Hewlett Packard Enterprise

HPE Performance Cluster Manager Installation Guide for Clusters Without Leader Nodes

Abstract

This publication describes how to install the HPE Performance Cluster Manager 1.10 software on an HPE cluster system that does not contain scalable unit (SU) leader nodes or ICE leader nodes. These clusters are sometimes referred to as flat clusters.

Part Number: P36610-007
Published: October 2023
Edition: 1

Notices

The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Confidential computer software. Valid license from Hewlett Packard Enterprise required for possession, use, or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Links to third-party websites take you outside the Hewlett Packard Enterprise website. Hewlett Packard Enterprise has no control over and is not responsible for information outside the Hewlett Packard Enterprise website.

Acknowledgments

Qualcomm® and the Qualcomm logo are trademarks of Qualcomm Incorporated, registered in the United States and other countries, used with permission.

Red Hat® is a registered trademark of Red Hat, Inc. in the United States and other countries.

SAP®, SAP HANA®, and SAP S/4HANA® are the trademarks or registered trademarks of SAP SE or its affiliates in Germany and in other countries.

sFlow® is a registered trademark of InMon Corp.

TOGAF® is a registered trademark of The Open Group. IT4IT™ is a trademark of The Open Group.

UNIX® is a registered trademark of The Open Group.

VMware®, VMware NSX®, VMware VCenter®, and VMware vSphere® are registered trademarks or trademarks of VMware, Inc. and its subsidiaries in the United States and other jurisdictions.

X/Open® is a registered trademark, and the X device is a trademark of X/Open Company Ltd. in the UK and other countries.

All third-party marks are property of their respective owners.

Product-specific acknowledgments

Altair, PBS, PBS Pro, and PBS Professional are trademarks or registered trademarks of Altair Engineering, Inc.

Cornelis and Omni-Path Express are trademarks of Cornelis Networks.

TimescaleDB is a trademark of Timescale, Inc.

Ubuntu® is a registered trademark of Canonical Ltd.

Revision history

Part number	Publication date	Edition	Summary of changes
P36610-007	October 2023	1	Supports the HPE Performance Cluster Manager 1.10 release.
P36610-006	March 2023	1	Supports the HPE Performance Cluster Manager 1.9 release.
P36610-005	September 2022	1	Supports the HPE Performance Cluster Manager 1.8 release.
P36610-004	April 2022	3	Supports the HPE Performance Cluster Manager 1.7 release. Edition 3 replaces Edition 2. Edition 3 includes miscellaneous corrections.
P36610-004	March 2022	2	Supports the HPE Performance Cluster Manager 1.7 release. Edition 2 replaces Edition 1. Edition 2 includes miscellaneous corrections.
P36610-004	March 2022	1	Supports the HPE Performance Cluster Manager 1.7 release.

Table Continued

Part number	Publication date	Edition	Summary of changes
P36610-003	November 2021	2	Supports the HPE Performance Cluster Manager 1.6 release. Edition 2 replaces Edition 1. Edition 2 includes information about the quorum high-availability software installation and includes other corrections and additions.
P36610-003	September 2021	1	Supports the HPE Performance Cluster Manager 1.6 release.
P36610-002	March 2021	1	Supports the HPE Performance Cluster Manager 1.5 release.
P36610-001	September 2020	1	Original publication. Supports the HPE Performance Cluster Manager 1.4 release.

Contents

Installing HPE Performance Cluster Manager

This manual is written for system administrators, data center administrators, and software developers. The procedures assume that you are familiar with Linux, clusters, and system administration.

HPE installs the operating system software and the cluster manager software on some cluster systems. If HPE installed and configured the operating system and the cluster software, and you want to keep the configuration, use the procedure in the following to attach the cluster to your network:

HPE Performance Cluster Manager Getting Started Guide

After you attach the cluster to your site network, you can return to this manual to reconfigure the cluster or add optional features.

To start a bare-metal installation, proceed to the following:

Installing the operating system and the cluster manager simultaneously on the admin node

HPE Performance Cluster Manager operating system releases supported

The following tables show the releases that the HPE Performance Cluster Manager 1.10 release supports.

Table 1: Operating systems for x86_64 architectures

Node type	Operating systems supported
Admin	RHEL 8.8
	SLES 15 SP5
	Rocky Linux 8.8
Compute	RHEL 9.2
	RHEL 9.1
	RHEL 8.8
	RHEL 8.7
	SLES 15 SP5 and HPE Cray operating system (COS) releases based on SLES 15 SP5
	SLES 15 SP4 and HPE Cray operating system (COS) releases based on SLES 15 SP4
	Rocky Linux 9.2
	Rocky Linux 9.1
	Rocky Linux 8.8
	Rocky Linux 8.7
	Tri-Lab Operating System Stack (TOSS) 4.6
	Tri-Lab Operating System Stack (TOSS) 4.5
	Ubuntu 22.04.*X*

Table 2: Operating systems for Arm (AArch64) architectures

Node types	Operating systems supported
Admin	Not applicable. The admin node is required to have an x86_64 architecture.
Compute	RHEL 9.2
	RHEL 8.8
	SLES 15 SP5 and HPE Cray operating system (COS) releases based on SLES 15 SP5
	Tri-Lab Operating System Stack (TOSS) 4.6

The following notes pertain to operating system support:

- For operating system availability information, see the HPE Support Center page for the HPE Performance Cluster Manager. Click **https://support.hpe.com** and search for **HPCM**. Also see the HPE Performance Cluster Manager release notes.
- In cluster manager documentation, you can assume that feature descriptions for RHEL platforms also pertain to Rocky Linux platforms, TOSS platforms, and Ubuntu platforms unless otherwise noted.
- HPE supports Infiniband networks on RHEL, Rocky Linux, SLES, TOSS, and Ubuntu platforms. HPE does not support the HPE Slingshot interconnect on Ubuntu platforms. For more information, see the cluster manager release notes and the data fabric documentation.
- Cluster manager operating system support for hardware platforms depends on support for the hardware in the operating system. For example, HPE does not support Ubuntu 22.04.*X* on Gen11 platforms. For more information, see the following:

  https://techlibrary.hpe.com/us/en/enterprise/servers/supportmatrix/).
- Within one cluster, compute nodes can be of a single architecture type or can be a mix of x86_64 and Arm (AArch64) architectures.

