rootfs/etc/yum.repos.d/
   + cp rhel.repo /tmp/build-temp-3909833592/rootfs/etc/yum.repos.d/
   INFO:    Running post scriptlet
   + yum -y install vim gcc
   RHEL-Base                                                                                                      77 MB/s |
   RHEL-AppStream                                                                                                 85 MB/s |
   Dependencies resolved.
   =========================================================================================================================
    Package                               Architecture              Version                                    Repository
   =========================================================================================================================
   Installing:
   ```

```
 gcc                                   x86_64                    8.5.0-10.el8                              AppStream
 vim-enhanced                          x86_64                    2:8.0.1763-16.el8_5.13                    AppStream
Installing dependencies:
 binutils                              x86_64                    2.30-113.el8                              Base
 cpp                                   x86_64                    8.5.0-10.el8                              AppStream
 glibc-devel                           x86_64                    2.28-189.1.el8                            Base
 glibc-headers                         x86_64                    2.28-189.1.el8                            Base
 gpm-libs                              x86_64                    1.20.7-17.el8                             AppStream
 isl                                   x86_64                    0.16.1-6.el8                              AppStream
 kernel-headers                        x86_64                    4.18.0-372.9.1.el8                        Base
 libmpc                                x86_64                    1.1.0-9.1.el8                             AppStream
 libpkgconf                            x86_64                    1.4.2-1.el8                               Base
 libxcrypt-devel                       x86_64                    4.1.1-6.el8                               Base
 pkgconf                               x86_64                    1.4.2-1.el8                               Base
 pkgconf-m4                            noarch                    1.4.2-1.el8                               Base
 pkgconf-pkg-config                    x86_64                    1.4.2-1.el8                               Base
 vim-common                            x86_64                    2:8.0.1763-16.el8_5.13                    AppStream
 vim-filesystem                        noarch                    2:8.0.1763-16.el8_5.13                    AppStream

Transaction Summary
=========================================================================================================================
Install  17 Packages

Total download size: 58 M
Installed size: 152 M
Downloading Packages:
(1/17): glibc-devel-2.28-189.1.el8.x86_64.rpm                                                                 4.3 MB/s |
(2/17): glibc-headers-2.28-189.1.el8.x86_64.rpm                                                               9.2 MB/s |
(3/17): libpkgconf-1.4.2-1.el8.x86_64.rpm                                                                     6.5 MB/s |
(4/17): libxcrypt-devel-4.1.1-6.el8.x86_64.rpm                                                                503 kB/s |
(5/17): pkgconf-1.4.2-1.el8.x86_64.rpm                                                                        1.1 MB/s |
(6/17): pkgconf-m4-1.4.2-1.el8.noarch.rpm                                                                     2.0 MB/s |
(7/17): pkgconf-pkg-config-1.4.2-1.el8.x86_64.rpm                                                             785 kB/s |
(8/17): binutils-2.30-113.el8.x86_64.rpm                                                                       29 MB/s |
(9/17): kernel-headers-4.18.0-372.9.1.el8.x86_64.rpm                                                           20 MB/s |
(10/17): cpp-8.5.0-10.el8.x86_64.rpm                                                                           29 MB/s |
(11/17): gpm-libs-1.20.7-17.el8.x86_64.rpm                                                                    886 kB/s |
(12/17): libmpc-1.1.0-9.1.el8.x86_64.rpm                                                                      4.4 MB/s |
(13/17): isl-0.16.1-6.el8.x86_64.rpm                                                                           33 MB/s |
(14/17): vim-enhanced-8.0.1763-16.el8_5.13.x86_64.rpm                                                          15 MB/s |
(15/17): vim-filesystem-8.0.1763-16.el8_5.13.noarch.rpm                                                       2.7 MB/s |
(16/17): vim-common-8.0.1763-16.el8_5.13.x86_64.rpm                                                            42 MB/s |
(17/17): gcc-8.5.0-10.el8.x86_64.rpm                                                                           37 MB/s |
-------------------------------------------------------------------------------------------------------------------------
Total                                                                                                          70 MB/s |
Running transaction check
Transaction check succeeded.
Running transaction test
Transaction test succeeded.
Running transaction
  Preparing        :
  Installing       : libmpc-1.1.0-9.1.el8.x86_64
  Installing       : cpp-8.5.0-10.el8.x86_64
  Running scriptlet: cpp-8.5.0-10.el8.x86_64
  Installing       : vim-filesystem-2:8.0.1763-16.el8_5.13.noarch
  Installing       : vim-common-2:8.0.1763-16.el8_5.13.x86_64
  Installing       : isl-0.16.1-6.el8.x86_64
  Running scriptlet: isl-0.16.1-6.el8.x86_64
  Installing       : gpm-libs-1.20.7-17.el8.x86_64
  Running scriptlet: gpm-libs-1.20.7-17.el8.x86_64
  Installing       : pkgconf-m4-1.4.2-1.el8.noarch
  Installing       : libpkgconf-1.4.2-1.el8.x86_64
  Installing       : pkgconf-1.4.2-1.el8.x86_64
  Installing       : pkgconf-pkg-config-1.4.2-1.el8.x86_64
  Installing       : kernel-headers-4.18.0-372.9.1.el8.x86_64
  Running scriptlet: glibc-headers-2.28-189.1.el8.x86_64
  Installing       : glibc-headers-2.28-189.1.el8.x86_64
  Installing       : libxcrypt-devel-4.1.1-6.el8.x86_64
  Installing       : glibc-devel-2.28-189.1.el8.x86_64
  Running scriptlet: glibc-devel-2.28-189.1.el8.x86_64
  Installing       : binutils-2.30-113.el8.x86_64
  Running scriptlet: binutils-2.30-113.el8.x86_64
  Installing       : gcc-8.5.0-10.el8.x86_64
  Running scriptlet: gcc-8.5.0-10.el8.x86_64
  Installing       : vim-enhanced-2:8.0.1763-16.el8_5.13.x86_64
  Running scriptlet: vim-enhanced-2:8.0.1763-16.el8_5.13.x86_64
  Running scriptlet: vim-common-2:8.0.1763-16.el8_5.13.x86_64
  Verifying        : binutils-2.30-113.el8.x86_64
  Verifying        : glibc-devel-2.28-189.1.el8.x86_64
  Verifying        : glibc-headers-2.28-189.1.el8.x86_64
  Verifying        : kernel-headers-4.18.0-372.9.1.el8.x86_64
  Verifying        : libpkgconf-1.4.2-1.el8.x86_64
  Verifying        : libxcrypt-devel-4.1.1-6.el8.x86_64
  Verifying        : pkgconf-1.4.2-1.el8.x86_64
  Verifying        : pkgconf-m4-1.4.2-1.el8.noarch
  Verifying        : pkgconf-pkg-config-1.4.2-1.el8.x86_64
  Verifying        : cpp-8.5.0-10.el8.x86_64
  Verifying        : gcc-8.5.0-10.el8.x86_64
  Verifying        : gpm-libs-1.20.7-17.el8.x86_64
  Verifying        : isl-0.16.1-6.el8.x86_64
  Verifying        : libmpc-1.1.0-9.1.el8.x86_64
  Verifying        : vim-common-2:8.0.1763-16.el8_5.13.x86_64
  Verifying        : vim-enhanced-2:8.0.1763-16.el8_5.13.x86_64
  Verifying        : vim-filesystem-2:8.0.1763-16.el8_5.13.noarch