Additional RHEL 9 and RHEL 8 requirements

The following information pertains to RHEL 9 and RHEL 8 support:

- RHEL 9 compute nodes include Linux kernel 5.14, which provides support for the Intel 64-bit (x86-64-v2) and AMD architectures. For more information, see the following:

  https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/9/html-single/9.0_release_notes/index
- Before you install RHEL 8 on any node, check the following website, and make sure that the cluster includes only supported hardware:

  https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/8/html/considerations_in_adopting_rhel_8/hardware-enablement_considerations-in-adopting-rhel-8
- If your cluster includes a high availability admin node, also note the supported SAS cards at the following website:

  https://access.redhat.com/solutions/4444321

Cluster manager requirements

The HPE Performance Cluster Manager requires you to observe the following requirements:

- Make sure that the physical admin node or nodes you want to include in the cluster are approved for use with the HPE Performance Cluster Manager. For more information, contact your sales representative.
- The cluster manager requires that the physical quorum high availability (HA) nodes and the physical system admin controller high availability (SAC HA) nodes be available to the cluster manager software for exclusive cluster manager use. Do not attempt to use these nodes for storage servers or for any other purpose.
- Quorum HA admin nodes use a Gluster file system. Hewlett Packard Enterprise does not test or plan for additional loads on these Gluster servers. Overloading a Gluster server can result in a slower file system. A slower file system can affect compute nodes, monitoring, and troubleshooting, all of which expect quick responses to file system access. In extreme situations, an overloaded Gluster file system can cause failovers. Cluster manager use cases include system management and programming environment deployments.
- The cluster manager supports RHEL 8.8, SLES 15 SP5, and Rocky Linux 8.8 on quorum HA admin nodes.

Cluster manager documentation

The following list shows the HPE Performance Cluster Manager documentation:

- The **HPE Performance Cluster Manager release notes** contain feature information, platform requirements, and other release-specific guidance.

  To access the release notes and other product information online, complete the following steps:

  1. Navigate to the following website:

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

  2. Click **Additional Resources** > **HPE Support Center**.
  3. In the search bar, enter **`HPE Performance Cluster Manager`**.

     The list of search results on the right-hand side of the display includes a link to the cluster manager release page.
  4. Click the cluster manager release page.
  5. On the release page, click the link to the release notes.

  To access this information on the product media, navigate to the release notes text file in the following directory:

  `/docs`

  Hewlett Packard Enterprise strongly recommends that you read the release notes, particularly the *Known Issues and 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 consumption management features included in the cluster manager:

  HPE Performance Cluster Manager Power Consumption 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`

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

References to operating system releases often include an X to represent parts of release identification information or service pack numbers. When following examples, replace the X shown in commands with the appropriate release-specific identification information or service pack information.

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`

Table Continued

Specification	Nodes affected
`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*BMC*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`.
BMC	A 1-digit baseboard management controller (BMC) identifier in the range 0 <= *BMC* <= 1. Examples: `b0`, `b1`. The cluster manager documentation defines a *BMC* as a management card, baseboard management controller, iLO device, or node controller (nC).
NODE	A 1-digit node identifier in the range 0 <= *NODE* <= 3. 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`.

Table Continued

Variable	Specification
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 switch 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`.

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:

   - `# cm system version`

     The preceding command is new in the HPE Performance Cluster Manager 1.10 release. It displays information about the cluster manager release that is installed.

   - `# cat /etc/*release`

     The preceding command displays information about operation system distribution files and the cluster manager release. The output includes information about the cluster manager release that is installed.

Installation flow diagrams

The following figure summarizes the procedural flow for cluster manager installations on clusters without leader nodes.

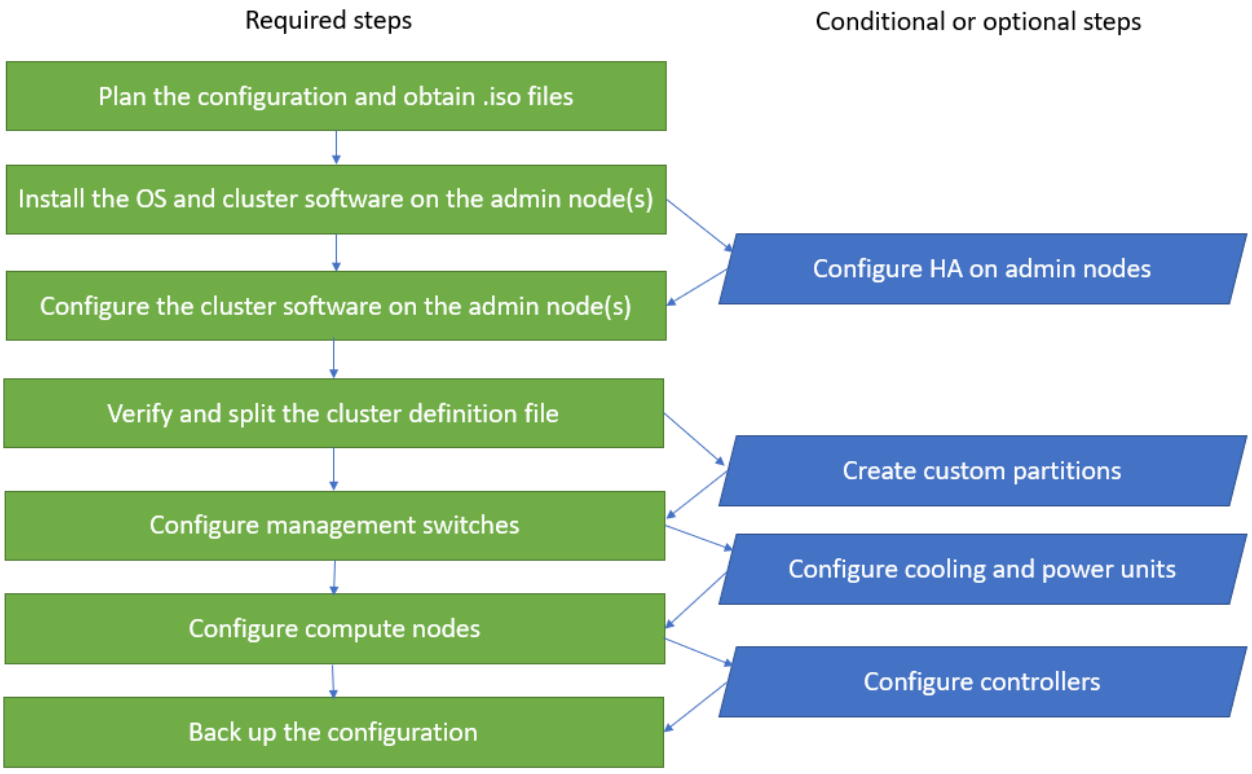

Figure 1: Installation process flow for clusters without leader nodes

The following figure shows tasks that the installer completes. The installation process for each cluster can vary depending on the cluster hardware and the software features configured.

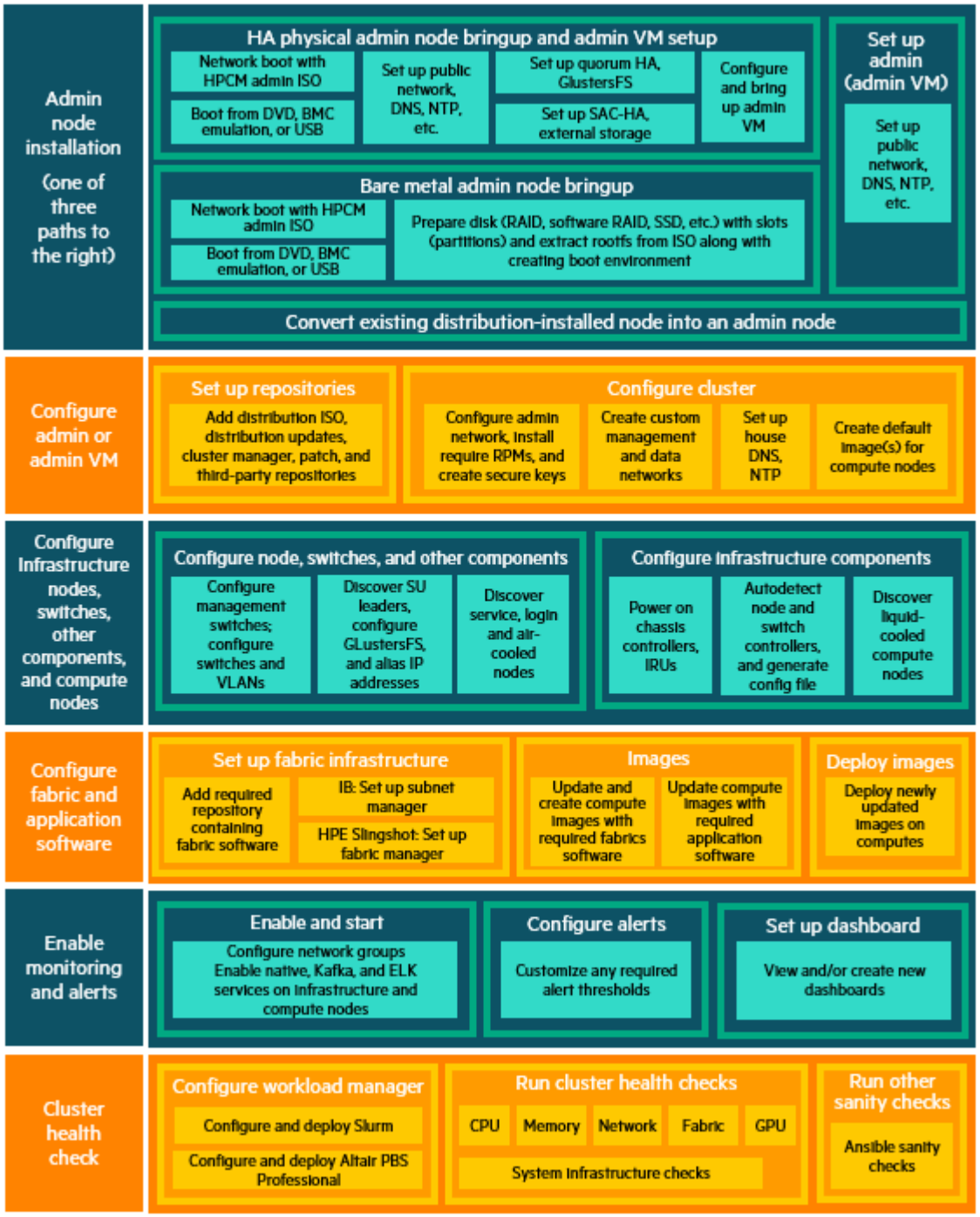

Figure 2: Installation process

Installing the operating system and the cluster manager simultaneously on the admin node

Procedure

1. **Preparing to install the operating system and the cluster manager simultaneously on the admin node**
2. **(Conditional) Preparing a USB device**
3. **(Optional) Configuring custom partitions on the admin node**
4. **Inserting the installation USB device and booting the admin node**
5. Installing an operating system. Use one of the following procedures:
   - **Configuring RHEL 8 on the admin node**
   - **Configuring SLES 15 on the admin node**
6. **(Conditional) Configuring the storage unit**
7. **(Conditional) Enabling an input-output memory management unit (IOMMU)**
8. **Verifying the configuration**

Preparing to install the operating system and the cluster manager simultaneously on the admin node

About this task

The admin installer `.iso` file is a bootable image that installs the operating system and cluster manager simultaneously. Hewlett Packard Enterprise recommends that you use this `.iso` file.

NOTE: As an alternative, you can install the operating system and the cluster manager by using the repository `.iso` file, which installs the cluster manager onto a pre-installed operating system and provides the cluster manager software needed for node images.

For information about this installation alternative, see one of the following:

Installing the operating system and the cluster manager separately

Upgrading the operating system and reinstalling the cluster manager

---

Procedure

1. Confirm that this procedure can work for you by making sure that at least one of the following is true for the cluster:
   - You received new, cabled hardware from HPE, but no software is installed on the cluster.
   - You want to configure custom partitions on the admin node, and you want the installer to configure the operating system and cluster manager together. This method assumes that you want to use the standard operating system installation parameters that are defined in the cluster manager software.

- You want the option to configure a high availability (HA) admin node.
- You want to configure two or more slots.
- A disaster occurred at your site, and you need to recover your cluster.

2. Contact your site network administrator to obtain network information for the node controller in the admin node.

   For the admin node controller, obtain the following:

   - (Optional) The current IP address of the node controller on the admin node. If you do not have this information, you can set the node controller address from a serial console.
   - The IP address you want to set for the node controller.
   - The netmask you want to set for the node controller.
   - The default gateway you want to set for the node controller.
   - A hostname.
   - The domain name.
   - An IP address.
   - The netmask.
   - The default route.
   - The root password.
   - The IP address of the local network time protocol (NTP) servers.