Installed:
  binutils-2.30-113.el8.x86_64                   cpp-8.5.0-10.el8.x86_64                          gcc-8.5.0-10.el8.x86_64
  glibc-devel-2.28-189.1.el8.x86_64              glibc-headers-2.28-189.1.el8.x86_64              gpm-libs-1.20.7-17.el8.
  isl-0.16.1-6.el8.x86_64                        kernel-headers-4.18.0-372.9.1.el8.x86_64         libmpc-1.1.0-9.1.el8.x8
  libpkgconf-1.4.2-1.el8.x86_64                  libxcrypt-devel-4.1.1-6.el8.x86_64               pkgconf-1.4.2-1.el8.x86
  pkgconf-m4-1.4.2-1.el8.noarch                  pkgconf-pkg-config-1.4.2-1.el8.x86_64            vim-common-2:8.0.1763-1
  vim-enhanced-2:8.0.1763-16.el8_5.13.x86_64     vim-filesystem-2:8.0.1763-16.el8_5.13.noarch

Complete!
INFO:    Adding environment to container
INFO:    Creating SIF file...
INFO:    Build complete: rhel-container
```

NOTE: The preceding output was truncated from the right for inclusion in this documentation.

Running a container

Prerequisites

Building a container

Procedure

1. Create a simple C program.

   For example:

   ```
   [root@cn01 ~]# cat << EOF >> hello_world.c
   #include <stdio.h>

   void main()
   {
     printf ("Hello World!\n");
   }
   EOF
   ```

2. Enter the following command to show that gcc is not available on the host:

   ```
   [root@cn01 ~]# gcc -o /tmp/hello hello_world.c
   -bash: gcc: command not found
   ```

3. Enter the following command to show that there is no binary:

   ```
   [root@cn01 ~]# ls -l /tmp/hello
   ls: cannot access '/tmp/hello': No such file or directory
   ```

4. Enter the following command to run the container to compile the C code:

   ```
   [root@cn01 ~]# singularity exec rhel-container gcc -o /tmp/hello /root/hello_world.c
   ```

5. Enter the following command to run the binary:

   ```
   [root@cn01 ~]# /tmp/hello
   Hello World!
   ```

Configuration manager framework

The configuration manager framework distributes configuration changes to all admin nodes, leader nodes, and compute nodes. The configuration manager consists of the following major components:

- The command-line interface, which is the only way to interact with the configuration manager. For example:

  ```
  # cm node update config --sync -n admin
  Configuration manager initiating node configuration.
  Node configuration complete.
  1 of 1 nodes completed in 1.130 seconds, averaging 0.945s per node.
  ```

- The configuration manager service, `config_manager`, which runs on the admin node. This service is the core of the configuration manager framework. It performs the following actions:
  - Accepting configuration change requests from the command-line interface through the network
  - Loading all the data for the configuration request
  - Sending that data over the network to the configuration client running on the node
- The configuration client service, `config_client`, which runs on admin nodes, leader nodes, and compute nodes. This service generates and applies the requested configuration files to the node upon which it is running.

For more information, enter the following command to display the manpage:

```
$ man cm-node-update-config
```

Preserving custom configuration changes

If you customize the cluster configuration, the customized changes can conflict with the configuration changes applied by the configuration manager. To protect your custom configuration files, edit the following exclude file and add the absolute path to the configuration file you want to preserve:

```
/etc/opt/sgi/conf.d/exclude-update-configs
```

By adding the absolute path of the configuration file you want to preserve to the `exclude-update-configs`, you prevent the configuration manager from updating that file. The exclude file acts as follows:

- To apply the restriction on all nodes, update the exclude file within a node image, and then provision all nodes with that image.
- To apply the restriction to one node, provision that node and then update the exclude file on that one node.

When you reprovision a node, the act of reprovisioning overwrites any changes you made.

Boot configuration framework

You can use the boot configuration framework to adjust the node setting that the cluster manager applies when it creates an image, during the boot process, and through the execution of the cluster-configuration script on the node being configured.

The following files control the boot configuration framework:

- The boot configuration execution script is in the following location:

  `/opt/clmgr/lib/cluster-configuration`

  When the boot configuration execution script runs, it runs the boot configuration scripts located in the following directory:

  `/etc/opt/sgi/conf.d`

- The individual boot configuration scripts are in the following location:

  `/etc/opt/sgi/conf.d`

  The cluster manager installs some scripts as part of the cluster manager software. You can add scripts as needed on individual nodes or within the images prior to provisioning.

  The boot configuration framework makes adjusting the execution of these scripts intuitive.

- The boot configuration exclude file is in the following location:

  `/etc/opt/sgi/conf.d/exclude`

  You can stop a script from running, without purging the script, by adding the name of the scriptlet to be excluded in the exclude file.

Observe the following guidelines when generating a local boot configuration script:

- Be sure the scripts are executable. The cluster manager does not run scripts that are not executable.
- Scripts must be tolerant of files that do not exist. For example, check that a `syslog` configuration file exists before trying to modify it.
- Scripts that end in a distribution name, or a distribution name with a specific distribution version, run only if the node in question is running that distribution.

  Example 1. The script called `/etc/opt/sgi/conf.d/99-foo.sles15spX` runs only when the node is running SLES 15 SPX.

  Example 2. This example shows the order of operations. Assume you have the following scripts:

  - `88-myscript.sles15spX`
  - `88-myscript.sles`
  - `88-myscript`

  These scripts are run as follows:

  - On a SLES 15 SPX system, `88-myscript.sles15spX` runs.
  - On a SLES system that is not sles15spX, `88-myscript.sles` runs.
  - On all other operating system platforms, `88-myscript` runs.

- To make a custom version of a script supplied by HPE, change the script suffix from the distribution name to `.local`. The `.local` suffix indicates that you want the local version to run in place of the one supplied by HPE.

  This naming convention allows you to customize the scripts provided by HPE while preserving the original, default script supplied by HPE. Scripts that end in `.local` have the highest precedence.

  For example, assume that you have the following scripts:

  - 88-myscript.sles
  - 88-myscript.local

  The script named 88-myscript.local runs in all cases, and other 88-myscript.*suffix* scripts never run.

Configuration information specific to ICE compute nodes

The cimage command pushes images to the ICE compute nodes. After pushing the images, the boot configuration framework runs on the leader node in a chroot environment.

In addition to the boot configuration framework, you can use per-host customization for ICE compute nodes, which runs at one of the following times:

- When an admin node pushes an image to the ICE leader nodes
- Upon demand. Use the --customizations-only option to the cimage command.

To alter images for ICE compute nodes, complete the tasks in the following order:

- Alter the image
- Use the cimage command to push the image to the ICE compute nodes

For more information, see the following:

Managing ICE compute node images

systemd presets

The cluster manager uses systemd presets to disable or enable various services by default. The systemd presets are in the default systemd directory within the image, as follows:

/opt/clmgr/image/images/*image*/usr/lib/systemd/system-preset

For example:

```
# ls *sgi*
80-sgi-array.preset  80-sgi-cpuset.preset  80-sgi-rhel8-compute.preset
```

If you want to override a preset, your options are as follows:

- Use a different preset with a higher priority
- Use a post-installation script
- Use a conf.d script

The presets set a balance between keeping needed services available while reducing the memory footprint and the load of the root filesystem. Root load is important for NFS write boot methods that are used.

A typical example is enabling munge, which 80-sgi-rhel8-compute.preset disables.

Updating admin nodes and scalable unit (SU) leader nodes

The admin node and the SU leader nodes host the cluster manager software, products from Hewlett Packard Enterprise, and products from other vendors. Because the SU leader nodes operate in concert, use the method that this chapter describes to update the SU leader nodes in a way that avoids taking down the cluster.

When you use the methods in this chapter, jobs continue to run even though the admin node or an SU leader node is down for the update maintenance period. Jobs can continue to run even if the admin node is powered off for a short time.

For example, from time to time, you might need to update the following:

- Linux distribution patches, updates, or upgrades.

  Use the instructions from the distribution, but use the method that this chapter describes.

  For example, these could be kernel updates from the operating system vendor.

- HPE Performance Cluster Manager.

  Use the instructions in the installation guide for your platform to perform upgrades from one release level to the another release level. For links to the installation guides, see the following:

  Cluster manager documentation

  If you need to apply a cluster manager patch, however, use the method in this chapter and the installation instructions that accompany the patch.

- Hardware.

  To replace a hardware component on an SU leader, use the method in this chapter and the instructions from in the component manufacturer.

- Firmware.

  To flash firmware, use the method in this chapter and the instructions from the firmware vendor.

  For example, to flash firmware from Hewlett Packard Enterprise, use the method in this chapter, but use the instructions in the following chapter:

  Managing firmware

Updating scalable unit (SU) leader nodes

This procedure explains how to update SU leader nodes with new software, firmware, or hardware components. For example, you can use this method to apply new kernels or to install new power supplies. Use the specific update or installation instructions from the provider.

NOTE: Do not use this method to update the CDTB software on SU leader nodes. For information about how to update the CTDB software, see the following:

Updating the cluster trivial database (CTDB) on scalable unit (SU) leader nodes

Procedure

1. Log into the admin node as the root user.

2. Map your cluster.

Make sure you know the SU leader trios that work together, and obtain the hostnames of each SU leader node. If necessary, enter the following command to display the SU leader node hostnames:

```
# cm node show -t role su-leader
leader1
leader2
leader3
leader4
leader5
leader6
```

For example, assume that you have six SU leader nodes, grouped as follows, with hostnames in the format of `leader`*`n`*:

- Trio 1:
  - `leader1`
  - `leader2`
  - `leader3`
- Trio 2:
  - `leader4`
  - `leader5`
  - `leader6`

3. Use the update instructions to apply the update to the SU leader nodes in the following order, which affects only one SU leader node in each trio at a time:

   Apply the update to the following SU leader nodes first:

   - `leader1`
   - `leader4`

   Apply the update to the following SU leader nodes second:

   - `leader2`
   - `leader5`

   Apply the update to the following SU leader nodes third:

   - `leader3`
   - `leader6`

   As this step shows, your goal is to update only one SU leader node from one trio at one time.

Updating the cluster trivial database (CTDB) on scalable unit (SU) leader nodes

The CTDB software is part of the operating system. The cluster manager requires the CTDB software on each SU leader node to be identical. Versions are not necessarily compatible. Use one of the following methods to update the CTDB software:

- Method One ensures that the CTDB software is identical across the cluster by locking the CTDB package so it is never available for an automatic update. This method ensures that the CTDB package remains at the same level that is installed presently. Plan to update the CTDB software during a scheduled downtime.

  Use this method, for example, if you do not want to take any compute nodes down for any length of time.

  Each operating system has a different method to lock a package to the currently installed version. For example:

  - RHEL 8.X systems require the `python3-dnf-plugin-versionlock` package. You might have to install this package to enable version locks.

    For information about how to add version locks and remove version locks, see the `dnf-versionlock` manpage.

  - SLES 15 SPX systems use the `addlock` feature documented on the `zypper` manpage.

  Sometimes, the lack of access to all update packages with cross dependencies can cause an upgrade operation to fail. The CTDB and Samba packages are one such case, where an update to only the Samba packages can introduce dependency errors in CTDB. For more information, see the HPE Performance Cluster Manager release notes.

- Method Two uses the `cm node install` command to update every SU leader node at the same time. Method Two updates all the SU leader nodes with a few commands in one session. Some compute nodes might be inaccessible for a short time while waiting for the SU leader nodes statuses to show `OK`.

The following steps explain how to implement Method Two.

Procedure

1. Log into the admin node as the root user.
2. Move the CTDB update software to the admin node.
3. Use the `cm node install` command to install the package on all SU leader nodes.

   For example, assume that you have SU leader nodes with hostnames `leader1`, `leader2`, `leader3`, `leader4`, `leader5`, and `leader6` organized into a custom group called `leader`.

   On RHEL systems, enter the following command:

   ```
   # cm node dnf -n @leader\* update ctdb
   ```

   On SLES systems, enter the following command:

   ```
   # cm node zypper -n @leader\* update ctdb
   ```

   Depending on the size of the cluster, this command can take up to a minute to run. During this period, the compute nodes are not in contact with their SU leader node.
4. Wait while the cluster manager updates the packages related to CTDB.

5. Manually restart the CTDB on all the nodes at once:

   ```
   # clush -g su-leader systemctl ctdb restart
   ```

6. Verify the status of the CTDB restart from one of the SU leader nodes:

   ```
   # ssh leader1 ctdb status
   Number of nodes:6
   pnn:0 172.23.100.1     OK (THIS NODE)
   pnn:1 172.23.100.2     OK
   pnn:2 172.23.100.3     OK
   pnn:3 172.23.100.4     OK
   pnn:4 172.23.100.5     OK
   pnn:5 172.23.100.6     OK
   .
   .
   .
   ```

   It might take a few minutes for the command to report `OK` status for each SU leader node.

   If necessary, examine the CTDB log file on each SU leader node. The location is as follows:

   ```
   /var/log/log.ctdb
   ```

Cluster manager ports

The topics that follow describe the communication ports that cluster manager uses by default during typical use.

An initial cluster manager installation requires additional specific ports. For information about the required ports for installation, see the installation guide for your platform. For links to the installation guides, see the following:

Cluster manager documentation

Configuring a custom port

Procedure

1. Log into the admin node as the root user.
2. Research the port number you want to use, and make sure that the port number is not in use at this time.

   For example, one method is to enter the following commands:

   ```
   # cm monitoring kafka get port
   ```

   And

   ```
   # cm monitoring elk get port
   ```

   To show all the options for these commands, add `-h` at the end.

   Alternatively, see the information about ports files in the following topics:

   - **General cluster manager ports**
   - **Fabric management node (FMN) ports**
   - **Service infrastructure monitoring (SIM) ports**

   For information about ports used for Elasticsearch, Kafka, and TimescaleDB, see the following:

   HPE Performance Cluster Manager System Monitoring Guide

   NOTE: The cluster manager does not check port numbers. Verify that the port you want to use is not already in use.

3. (Optional) Add the port you want to configure to the following file:

   `/etc/opt/sgi/conf.d/exclude`

4. Use the documentation for your operating system to configure the custom port.

General cluster manager ports

The following files include information about cluster manager ports:

- `/opt/clmgr/etc/cmuserver.conf`. For example, this file includes the setting `CMU_HTTP_PORT=81`.
- `/etc/opt/sgi/conf.d/80-reserve-ports`

The following table provides more information about the ports in the preceding files.

Table 3: Cluster manager ports

Service	TCP/UDP ports
Configuration manager ports	- 1030 TCP control port. Used for communication between the CM CLI and the configuration manager service on the admin node. All traffic on this port is encrypted using TLS. - 1031 TCP event port. Used to receive database update events from the cluster manager back-end service. Traffic on this port is not encrypted. - 1032 TCP client registration port. Used by configuration clients to register themselves with the configuration manager. All traffic on the port is encrypted using TLS.
Mosquitto (MQTT) client connections	1883
Redfish subscription callbacks	1890
Grafana server	3000 Used to access the Grafana web.
Cluster trivial database (CTDB)	4379
MLflow	5000
PostgresSQL	5432
Cluster health check	8082
Power service ports For example, `cm power`, `cpower`, and `mpower` commands.	502, 1041, 8585, 8587, 8654, 8800, 8888 (TCP ports) 1319 (UDP port)
Flamethrower directory	9000 You can specify this port number on the `admin_udpcast_portbase` configuration attribute.
Alerta	9090
Kafka broker port	9092

Table Continued

Service	TCP/UDP ports
Gluster	- 24007 and 24008 Used for server communication. - 49152-49155 One port used for each brick in the 4-brick configuration.
UDPcast global port base	43124 This port number increases by 2 for each available image. You can specify this port number on the `udpcast_portbase` configuration attribute.

For information about ports used for system monitoring, see the following:

HPE Performance Cluster Manager System Monitoring Guide

Fabric management node (FMN) ports

The HPE Slingshot fabric on the fabric management node uses port 80 and port 443.

Service infrastructure monitoring (SIM) ports

The following table shows the ports that SIM and the `cm sim` command use.

Name	Ports
`node-exporter`	9100
`process-exporter`	9256
`postgres-exporter`	9187 Monitors the TimescaleDB component.
`snmp-exporter`	9116
`elasticsearch-exporter`	9200
`mosquitto-exporter`	9234
`logstash-exporter`	9304
`ctdb-exporter`	9727

Table Continued

Name	Ports
gluster-exporter	9713
zookeeper-jmx-exporter	7070
kafka-jmx-exporter	7071
kafka-cnt-jmx-exporter	7072
schema-registry-jmx-exporter	7073
ksql-jmx-exporter	7074
kibana-exporter	5601
aiops-exporter	5101, 5102, 5111, and 5112
alertmanager	9300

Configuring an HPE Superdome Flex server as a cluster node

Prerequisites

Verify that the following prerequisites are met before you configure a HPE Superdome Flex server as a cluster node:

- The HPE Performance Cluster Manager software is installed and working on the cluster.
- On the HPE Superdome Flex server, the rack management controller (RMC) management interface or the eRMC management interface is connected to the cluster manager BMC management network. The eRMC or RMC is also referred to as the *base I/O BMC external Ethernet port*.
- The server is an HPE Superdome Flex server or an HPE Superdome Flex 280 server.
- The operating system that resides on the HPE Superdome Flex server is one of the operating systems that the cluster manager supports. For information about supported operating systems, see the installation guide that pertains to the cluster platform. For links to the installation guides, see the following:

  Cluster manager documentation

Procedure

1. **Creating a cluster definition file for an HPE Superdome Flex server**
2. **Adding an HPE Superdome Flex server to a cluster**
3. **(Conditional) Creating an image for an HPE Superdome Flex server and deploying the image**

Creating a cluster definition file for an HPE Superdome Flex server

A **cluster definition file** is a text file that includes configuration attributes for one or more cluster components. The goal of this procedure is to create a cluster definition file for the HPE Superdome Flex node.

Procedure

1. Log into the cluster admin node as the root user.
2. Open a text file within an editor, and copy the following information into the file:

```
[discover]
internal_name=XXXX,
hostname1=XXXX,
mgmt_bmc_net_name=XXXX,
mgmt_bmc_net_macs="XXXX",
mgmt_net_name=head,
mgmt_net_bonding_master=bond0,
mgmt_net_bonding_mode=active-backup,
mgmt_net_macs="XXXX",
mgmt_net_interfaces="XXXX",
predictable_net_names=yes,
rootfs=disk,
transport=rsync,
disk_bootloader=no,
dhcp_bootfile=grub2,
card_type=SDFlex,
bmc_username=XXXX,
bmc_password=XXXX
console_device=ttyS0,
architecture=x86_64,
conserver_logging=yes,
conserver_ondemand=yes,
tpm_boot=no,
switch_mgmt_network=no,
redundant_mgmt_network=no
image=XXXX
```

NOTE: For readability, the preceding example shows each configuration attribute on its own line. The cluster manager, however, requires the file to consist of one continuous line. The procedure in this topic includes a step to collapse the configuration attributes into that one line.

For example, you can copy the preceding lines to a file called `template.file` because the preceding lines create a template file that you can modify.

In the file, note the following:

- The following configuration attributes have values of `XXXX`:
  - `internal_name=XXXX`
  - `hostname1=XXXX`
  - `mgmt_bmc_net_name=XXXX`
  - `mgmt_bmc_net_macs="XXXX"`
  - `mgmt_net_macs="XXXX"`
  - `mgmt_net_interfaces="XXXX"`
  - `bmc_username=XXXX`
  - `bmc_password=XXXX`
  - `image=XXXX`

  This procedure explains how to define these configuration attributes with values from your HPE Superdome Flex server.
- The `card_type=SDFlex` configuration attribute directs the cluster manager to manage the node as an HPE Superdome Flex node.

3. (Optional) Enter the following command to write the current, existing cluster definition file to an output file of your choice, and examine its contents:

   ```
   discover --show-configfile > outputfile
   ```

   For *outputfile*, specify a name for the output file. It can be any name. For example, `existing.config`.

   Use this file for reference. When you configure the HPE Superdome Flex server as a cluster node, take care not to duplicate any hostnames or other information that is already specified in the cluster definition file.

4. Edit the template file, and add information to the following fields:

Field	Contents
internal_name	Defines the function, or role, of the node. For example, service7. This name must be unique within the cluster. Do not specify a name that already appears in the existing cluster definition file as the internal_name for an existing node.
hostname1	The hostname that you want to assign to the HPE Superdome Flex system. This is the name user can enter when they want to log into the HPE Superdome Flex node. This name can be the same as the name in the internal_name field.
mgmt_bmc_net_name	The name of the cluster manager BMC management network. Typically head-bmc.
bmc_username	The eRMC or the RMC login credentials on the HPE Superdome Flex system. Typically administrator.
bmc_password	The BMC password for the eRMC or the RMC.

Keep this window up, and keep the template file open. Subsequent steps in this procedure require you to provide more information in the template file.

5. Open a window to the eRMC or the RMC on the HPE Superdome Flex system, and log in as the `administrator` user.

6. Complete the following steps to retrieve information from the eRMC or the RMC, and populate the `mgmt_bmc_net_macs` field in the template:

   a. Use the `baseiolist` command to display the hardware MAC address of the `eth1` management interface. For example:

   ```
   eRMC:r001u01c cli> baseiolist
   P000 [r001u01b]: 172.24.0.3 [94:40:c9:d6:09:07]
                    fe80::9640:c9ff:fed6:907/64 Scope:Link
   ```

   b. Use the `show network` command to confirm the hardware MAC address and the hostname of the management interface. For example:

   ```
   eRMC:r001u01c cli> show network
   .
   .
   .
   -- Network Information --
   eth1  Link encap:Ethernet HWaddr 94:40:c9:d6:09:07
         inet addr:172.24.0.3 Bcast:172.24.255.255 Mask:255.255.0.0
         inet6 addr: fe80::9640:c9ff:fed6:907/64 Scope:Link
         UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
         RX packets:170987038 errors:1 dropped:653709 overruns:0 frame:0
         TX packets:103466013 errors:0 dropped:0 overruns:0 carrier:0
   ```

```
          collisions:0 txqueuelen:1000
          RX bytes:1104269157 (1.0 GiB) TX bytes:2118113821 (1.9 GiB)
          Interrupt:33
```

c. Enter the MAC address of `eth1` into the template file as the value for the `mgmt_bmc_net_macs` field.

In the examples shown in this step, the value is `94:40:c9:d6:09:07`.

7. Open a console window to the HPE Superdome Flex server itself (not the eRMC or RMC), and enter the following command to invoke the EFI shell:

```
RMC> console
```

Wait 5-10 minutes for the command to return boot-to-shell information and display the `Shell>` prompt.

8. Complete the following steps to retrieve information from the HPE Superdome Flex server, and populate the `mgmt_net_macs` field in the template:

a. At the `Shell>` prompt, enter the following command to display NIC hostnames and MAC addresses:

```
Shell> ifconfig -l
-----------------------------------------------------------------
name         : eth0
Media State  : Media disconnected
policy       : static
mac addr     : 94:40:C9:D6:09:08
ipv4 address : 0.0.0.0
subnet mask  : 0.0.0.0
default gateway: 0.0.0.0
  Routes (0 entries):
DNS server     :
-----------------------------------------------------------------
name         : eth1
Media State  : Media present
policy       : static
mac addr     : 94:40:C9:D6:09:09
ipv4 address : 0.0.0.0
subnet mask  : 0.0.0.0
default gateway: 0.0.0.0
  Routes (0 entries):
DNS server   :
-----------------------------------------------------------------
name         : eth2
Media State  : Media disconnected
policy       : dhcp
mac addr     : B8:83:03:8D:9F:E8
ipv4 address : 0.0.0.0
subnet mask  : 0.0.0.0
default gateway: 0.0.0.0
  Routes (0 entries):
DNS server   :
```

In the command output, look for a line that reads `Media State: : Media present`. This line identifies the NIC that is physically connected to the cluster management network. In the preceding example, the NIC to choose is `eth1`. The MAC address for `eth1` is `94:40:C9:D6:09:09`.

b. Use the `ifconfig` command to verify the MAC address of the NIC you chose. For example:

```
Shell> ifconfig -l eth1
name : eth1
Media State : Media present
policy : static
mac addr : 94:40:C9:D6:09:09
ipv4 address : 0.0.0.0
subnet mask : 0.0.0.0
default gateway: 0.0.0.0
Routes (0 entries):
DNS server :
```

c. Enter the MAC address of the interface you choose into the `mgmt_net_macs` field of the template file.

The MAC address in this example is `94:40:C9:D6:09:09`.

9. Complete the following steps to determine the hostname of the NIC that corresponds to the MAC address you chose in Step **8** and to specify that hostname in the `mgmt_net_interfaces` field of the template file.

a. The following table shows the hostname to specify based on the hostname that appears at EFI.

NIC hostname as it appears at EFI in Step 8	NIC hostname to specify for HPE Superdome Flex servers	NIC hostname to specify for HPE Superdome Flex 280 servers
`eth0`	`eno1`	`enp1s0`
`eth1`	`en02`	`enp2s0`
`eth2`	`en03`	`enp3s0`
`eth3`	`en04`	`enp4s0`

For example, if you have an HPE Superdome Flex 280 server and you determined that `eth1` is the NIC that is physically connected to the cluster management network, the hostname to specify is `enp2s0`.

b. (Conditional) Verify the NIC hostname.

Complete this step if an operating system is installed on the HPE Superdome Flex server at this time. If there is no operating system installed on the HPE Superdome Flex server, it is not possible to complete this step.

From the operating system prompt on the HPE Superdome Flex server, enter the `ip` command. In the output, look for the MAC address of `enp2s0`.

In the following example, the MAC address matches the MAC address of the NIC you chose in Step **8**:

```
# ip l sh enp2s0
5: enp2s0: <BROADCAST,MULTICAST,SLAVE,UP,LOWER_UP> mtu 1500 qdisc
mq master bond0 state UP mode DEFAULT group default qlen 1000
    link/ether 94:40:c9:d6:09:09 brd ff:ff:ff:ff:ff:ff
```

The operating system-level hostname of the NIC depends on whether the network uses persistent names and whether the system is an HPE Superdome Flex server or an HPE Superdome Flex 280 server. The preceding example shows an HPE Superdome Flex 280 server network with persistent names.

c. Enter the hostname of the NIC into the `mgmt_net_interfaces` field of the template file.

For example, specify `enp2s0` in the `mgmt_net_interfaces` field.

10. Specify an image, or remove the `image=XXXX` line from the template file.

The cluster manager can install an image on a node as part of the configuration process. If there is an image on the cluster that you want to deploy on the HPE Superdome Flex server, edit the `image=XXXX` line in the template file and specify the image name. For example,. `image=rhel8X-hfs`, where `X` is a RHEL version identifier.

If you do not specify an image in the template file, remove the `image=XXXX` line from the template file and plan to complete the following procedure:

(Conditional) Creating an image for an HPE Superdome Flex server and deploying the image

11. Check the cluster definition template file.

The configuration attributes in the template file are those needed to create a cluster definition file for an HPE Superdome Flex node. You might need to add more configuration attributes for your site.

For more information about configuration attributes, see the installation guide that pertains to the cluster platform. For links to the installation guides, see the following:

Cluster manager documentation

12. Remove the line breaks in the cluster definition template file.

This action changes the file so that it appears as one continuous line. For an individual node, the cluster manager requires that the node definition consist of a series of configuration attributes all on one line. Use a comma character (`,`) and a space to separate each individual configuration attribute.

For example:

```
[discover]
internal_name=service0, hostname1=sdflex280, mgmt_bmc_net_name=head-bmc,
mgmt_bmc_net_macs="94:40:c9:d6:09:07",
mgmt_net_name=head, mgmt_net_bonding_master=bond0, mgmt_net_bonding_mode=active-backup,
mgmt_net_macs="94:40:C9:D6:09:09", mgmt_net_interfaces="enp2s0", predictable_net_names=yes, rootfs=disk,
transport=rsync, disk_bootloader=no, dhcp_bootfile=grub2, card_type=SDFlex, bmc_username=administrator,
bmc_password=mypwd, console_device=ttyS0, architecture=x86_64, conserver_logging=yes, conserver_ondemand=yes,
tpm_boot=no, switch_mgmt_network=no, redundant_mgmt_network=no, image=rhel8X-hfs
```

13. Save and close the cluster definition template file.

14. Enter the following command to set the eRMC or RMC management interface in DHCP mode:

```
eRMC:r001u01c cli> set network addressing=dhcp
eRMC:r001u01c cli> reboot rmc
```

As an alternative, you can delay this step until you are ready to configure the HPE Superdome Flex server into the cluster.

Adding an HPE Superdome Flex server to a cluster

Prerequisites

Creating a cluster definition file for an HPE Superdome Flex server

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node add` command in the following format to add the HPE Superdome Flex server as a cluster node:

   ```
   cm node add -c cluster_definition_file
   ```

   For *cluster_definition_file*, specify the name of the cluster definition file you created in the following procedure:

   Creating a cluster definition file for an HPE Superdome Flex server

   For example:

   ```
   $ cm node add -c sdflex280.node
   ```

3. Use the `cm node show` command in the following format to display the information in the cluster manager regarding the node:

   ```
   cm node show -Aj -n hostname
   ```

   For *hostname*, specify the hostname as specified in the `hostname1` field in the cluster definition file.

   For example, note the **"managed": true** line in the output:

   ```
   $ cm node show -Aj -n sdflex280
   [
       {
           "name": "sdflex280",
           "aliases": {},
           "network": {
               "name": null,
               "defaultGateway": "default",
               "nics": [
                 {
                   "name": "enp2s0",
                   "ipAddress": "172.23.0.3",
                   "macAddress": "94:40:c9:d6:09:09",
                   "interfaceName": "sdflex280",
                   "managed": true,
                   "type": "mgmt"
                 },
                 {
                   "name": "bmc0",
                   "ipAddress": "172.24.0.3",
                   "macAddress": "94:40:c9:d6:09:07",
                   "interfaceName": "sdflex280-bmc",
                   "managed": true,
                   "type": "mgmt-bmc"
                 }
               ],
               "ipAddress": "172.23.0.3",
               "macAddress": "94:40:c9:d6:09:09",
               "subnetMask": "255.255.0.0",
               "mgmtServerIp": "default"
           },
           "image": {
             "name": "rhel8X-hfs",
             "kernel": "4.18.0-305.el8.x86_64",
             "cloningBlockDevice": "default",
   ```