   Also obtain the IP addresses of the domain name servers (DNSs) on your site network.

   NOTE: To configure two nodes, as part of a two-node HA admin node configuration, make sure to obtain the necessary configuration information for both nodes.

3. Obtain operating system software directly from your operating system software vendor.

   After you obtain the operating system software, write the `.iso` file to a USB device or to a network location from which you can install the software.

   Make sure that the system is subscribed to the operating system vendor for online operating system updates.

4. Obtain the cluster manager software from HPE.

   From the customer portal, you can obtain the cluster manager installation software, including patches and updates, from the following website:

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

   The website requires you to log in with your HPE Passport account.

   If you want your initial installation to include all available cluster manager patches, log into a separate system, download the cluster manager patches, and have them available during this initial installation. For more information, see the release notes.

   As an alternative to downloading the cluster manager software, you can obtain a media kit from HPE. This media kit includes installation DVDs . If you obtain a media kit, use the instructions in the `README` file on the installation DVDs to create a `cm-admin-install .iso` file.

5. (Conditional) Configure the storage unit hardware and software for a system admin controller high availability (SAC HA) admin node.

Complete this step only if you want a SAC HA admin node.

Do not complete this step if you want to configure a quorum HA admin node.

The SAC HA admin node environment requires an HPE MSA 2050 storage unit and associated software.

When you configure the storage unit, configure one LUN per slot. If your cluster was configured at the HPE factory, the factory configured the storage unit for one LUN per slot.

You can manage the storage unit from one of the physical admin nodes or from another computer. For example, you can use a laptop to manage the storage unit.

After you install the storage unit software, start the storage unit software GUI to add the addresses and passwords of the storage controllers.

6. Attach the cluster to your site network.

   Use the procedure in the following:

   HPE Performance Cluster Manager Getting Started Guide

7. Gather information about the cluster components.

   Ideally, obtain the cluster definition file for this cluster. Proceed as follows:

   - If a cluster definition file is available, retrieve the cluster definition file for this cluster. The configuration file contains system data, for example, the MAC address information for the nodes. If you have these addresses, the node discovery process can complete more quickly.

     The cluster definition file can reside in any directory, under any name, on the cluster.

     Use the following command to create a cluster definition file and write it to a location of your own choosing:

     ```
     cm system show configfile --all > filename
     ```

     For *filename*, specify the output file name. This command writes the cluster definition file to *filename*.

     If you backed up the cluster definition file, use the backup copy at your site. If necessary, you can obtain a copy of the original cluster definition file from the HPE factory.

   - If no cluster definition file is available, plan to use the `cm node discover` command to configure nodes into the cluster.

8. (Optional) Configure a software RAID on the admin node.

   The cluster manager supports the following admin node RAID configurations:

   - A Linux RAID admin node.
   - An MD RAID 1 admin node. The following topic includes a step that allows you to configure MD RAID 1 on the admin node:

     Inserting the installation USB device and booting the admin node

   - BIOS software RAID. To implement this configuration, use the BIOS documentation to configure the BIOS RAID at this time.

   For more information, see the following:

   - The README file for the cluster manager admin node installation software.
   - **(Optional) Configuring software RAID on cluster nodes**

(Conditional) Preparing a USB device

About this task

Complete the procedure in this topic if you downloaded the cluster manager software to a network location at your site or if you assembled an installation `.iso` file from the HPE Performance Cluster Manager media kit.

NOTE: The cluster manager installation instructions assume that the operating system software is written to physical media in the form of a USB device. If you want to install the cluster manager from a network location, use your site practices to access the operating system software installation files and modify the installation instructions accordingly.

Procedure

1. Plug a USB device into the server to which you downloaded the `.iso` installation files.

   Make sure that the USB stick has a capacity of 16 GB or more.

2. Use either the Method 1 (for Linux) or Method 2 (for Windows) to write the cluster manager software to the USB device:

   Method 1 - Writing the cluster manager software from a Linux server to the USB device

   a. Plug the USB device into the Linux server to which you downloaded the ISO.

      Make sure that the USB stick has a capacity of 16 GB or more.

   b. In a terminal window, use the following command to retrieve the device name:

      ```
      # dmesg | tail [-20]
      ```

      Specify `-20` on the command if you want the full identity on the USB.

      For example:

      ```
      # dmesg | tail
      [876318.185357] scsi 10:0:0:0: Direct-Access Lexar USB Flash Drive 1100 PQ: 0 ANSI: 6
      [876318.185478] scsi 10:0:0:0: alua: supports implicit and explicit TPGS
      [876318.185481] scsi 10:0:0:0: alua: No target port descriptors found
      [876318.185774] sd 10:0:0:0: Attached scsi generic sg5 type 0
      [876318.186994] sd 10:0:0:0: [sdd] 31285248 512-byte logical blocks: (16.0 GB/14.9 GiB)
      [876318.187603] sd 10:0:0:0: [sdd] Write Protect is off
      [876318.187609] sd 10:0:0:0: [sdd] Mode Sense: 43 00 00 00
      [876318.188181] sd 10:0:0:0: [sdd] Write cache: enabled, read cache: enabled, doesn't support DPO or FUA
      [876318.198875] sdd: sdd1 sdd2 sdd3
      [876318.201520] sd 10:0:0:0: [sdd] Attached SCSI removable disk
      ```

      In the preceding example, the device name is `sdd`.

   c. Enter the following commands to find the `/dev/sdX` of the USB device:

      ```
      # dd if=/dev/zero of=/dev/sdX bs=512 count=65536
      # dd if=cm-admin-install-1.8-os.iso of=/dev/sdX bs=1024
      ```

      For *os*, specify the operating system.

   d. Extract the USB device and plug it in again.

   e. Enter the `parted` command as shown in the following example, and at the `parted` prompt, enter `p` to print the partition map:

      ```
      # parted /dev/sdX
      GNU Parted 3.2 Using /dev/sdd Welcome to GNU Parted! Type 'help' to view a list of commands.
      (parted) p
      ```

   f. (Conditional) Enter `F` to fix the error if there is an error notification.

If the following message appears, enter `F` to fix:

```
Warning: Not all of the space available to /dev/sdd appears to be used, you can fix the
GPT to use all of the space (an extra 17098052 blocks) or continue with the current setting?
Fix/Ignore? F
```

g. Enter `q` to quit.

Method 2 - Writing the cluster manager software from a Windows server to the USB device

a. Plug the USB device into the Windows system to which you downloaded the ISO.

b. Start Win32DiskImager.

c. Click the file folder icon.

d. In the **Select a disk image** popup, browse to the `.iso` file, select the `.iso` file, and click **Open**.

e. In the **Image File** field, verify the path to the location of the `.iso` file.

f. In the **Device** field, verify the destination device.

g. Click **Write**.

NOTE: If a popup window prompts you to format the disk, select **Cancel**. This window can appear multiple times.

h. When the **Complete** popup appears, click **OK**.

(Optional) Configuring custom partitions on the admin node

About this task

Complete the procedure in this topic if the default partitioning scheme does not suit the needs of this cluster. This procedure lets you choose your own layout for the system disk.

Using custom partitions on the admin node does not affect your ability to use custom partitions on the compute nodes. Likewise, using custom partitions on compute nodes does not affect your ability to use custom partitions on the admin node.

If you use custom partitions on the admin node, the cluster behaves as if it has just one root slot.

If you use custom partitions on compute nodes, you can create partitions that are different from the partitions on the admin node. If you accept default partitions on the admin node, you can still create custom partitions on the compute nodes. The following information pertains to custom partitions on compute nodes:

- If the admin node is configured to use default partitions, you can create custom partitions on compute nodes.
- If the admin node is configured to use custom partitions, you can create custom partitions on compute nodes that use a different partitioning scheme.
- You can create custom partitions on any compute node, and the partitions can be different on each compute node. Create one custom partitioning file for each partitioning scheme that you want to impose on one or more compute nodes.

The procedure in this topic explains how to specify custom partitions for the admin node. When the admin node boots, the boot process creates the partitions. The node discovery commands configure the nodes. When you run the node discovery commands, you can create the same (or different) custom partitions on the compute nodes.

NOTE: The following notes apply to custom partitions:

- If you choose to implement custom partitions on the admin node, the admin node is reduced to one slot. Keep this caveat in mind if you want to configure custom partitions on the admin node.

  Custom partitions do not apply to compute nodes configured with an NFS root file system or a `tmpfs` root file system.

  The cluster manager does not support custom admin node partitions on clusters with HA admin nodes.

  For information about the default cluster partitioning scheme, see the following:

  Default partition layout information

- Do not use the custom partitioning feature to specify additional storage. Do not custom partitioning if you need more than one slot.
- Custom partitions do not apply to compute nodes configured with an NFS file system or a `tmpfs` file system. In addition, custom partitions do not apply to compute nodes installed by using AutoYaST or Kickstart.

The following procedure explains how to create custom partitions on the admin node.

Procedure

1. Mount the cluster manager installation USB into the USB drive of a local computer at your site.

   Do not mount the installation USB into the USB drive on the admin node.

2. Read all the information in `README.install` file.

   This file resides in the root directory of the installation USB.

   This file includes general installation and custom partitioning information.

3. Read all the information in `custom_partitions_example.cfg`.

   This file resides in the root directory of the installation USB.

   This file contains information about how to use the file and about the effect of custom partitions on cluster operations.

   When you install an admin node with custom partitions, the installer destroys all other data. The destroyed data includes any slot specifications that might reside on the admin node hard disk. In other words, when you install an admin node with custom partitions, you no longer have a cluster with slots. By extension, when the admin node is configured with custom partitions, you cannot have compute nodes with multiple slots.

4. Decide where you want `custom_partitions_example.cfg` to reside.

   Typically, you write the configuration file to an NFS server at your site. Use an existing server. A later procedure explains how to specify the location to the installer at boot time.

   Alternatively, you can write the configuration file to the installation media, but this requires assistance from Hewlett Packard Enterprise.

5. Open file `custom_partitions_example.cfg` in a text editor, and specify the partitions you want for the admin node.

   The `custom_partitions_example.cfg` file consists of columns of data separated by vertical bar (`|`) characters, which separate the fields into columns. Be careful with the columns in this file. All vertical bar characters must align in order for the partitioning to complete correctly.

   For the `/opt` partition, make sure to specify enough size to create and host the images you need for the nodes.

   The file system specifications that the cluster manager supports are as follows:

- `XFS`, which is the default root file system for the cluster manager
- `ext4`
- `ext3`

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 filesystem 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.

6. Save and close the file as `custom_partitions.cfg`
7. (Conditional) Repeat the steps in this procedure on the second or third admin node.

   Repeat this procedure on the additional admin nodes that you plan to configure into a high availability (HA) admin node configuration.

Inserting the installation USB device and booting the admin node

Procedure

1. Ensure that the admin node is configured to boot from a bootable USB device or from a network location.
2. Power on the admin node.
3. (Conditional) Insert the USB device you created into the USB port on the admin node.

   Complete this step if you wrote the software to a USB device.
4. Use the arrow keys to select **Display Instructions**, and read the instructions carefully.
5. Use the arrow keys to select one of the boot options, press Enter, and monitor the installation.

   Each boot option has a set of default behaviors. Some boot options permit you to specify custom boot parameters. The options are as follows:

   - **Display Instructions**

     Select this option if you want information about custom boot parameters. This option displays information about the actionable parameters and returns to the boot menu.
   - **Install: Install to Designated Slot**

     Select this option if you have an open slot on your cluster, and you want to recreate an operating system in that open slot. If you select this option, only the open slot is affected. All other slots remain as configured.

     This boot option permits you to specify custom boot parameters.
   - **Install: Wipe Out and Start Over: Prompted**

     Select this option if you want to add slots.

     This option destroys all information currently on the cluster. The installer partitions the admin node with the specified number of slots, and the installer writes the initial installation to the designated slot. For example, for an initial installation, select this option.

- **Rescue: Prompted**

  To create a troubleshooting environment, select this option.

- **Install: Custom, type 'e' to edit kernel parameters**

  Select this option if you want to customize the installation. This option lets you supply all boot options as command-line parameters. Unlike the other boot methods, there are no system prompts for boot options. More information is available in **Display Instructions**.

  This boot option permits you to specify custom boot parameters. Hewlett Packard Enterprise recommends this option only for users with installation experience.

Example 1. To specify `console=` or any other custom boot parameter, select the **Display Instructions** option. Familiarize yourself with the parameters you want to use before you select an actionable option.