```
      "cloningDate": "2022-03-10T12:25:18.138+0000"
},
"platform": {
"name": "generic",
"architecture": "x86_64",
"serialPort": "ttyS0",
"serialPortSpeed": "default",
"vendorsArgs": "default"
},
"management": {
  "cardType": "SDFlex",
  "cardIpAddress": "172.24.0.3",
  "cardMacAddress": "94:40:c9:d6:09:07",
  "protocol": "Apache,NO_IPMI,None,SDFlex,redfish",
  "username": "administrator",
  "password": "<administrator_encrypted_password>"
},
```

(Conditional) Creating an image for an HPE Superdome Flex server and deploying the image

Prerequisites

Adding an HPE Superdome Flex server to a cluster

Procedure

1. Log into the cluster admin node as the root user.
2. Enter the `cm repo show` command.

   For example:

   ```
   # cm repo show
   Cluster-Manager-1.8-rhel8X-x86_64 : /opt/clmgr/repos/cm/Cluster-Manager-1.8-rhel8X-x86_64
   Cluster-Manager-AIOps-1.8-rhel8X-x86_64 : /opt/clmgr/repos/cm/Cluster-Manager-AIOps-1.8-
   rhel8Xx86_
   64
   Red-Hat-Enterprise-Linux-8.X.X-x86_64 : /opt/clmgr/repos/distro/rhel8.X.X-x86_64
   HPE-Foundation-Software-X.X.X-x86_64 : /opt/clmgr/repos/cm/HPE-Foundation-Software-X.X.X-x86_64
   ```

3. Create a repository group to include matching operating system, cluster manager, and HPE Foundation Software repositories.

   For example:

   ```
   # cm repo group add rhel8X-hfs --repos \
   Red-Hat-Enterprise-Linux-8.X.X-x86_64 \
   HPE-Foundation-Software-X.X.X-x86_64 \
   Cluster-Manager-1.8-rhel8X-x86_64 \
   Cluster-Manager-AIOps-1.8-rhel8X-x86_64
   Repo Red-Hat-Enterprise-Linux-8.X.X-x86_64 added to group rhel8X-hfs
   Repo HPE-Foundation-Software-X.X.X-x86_64 added to group rhel8X-hfs
   Repo Cluster-Manager-1.8-rhel8X-x86_64 added to group rhel8X-hfs
   Repo Cluster-Manager-AIOps-1.8-rhel8X-x86_64 added to group rhel8X-hfs
   Removing: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfsadmin.
   rpmlist
   Removing: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs-
   ```

```
ice.rpmlist
Removing: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs-
lead.rpmlist
Removing: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs.rpmlist
Updating: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs.rpmlist
Updating: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs-
ice.rpmlist
Updating: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfslead.
rpmlist
Updating: /opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfsadmin.
rpmlist
```

4. Display the contents of the repository group.

   For example:

```
# cm repo group show rhel8X-hfs
Group: rhel8X-hfs
Cluster-Manager-1.8-rhel8X-x86_64
Cluster-Manager-AIOps-1.8-rhel8X-x86_64
HPE-Foundation-Software-X.X.X-x86_64
Red-Hat-Enterprise-Linux-8.X.X-x86_64
```

5. Create an image using the repository group as the source.

   For example:

```
# cm image create -i rhel8X-hfs --repo-group \
--repo-group rhel8X-hfs --rpmlist --rpmlist \
/opt/clmgr/image/rpmlists/generated/generated-group-rhel8X-hfs.rpmlist
Repo group rhel8X-hfs specified, using repos: Red-Hat-Enterprise-
Linux-8.4.0-x86_64 HPE-Foundation-Software-X.X.X-x86_64 Cluster-
Manager-1.8-rhel8X-x86_64
Cluster-M
anager-AIOps-1.8-rhel8X-x86_64
----- bootstrap image -----
Seeding with bootstrap tarball: rhel8.X-x86_64-bootstrap.tar.gz
Bootstrap
...
```

6. Add the HPE Foundation Software package to the image.

   The command to use depends on your operating system, as follows:

   - For RHEL images, use the following command:

     ```
     cm image yum -i image --repo-group repo_group 'groups install "HPE\ Foundation\
     Software"'
     ```

   - For SLES images, use the following command:

     ```
     cm image zypper -i image --repo-group repo_group 'in -t pattern HPE-
     Foundation'
     ```

   The variables are as follows:

Variable	Specification
image	The name of the image that you created in Step **5**
repo_group	The name of the repository group that you created in Step **3**

For example, enter the following command to add HPE Foundation Software to a RHEL 8.X image:

```
# cm image yum -i rhel8X-hfs --repo-group rhel8X-hfs 'groups install "HPE\ Foundation\ Software"'
...
=================================================================================================
==
Package Architecture Version Repository
Size
=================================================================================================
==
Installing group/module packages:
hpe-auto-config x86_64 1.4-740.0840.220307T0100.a.rhel8Xhpe cm_HPE-Foundation-Software-X.X.X-
x86_64_3
44 k
hpe-dcd x86_64 3.5-5.1 cm_HPE-Foundation-Software-X.X.X-x86_64_3
2.2 M
...
```

7. (Optional) Use the `cm image show` command to display the image characteristics.

   For example:

```
# cm image show -jd -i rhel8X-hfs
[
  {
    "name": "rhel8X-hfs",
    "type": "generic",
    "architecture": "x86_64",
    "baseOsName": "rhel",
    "baseOsVersion": "8",
    "kernels": [
    "0-rescue-11c26020fca44c488d770ceebd9e39e3",
    "4.18.0-305.el8.x86_64"
    ],
    "target": "compute",
    "revisions": "1"
  }
]
```

8. Use the `cm image set` command to specify a crash kernel boot parameter size of 512M for the image.

   For example:

```
# cm image set -i rhel8X-hfs --crashkernel='512M,high'
```

9. Use the `cm image show` command to display the characteristics of the image, including the crash kernel size.

   For example:

```
# cm image show -s -i rhel8X-hfs
custom-partitions = Undefined
crashkernel = crashkernel=512M,high
hard-quota = Undefined
kernel-extra-params =
kernel-distro-params = ro root=dhcp selinux=0 biosdevname=0
numa_balancing=disable
kernel-leader-params =
nfsroot-extra-params = Undefined
quota-timer = Undefined
repo-group = Undefined
soft-quota = Undefined
```

10. (Conditional) Use the `cm image set` command to set the extra kernel parameters that are needed for PXE booting in the HPE Superdome Flex image.

    Complete this step if you plan to PXE boot the HPE Superdome Flex server, rather than to configure the HPE Superdome Flex server to boot from disk after you provision it.

    For example:

    ```
    # cm image set -i rhel8X-hfs --kernel-extra-params \
    'earlyprintk=ttyS0,115200 log_buf_len=8M nmi_watchdog=0 mce=2 \
    uv_nmi.action=kdump bau=0 pci=nobar udev.children-max=32'
    ```

11. Use the `cm image show` command to display the characteristics of the image.

    For example:

    ```
    # cm image show -s -i rhel8X-hfs
    custom-partitions = Undefined
    crashkernel = crashkernel=512M,high
    hard-quota = Undefined
    kernel-extra-params = earlyprintk=ttyS0,115200 log_buf_len=8M
    nmi_watchdog=0 mce=2 uv_nmi.action=kdump bau=0 pci=nobar udev.childrenmax=
    32
    kernel-distro-params = ro root=dhcp selinux=0 biosdevname=0
    numa_balancing=disable
    kernel-leader-params =
    nfsroot-extra-params = Undefined
    quota-timer = Undefined
    repo-group = Undefined
    soft-quota = Undefined
    ```

12. Use the `cm node provision` command in the following format to provision the node with the new image the next time the node boots:

    ```
    cm node provision -s -i image -n hostname
    ```

    The variables are as follows:

Variable	Specification
image	The name of the image that you created in Step **5**
hostname	The hostname of the HPE Superdome Flex server as specified in the cluster definition file.

    For example:

    ```
    # cm node provision -s -i rhel8X-hfs -n sdflex280
    Assigning image "rhel8X-hfs" and kernel "4.18.0-305.el8.x86_64" to the
    nodes...
    ```

```
Configuration manager initiating node configuration.
1 of 1 nodes completed in 1.0 seconds, averaging 0.3s per node
Node configuration complete.
Setting non-autoinstall nodes to provision on their next boot...
```

13. Log into the HPE Superdome Flex server, and boot the server.

There are multiple ways to boot the HPE Superdome Flex server in a way that causes the node to be provisioned with an image. This step shows one method. Complete the following steps:

a. Log into the HPE Superdome Flex server, and invoke the EFI shell:

```
RMC> console
```

Wait 5-10 minutes for the command to return boot-to-shell information and display the **Shell>** prompt.

b. Use the `lanboot select` command to initiate the boot.

For example:

```
Shell> lanboot select -index 2
M 001 PciRoot(0x0)/Pci(0x1C,0x0)/Pci(0x0,0x0)/MAC(9440C9D60908,0x1)/
IPv4(0.0.0.0)
M 002 PciRoot(0x0)/Pci(0x1C,0x5)/Pci(0x0,0x0)/MAC(9440C9D60909,0x1)/
IPv4(0.0.0.0)
M 003 PciRoot(0x4)/Pci(0x0,0x0)/Pci(0x0,0x0)/MAC(B883038D9FE8,0x1)/
IPv4(0.0.0.0)
Select Desired LAN:
```

Notice that the third line of the preceding output, which starts with `M 002`, includes the MAC address that is specified in the `mgmt_net_macs` field in the cluster definition file for this HPE Superdome Flex node.

c. At the **Select Desired LAN:** prompt, to choose **M 002**, type the number **2**, and press Enter.

The system displays the following additional information:

```
Performing a non-directed lanboot...
>>Start PXE over IPv4.
Station IP address is 172.23.0.3
Server IP address is 172.23.0.1
NBP filename is grub2/x86_64-efi/grub-cm-x86_64.efi
NBP filesize is 1086464 Bytes
Downloading NBP file...
NBP file downloaded successfully.
Optional Data: NONE
Cluster Manager GRUB2 Network Boot Environment
GRUB CPU: x86_64 GRUB Platform: efi
PXE Client IP address: 172.23.0.3
PXE Client MAC address: 94:40:c9:d6:09:09
Loading kernel (linuxefi)...
Loading initrd...
Booting...
...
```

14. (Optional) Configure the HPE Superdome Flex node to PXE boot or to boot from its disk by default.

On the HPE Superdome Flex node, you can use the `efibootmgr` operating system command to configure booting.

For example, enter the following command to retrieve information:

```
# efibootmgr
BootCurrent: 0001
Timeout: 2 seconds
BootOrder:
0001,0018,000E,0002,0003,0004,0005,0006,0007,0008,0009,000A,000B,000C,000D
,000F 0010,0011,0012,0013,0014,0015,0017,0000
Boot0000* Enter Setup
Boot0001* UEFI Internal Shell
Boot0002* Virtual CDROM Device 0 LUN 0
Boot0003* Virtual CDROM Device 0 LUN 1
Boot0004* Virtual CDROM Device 0 LUN 2
Boot0005* Virtual CDROM Device 0 LUN 3
Boot0006* Virtual HardDisk Device 0 LUN 0
Boot0007* Virtual HardDisk Device 0 LUN 1
Boot0008* Virtual HardDisk Device 0 LUN 2
Boot0009* Virtual HardDisk Device 0 LUN 3
Boot000A* BASEIO r001u01b (PXEv4) - MAC:9440C9D60908
Boot000B* BASEIO r001u01b (PXEv6) - MAC:9440C9D60908
Boot000C* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60908
Boot000D* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60908
Boot000E* BASEIO r001u01b (PXEv4) - MAC:9440C9D60909
Boot000F* BASEIO r001u01b (PXEv6) - MAC:9440C9D60909
Boot0010* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60909
Boot0011* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60909
Boot0012* PCIe Slot r001u01p0i05 (PXEv4) - MAC:B883038D9FE8
Boot0013* PCIe Slot r001u01p0i05 (PXEv6) - MAC:B883038D9FE8
Boot0014* PCIe Slot r001u01p0i05 (HTTPv4) - MAC:B883038D9FE8
Boot0015* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0016* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0017* Slot 1
Boot0018* SGI Slot Chooser
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

To configure the HPE Superdome Flex server to PXE boot over NIC 9440C9D60909 by default, enter the following command:

```
# efibootmgr \
--bootorder E,18,1,2,3,4,5,6,7,8,9,A,B,C,D,F,10,11,12,13,14,15,17,0
BootCurrent: 0001
Timeout: 2 seconds
BootOrder:
000E,0018,0001,0002,0003,0004,0005,0006,0007,0008,0009,000A,000B,000C,000D
,000F,0010,0011,0012,0013,0014,0015,0017,0000
Boot0000* Enter Setup
Boot0001* UEFI Internal Shell
Boot0002* Virtual CDROM Device 0 LUN 0
Boot0003* Virtual CDROM Device 0 LUN 1
Boot0004* Virtual CDROM Device 0 LUN 2
Boot0005* Virtual CDROM Device 0 LUN 3
Boot0006* Virtual HardDisk Device 0 LUN 0
Boot0007* Virtual HardDisk Device 0 LUN 1
Boot0008* Virtual HardDisk Device 0 LUN 2
Boot0009* Virtual HardDisk Device 0 LUN 3
```

```
Boot000A* BASEIO r001u01b (PXEv4) - MAC:9440C9D60908
Boot000B* BASEIO r001u01b (PXEv6) - MAC:9440C9D60908
Boot000C* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60908
Boot000D* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60908
Boot000E* BASEIO r001u01b (PXEv4) - MAC:9440C9D60909
Boot000F* BASEIO r001u01b (PXEv6) - MAC:9440C9D60909
Boot0010* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60909
Boot0011* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60909
Boot0012* PCIe Slot r001u01p0i05 (PXEv4) - MAC:B883038D9FE8
Boot0013* PCIe Slot r001u01p0i05 (PXEv6) - MAC:B883038D9FE8
Boot0014* PCIe Slot r001u01p0i05 (HTTPv4) - MAC:B883038D9FE8
Boot0015* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0016* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0017* Slot 1
Boot0018* SGI Slot Chooser
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

To configure the HPE Superdome Flex server to boot from disk by default, enter the following command:

```
# efibootmgr \
--bootorder 18,E,1,2,3,4,5,6,7,8,9,A,B,C,D,F,10,11,12,13,14,15,17,0
BootCurrent: 0001
Timeout: 2 seconds
BootOrder:
0018,000E,0001,0002,0003,0004,0005,0006,0007,0008,0009,000A,000B,000C,000D
,000F,0010,0011,0012,0013,0014,0015,0017,0000
Boot0000* Enter Setup
Boot0001* UEFI Internal Shell
Boot0002* Virtual CDROM Device 0 LUN 0
Boot0003* Virtual CDROM Device 0 LUN 1
Boot0004* Virtual CDROM Device 0 LUN 2
Boot0005* Virtual CDROM Device 0 LUN 3
Boot0006* Virtual HardDisk Device 0 LUN 0
Boot0007* Virtual HardDisk Device 0 LUN 1
Boot0008* Virtual HardDisk Device 0 LUN 2
Boot0009* Virtual HardDisk Device 0 LUN 3
Boot000A* BASEIO r001u01b (PXEv4) - MAC:9440C9D60908
Boot000B* BASEIO r001u01b (PXEv6) - MAC:9440C9D60908
Boot000C* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60908
Boot000D* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60908
Boot000E* BASEIO r001u01b (PXEv4) - MAC:9440C9D60909
Boot000F* BASEIO r001u01b (PXEv6) - MAC:9440C9D60909
Boot0010* BASEIO r001u01b (HTTPv4) - MAC:9440C9D60909
Boot0011* BASEIO r001u01b (HTTPv6) - MAC:9440C9D60909
Boot0012* PCIe Slot r001u01p0i05 (PXEv4) - MAC:B883038D9FE8
Boot0013* PCIe Slot r001u01p0i05 (PXEv6) - MAC:B883038D9FE8
Boot0014* PCIe Slot r001u01p0i05 (HTTPv4) - MAC:B883038D9FE8
Boot0015* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0016* PCIe Slot r001u01p0i05 (HTTPv6) - MAC:B883038D9FE8
Boot0017* Slot 1
Boot0018* SGI Slot Chooser
MirroredPercentageAbove4G: 0.00
MirrorMemoryBelow4GB: false
```

Manpages

The cluster manager online manpages reside in the following directories:

- `/opt/clmgr/man`
- `/opt/sgi/share/man`

To retrieve a manpage, peruse the preceding directories for the command that interests you and use the `man`(1) command to display the output.

For example, enter the following command:

```
# man cm
```

The `cm` manpage lists and describes all the `cm` subcommands in the `SEE ALSO` section. The following examples show how to invoke a subcommand:

```
# cm inventory      # display information on field replaceable units (FRUs)
.
.
.
# cm health check   # display system monitoring information
.
.
.
# cm image copy     # display information about how to copy images
.
.
.
```

YaST navigation

The following table shows SLES YaST navigation key sequences.

Key	Action
Tab **Alt** + **Tab** **Esc** + **Tab** **Shift** + **Tab**	Moves you from label to label or from list to list.
Ctrl + **L**	Refreshes the screen.
Enter	Starts a module from a selected category, runs an action, or activates a menu item.
Up arrow	Changes the category. Selects the next category up.
Down arrow	Changes the category. Selects the next category down.
Right arrow	Starts a module from the selected category.
Shift + **right arrow** **Ctrl** + **A**	Scrolls horizontally to the right. Useful in screens if use of the **left arrow** key would otherwise change the active pane or current selection list.
Alt + *letter* **Esc** + *letter*	Selects the label or action that begins with the *letter* you select. Labels and selected fields in the display contain a highlighted *letter*.
Exit	Quits the YaST interface.

Support and other resources

Accessing Hewlett Packard Enterprise Support

- For live assistance, go to the Contact Hewlett Packard Enterprise Worldwide website:

  https://www.hpe.com/info/assistance

- To access documentation and support services, go to the Hewlett Packard Enterprise Support Center website:

  https://www.hpe.com/support/hpesc

Information to collect

- Technical support registration number (if applicable)
- Product name, model or version, and serial number
- Operating system name and version
- Firmware version
- Error messages
- Product-specific reports and logs
- Add-on products or components
- Third-party products or components

Accessing updates

- Some software products provide a mechanism for accessing software updates through the product interface. Review your product documentation to identify the recommended software update method.
- To download product updates:

  Hewlett Packard Enterprise Support Center

  https://www.hpe.com/support/hpesc

  Hewlett Packard Enterprise Support Center: Software downloads

  https://www.hpe.com/support/downloads

  My HPE Software Center

  https://www.hpe.com/software/hpesoftwarecenter

- To subscribe to eNewsletters and alerts:

  https://www.hpe.com/support/e-updates

- To view and update your entitlements, and to link your contracts and warranties with your profile, go to the Hewlett Packard Enterprise Support Center **More Information on Access to Support Materials** page:

  https://www.hpe.com/support/AccessToSupportMaterials

ⓘ **IMPORTANT:** Access to some updates might require product entitlement when accessed through the Hewlett Packard Enterprise Support Center. You must have an HPE Onepass set up with relevant entitlements.

Remote support

Remote support is available with supported devices as part of your warranty or contractual support agreement. It provides intelligent event diagnosis, and automatic, secure submission of hardware event notifications to Hewlett Packard Enterprise, which initiates a fast and accurate resolution based on the service level of your product. Hewlett Packard Enterprise strongly recommends that you register your device for remote support.

If your product includes additional remote support details, use search to locate that information.

HPE Get Connected

https://www.hpe.com/services/getconnected

HPE Pointnext Tech Care

https://www.hpe.com/services/techcare

HPE Complete Care

https://www.hpe.com/services/completecare

Customer self repair

Hewlett Packard Enterprise customer self repair (CSR) programs allow you to repair your product. If a CSR part needs to be replaced, it will be shipped directly to you so that you can install it at your convenience. Some parts do not qualify for CSR. Your Hewlett Packard Enterprise authorized service provider will determine whether a repair can be accomplished by CSR.

For more information about CSR, contact your local service provider.

Warranty information

To view the warranty information for your product, see the links provided below:

HPE ProLiant and IA-32 Servers and Options

https://www.hpe.com/support/ProLiantServers-Warranties

HPE Enterprise and Cloudline Servers

https://www.hpe.com/support/EnterpriseServers-Warranties

HPE Storage Products

https://www.hpe.com/support/Storage-Warranties

HPE Networking Products

https://www.hpe.com/support/Networking-Warranties

Regulatory information

To view the regulatory information for your product, view the *Safety and Compliance Information for Server, Storage, Power, Networking, and Rack Products*, available at the Hewlett Packard Enterprise Support Center:

https://www.hpe.com/support/Safety-Compliance-EnterpriseProducts

Additional regulatory information

Hewlett Packard Enterprise is committed to providing our customers with information about the chemical substances in our products as needed to comply with legal requirements such as REACH (Regulation EC No 1907/2006 of the European Parliament and the Council). A chemical information report for this product can be found at:

https://www.hpe.com/info/reach

For Hewlett Packard Enterprise product environmental and safety information and compliance data, including RoHS and REACH, see:

https://www.hpe.com/info/ecodata

For Hewlett Packard Enterprise environmental information, including company programs, product recycling, and energy efficiency, see:

https://www.hpe.com/info/environment

Documentation feedback

Hewlett Packard Enterprise is committed to providing documentation that meets your needs. To help us improve the documentation, use the **Feedback** button and icons (located at the bottom of an opened document) on the Hewlett Packard Enterprise Support Center portal (**https://www.hpe.com/support/hpesc**) to send any errors, suggestions, or comments. All document information is captured by the process.