Example 2. To allocate scratch disk space on the system disk of the admin node, add the following parameters to the kernel parameter list:

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

For *size*, specify a size in GiB. The cluster manager creates the scratch disk space in partition 61, but you must otherwise structure the scratch disk space. That is, you create the file system, add the `fstab` entries, and so on. For more information about how to create scratch disk space for a node, see the following:

HPE Performance Cluster Manager Administration Guide

Example 3. To configure this node as one of the physical nodes in an HA admin node, select **Install: Wipe Out and Start Over: Prompted**.

6. Respond to the questions on the installation menus.

   All the options launch you into an installation dialog. At the end of the dialog, the final question asks you to confirm your choices. In this way, you have the chance to cancel your choices and return to the GNU GRUB boot menu to start over. The following are some of the installation dialog prompts that appear when you select a boot option:

   - **Enter number of slots to allow space for: (1-10):**

     Enter `1`, `2`, `3`, `4`, `5`, `6`, `7`, `8`, `9`, or `10`.

     This question appears only if you select **Install: Wipe Out and Start Over: Prompted** from the GNU GRUB menu. Typically, you want at least two slots.

     For more information, see the following:

     Slots

   - **Enter which slot to install to:**

     Enter `1`, `2`, `3`, `4`, `5`, `6`, `7`, `8`, `9`, or `10`.

     This question appears only if you select **Install: Install to Designated Slot** from the GNU GRUB menu.

     If you selected **Install: Wipe Out and Start Over: Prompted**, you can select slot 1.

   - **Destructively bypass sanity checks? (y/n):**

     If you enter **`y`** and press Enter, the installer proceeds without checking to see if there is any data in the partition.

     If you enter **`n`** and press Enter, the installer checks to see if there is data in the partition.

- **Is this an SAC-HA or Quorum-HA Physical Host? (normally no) (y/n):**

  To configure this node as part of an HA admin node configuration, enter **y** and press Enter.

  If this node is a standalone, non-HA admin node, enter **n** and press Enter.

- **Configure MD RAID 1 for root? (normally no) (y/n)**

  By default, the cluster manager does not create an MD RAID 1 boot disk for the admin node.

  To create a Linux software MD RAID 1 boot disk for the admin node, enter **y**. The admin node can boot from this RAID.

- **Supply devices(s) for root, comma separated (optional)**

  By default, the cluster manager chooses the disk from which the admin node can boot.

  When you specify disks, you can avoid unpredictable results by specifying persistent disk names, which include a `by-path` identifier or a `by-id` identifier.

  For example:

  ```
  /dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:1
  ```

  For example, to configure a Linux software MD RAID 1 for the admin node boot disk, specify a comma-separated list of disk devices to include in the RAID.

- **Use predictable network names for the admin node? (normally yes) (y/n):**

  This question determines whether predictable names or legacy names are assigned to the network interface cards (NICs) in the node.

  To configure the admin node with predictable names, enter **y** and press Enter. The following types of nodes can use predictable names:

  - Standalone admin nodes
  - The virtual machine admin node that is part of a high availability (HA) admin node

  Hewlett Packard Enterprise recommends that you enter **y** when possible.

  To configure a physical admin node as part of a two-node HA admin node, enter **n** and press Enter. This action configures the node with legacy names. A physical admin node that is part of an HA admin node requires legacy names.

  For information about predictable network names, see the following:
  Predictable network interface card (NIC) names

- **Start the network (DHCP) and sshd on port 40? (y/n)n**

  If you enter `n`, or accept the default of `n`, the installer does not use a network when it installs the admin node software. The installer does not start an `ssh` client.

  If you enter `y`, you can `ssh` into the admin node to troubleshoot an installation. The root password is `cmdefault`. Typically, if you are troubleshooting an admin node installation, you can access the admin node only through the video screen or the serial console. If you enter `y` at this prompt, you gain the following:

  - An additional method for observing the admin node installation.
  - More rescue mode options.

- **Additional parameters (like console=, etc):**

  Supply additional parameters as follows:

- To configure an HA admin node based on an HPE ProLiant DL360 or an HPE ProLiant DL325, specify a `console=`*`type`* parameter. You need to determine whether to specify `console=ttyS0` or `console=ttyS1` parameter based on the admin node type. For most HPE ProLiant admin nodes, specify `console=ttyS0`.

  For example:

  ```
  console=ttyS0,115200n8
  ```

- To specify additional boot parameters, enter them in a space-separated list and press Enter.
- To configure custom partitions, add the target for the custom partitions file, as follows:

  `custom_partitions=`*`NFS_server_address`*`:/`*`path_to_custom_partitions`*`.cfg`

  The variables are as follows:

Variable	Specification
NFS_server_address	The identifier of your site NFS server. This address can be an IP address or a hostname.
path_to_custom_partitions`.cfg`	The full path to the custom partitioning file.

The cluster manager documentation does not describe a network install of an HA admin node. However, if you install an HA admin node over the network, specify `sac_ha=1` as a boot parameter.

For information about all the boot parameters that are available, select **Display Instructions** from the GNU GRUB menu and press Enter.

- **OK to proceed? (y/n):**

  If you enter **`y`** and press Enter, the boot proceeds.

  If you enter **`n`** and press enter, the menu returns you to the main GNU GRUB menu.

7. Wait for the installation to complete.

   The installation can take several minutes.

   If the system issues a failure message, scroll up to the top of the message to display the steps that explain how to recover the installation. Complete the recovery steps, and continue with this procedure.

8. Remove the operating system installation USB device.

9. At the # prompt, enter **`reboot`**.

   This boot is the first boot from the admin node hard disk.

10. (Conditional) Repeat the steps in this procedure on the second or third admin node.

    Repeat this procedure on the additional admin nodes that you plan to configure into a high availability (HA) admin node configuration.

Configuring RHEL 8 on the admin node

Procedure

1. Use one of the following methods to log into each physical admin node as the root user:

- Use the intelligent platform management interface (IPMI) tool
- Use the keyboard, video display terminal, and mouse (KVM) equipment attached to the console or attach your own KVM equipment to the cluster

To configure a high availability (HA) admin node, complete the steps in this procedure on all physical admin nodes.

2. Enter the following command to retrieve current time zone information:

```
# date
Fri Apr 20 10:12:50 CDT 20XX
```

The previous output is an example that shows the admin node set to US central daylight time. If the output you see is **not** correct for this cluster, complete the following steps:

a. Enter the following command to display a list of time zones:

```
# timedatectl list-timezones
```

b. Use the following command to set the time zone:

```
timedatectl set-timezone time_zone
```

For *time_zone*, specify one of the time zones from the `timedatectl list-timezones` command output.

When finished, you can use the `timedatectl` command to display the time zone information you configured. For example:

```
# timedatectl
Local time: Fri 20XX-04-15 14:55:33 PDT
Universal time: Fri 20XX-04-15 21:55:33 UTC
RTC time: Fri 20XX-04-15 21:55:33
Time zone: America/Los_Angeles (PDT, -0700)
NTP enabled: yes
NTP synchronized: yes
RTC in local TZ: no
DST active: yes
Last DST change: DST began at
Sun 20XX-03-13 01:59:59 PST
Sun 20XX-03-13 03:00:00 PDT
Next DST change: DST ends (the clock jumps one hour backwards) at
Sun 20XX-11-06 01:59:59 PDT
Sun 20XX-11-06 01:00:00 PST
```

3. Enter the following command to set the admin node hostname:

```
# hostnamectl set-hostname admin_node_hostname
```

For *admin_node_hostname*, make sure to enter the hostname, which is the short name. Do not enter the fully qualified domain name (FQDN), which is the longer name.

If you complete this step as part of an HA admin node configuration, specify the *admin_node_hostname* of the node you are configuring at this time.

4. Complete the following steps to direct network time protocol (NTP) server requests to the server at your site rather than the public time servers of the `pool.ntp.org` project:

a. Use a text editor to open file `/etc/chrony.conf`.

b. Insert a pound character (#) into column 1 of each line that includes `rhel.pool.ntp.org`.

c. At the end of the file, add lines for the following:

- Identification for the NTP servers at your site. Add `server xxx.xxx.xxx.xxx iburst` lines.
- NTP broadcasting to the management network and the baseboard management controller (BMC) network . Add `allow 172.2x.0` lines.

NOTE: Specify the IP addresses of the public time servers at your site. Do not specify the DNS hostnames or FQDNs for the public time servers at your site.

For example:

```
server 150.166.33.20    iburst
server 150.166.33.25    iburst
server 150.166.33.89    iburst
allow 172.23.0
allow 172.24.0
```

d. Save and close the file.

5. Use a text editor to open file `/etc/hosts`.

6. For each physical admin node, add an entry in the `/etc/hosts` file that contains the network address and the FQDN for each node.

   Use the following format:

   admin_node_IP admin_node_FQDN admin_node_hostname

   The variables in the line are as follows:

Variable	Specification
admin_node_IP1	The IP address of a physical admin node.
[*admin_node_IP2*]	For a single-node admin node, specify the IP address.
[*admin_node_IP3*]	For system admin controller high availability (SAC HA) admin nodes, create an additional line for the second admin node. For quorum HA admin nodes, create additional lines for the second admin node and the third admin node.
admin_node_FQDN	The FQDN of the physical admin node.
admin_node_hostname	The hostname of the physical admin node.

Example 1. If you have one physical admin node, add a line similar to the following:

```
# physical node address
100.162.244.251 acme-admin.acme.usa.com acme-admin
```

Example 2. If you have two physical admin nodes as part of a SAC-HA configuration, add lines similar to the following:

```
# physical node addresses
   100.162.244.251   acme-admin1.acme.usa.com    acme-admin1
   100.162.244.252   acme-admin2.acme.usa.com    acme-admin2
```

Example 3. If you have three physical admin nodes as part of a quorum HA configuration, add lines similar to the following:

```
# physical node addresses
   100.162.244.251   acme-admin1.acme.usa.com    acme-admin1
   100.162.244.252   acme-admin2.acme.usa.com    acme-admin2
   100.162.244.253   acme-admin3.acme.usa.com    acme-admin3
```

7. Save and close file `/etc/hosts`.

8. Use a text editor to create the following file:

`/etc/resolv.conf`

Add the following information to `resolv.conf`, and then save and close the file:

- The `nameserver` keyword and the IP address of the name server at your site.
- The `search` keyword and the FQDN.

For example:

```
search cluster.publicdomain.com publicdomain.com
nameserver 150.150.39.101
```

9. Use the `ip addr show` command to determine the following:

- The name of the network interface card (NIC) that connects the admin node to the house network.
- The MAC address of the NIC that connects the admin node to the house network.

For example, in the following output, the NIC name is `ens20f0` and the MAC address is `00:25:90:fd:3d:a8`:

```
admin # ip addr show
1: lo:  mtu 65536 qdisc noqueue state UNKNOWN qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: ens20f0:  mtu 1500 qdisc mq state UP qlen 1000
    link/ether 00:25:90:fd:3d:a8 brd ff:ff:ff:ff:ff:ff
    inet 192.168.8.1/24 brd 192.168.8.2 scope global ens20f0
       valid_lft forever preferred_lft forever
    inet6 fe80::225:90ff:fefd:3da8/64 scope link
       valid_lft forever preferred_lft forever
3: ens20f1:  mtu 1500 qdisc mq master bond0 state UP qlen 1000
    link/ether 00:25:90:fd:3d:a9 brd ff:ff:ff:ff:ff:ff
4: ens20f2:  mtu 1500 qdisc mq master bond0 state DOWN qlen 1000
    link/ether 00:25:90:fd:3d:a9 brd ff:ff:ff:ff:ff:ff
5: ens20f3:  mtu 1500 qdisc mq state DOWN qlen 1000
```

```
    link/ether 00:25:90:fd:3d:ab brd ff:ff:ff:ff:ff:ff
.
.
.
```

10. Use a text editor to open file `ifcfg-name`.

File `ifcfg-name` is the configuration file for the NIC that is connected to the house network. For example, `ifcfg-ens20f0`.

The path to this file is as follows:

`/etc/sysconfig/network-scripts/ifcfg-name`

11. In the `ifcfg-name` file, update the following lines with the information for this cluster:

```
NAME=name      # Add the name of the house NIC
DEVICE=name    # Add the name of the house NIC
IPADDR=        # Add the IP address of this admin node
PREFIX=        # Add your site netmask setting.  For example:  24
GATEWAY=       # Add your site gateway
DNS1=          # Add your site primary DNS IP address
DNS2=          # Add your site secondary DNS IP address
DOMAIN=        # Add your site domain
HWADDR=        # Add the NIC MAC address
BOOTPROTO=     # Set to "none"
ONBOOT=        # Set to "yes"
DEFROUTE=      # Set to "yes"
TYPE=          # Set to "Ethernet"
UUID=          # Enter "nmcli connection show" to retrieve the UUID value
```

For information about how the inputs to the `ifcfg-name` file, see the following:

- Your RHEL documentation
- The `nm-settings-ifcfg-rh` manpage

NOTE: The cluster software does not support IPV6 on the public NIC in the admin node. The following line is needed for this installation:

`IPV6INIT="no"`

You can remove the lines that start with `IPV6_` from the `ifcfg-name` file, or you can retain those lines for completeness.

---

12. Save and close file `ifcfg-name`.

13. Reboot the physical admin node, and watch the reboot on the console..

Wait for the physical admin node to come back up.

14. Log into each physical admin node as the root user.

15. Enter a `ping` command to another server at your site to make sure that the network is functioning.

Wait for the `ping` to return.

16. Log out of the console window, and use the `ssh` command to log into the admin node as the root user.

Configuring SLES 15 on the admin node

About this task

The procedure in this topic uses the SLES YaST interface. To navigate YaST, use key combinations such as the following:

- Press tab to move the cursor forward.
- Press Shift + tab to move the cursor backward.
- Press the arrow keys to move the cursor up, down, left, and right.
- To use shortcuts, press the Alt key + the highlighted letter.
- Press Enter to complete or confirm an action.
- Press Ctrl + L to refresh the screen.

Procedure

1. Use one of the following methods to log into each physical admin node as the root user:

   - Use the intelligent platform management interface (IPMI) tool
   - Use the keyboard, video display terminal, and mouse (KVM) equipment attached to the console or attach your own KVM equipment to the cluster

   To configure a high availability (HA) admin node, complete the steps in this procedure on all physical admin nodes.

2. Enter the following command to retrieve current time zone information:

   ```
   # date
   Fri Apr 20 10:12:50 CDT 20XX
   ```

   The previous output is an example that shows the admin node set to US central daylight time. If the output you see is **not** correct for this cluster, complete the following steps:

   a. Enter the following command to display a list of time zones:

      ```
      # timedatectl list-timezones
      ```

   b. Use the following command to set the time zone:

      ```
      timedatectl set-timezone time_zone
      ```

      For *time_zone*, specify one of the time zones from the `timedatectl list-timezones` command output.

      When finished, you can use the `timedatectl` command to display the time zone information you configured. For example:

      ```
      # timedatectl
      Local time: Fri 20XX-04-15 14:55:33 PDT
      Universal time: Fri 20XX-04-15 21:55:33 UTC
      RTC time: Fri 20XX-04-15 21:55:33
      Time zone: America/Los_Angeles (PDT, -0700)
      NTP enabled: yes
      NTP synchronized: yes
      ```

```
RTC in local TZ: no
DST active: yes
Last DST change: DST began at
Sun 20XX-03-13 01:59:59 PST
Sun 20XX-03-13 03:00:00 PDT
Next DST change: DST ends (the clock jumps one hour backwards) at
Sun 20XX-11-06 01:59:59 PDT
Sun 20XX-11-06 01:00:00 PST
```

3. Enter the following command to set the admin node hostname:

   ```
   # hostnamectl set-hostname admin_node_hostname
   ```

   For *admin_node_hostname*, make sure to enter the hostname, which is the short name. Do not enter the fully qualified domain name (FQDN), which is the longer name.

   If you complete this step as part of a high availability HA admin node configuration, specify the *admin_node_hostname* of the node you are configuring at this time.

4. Complete the following steps to direct network time protocol (NTP) server requests to the server at your site rather than the public time servers of the `pool.ntp.org` project:

   a. Use a text editor to open file `/etc/chrony.conf`.

   b. Search for the following line in `/etc/chrony.conf`:

      ```
      ! pool pool.ntp.org iburst
      ```

      In column 1, replace the exclamation point character (`!`) with a pound sign character (`#`).

   c. At the end of the file, add lines for the following:

      - Identification for the NTP servers at your site. Add `server xxx.xxx.xxx.xxx iburst` lines.
      - NTP broadcasting to the management network and the baseboard management controller (BMC) network . Add `allow 172.2x.0` lines.

      NOTE: Specify the IP addresses of the public time servers at your site. Do not specify the DNS hostnames or FQDNs for the public time servers at your site.

      For example:

      ```
      server 150.166.33.20    iburst
      server 150.166.33.25    iburst
      server 150.166.33.89    iburst
      allow 172.23.0
      allow 172.24.0
      ```

   d. Save and close the file.

5. Enter the following commands to start YaST2:

   ```
   # export Textmode=1
   # export TERM=xterm
   # /usr/lib/YaST2/startup/YaST2.Firstboot
   ```

In addition, you might need to alter your environment. For example to run YaST from a PuTTY window, also enter the following:

```
# export NCURSES_NO_UTF8_ACS=1
```

For information about navigation, see the following:

YaST navigation

6. On the **Language and Keyboard Layout** screen, complete the following steps:

   a. Select your language.

   b. Select your keyboard layout.

   c. Select **Next**.

7. On the **Welcome** screen, select **Next**.

8. On the **License Agreement** screen for the operating system, complete the following steps:

   a. Tab to the box (`[ ] I Agree ...`).

   b. Press the space bar to accept the license terms. This action puts an `x` in the box, so it looks like this: `[x]`.

   c. Select **Next**.

   d. (Conditional) If there are more license agreement screens, select **Next** again.

9. On the **Network Settings** screen, prepare to specify the NIC information.

   Complete the following steps:

   a. Highlight the NIC with the lowest MAC address. Look at the final octet in each MAC address.

      For example, if the node includes the following NICs, highlight the NIC numbered `ec:eb:b8:89:f2:90`:

      ```
      hikari2:~ # ip addr | grep ether
          link/ether ec:eb:b8:89:f2:90 brd ff:ff:ff:ff:ff:ff    # lowest
          link/ether ec:eb:b8:89:f2:91 brd ff:ff:ff:ff:ff:ff
          link/ether ec:eb:b8:89:f2:92 brd ff:ff:ff:ff:ff:ff
          link/ether ec:eb:b8:89:f2:93 brd ff:ff:ff:ff:ff:ff
      ```

   b. Select **Edit**.

10. On the **Network Card Setup** screen, complete the following steps to specify the admin node public NIC:

    a. Select **Statically Assigned IP Address**. Hewlett Packard Enterprise recommends a static IP address, not DHCP, for the admin node.

    b. In the **IP Address** field, enter the admin node IP address. This IP address is the IP address for users to use when they want to access the cluster.

    c. In the **Subnet Mask** field, enter the admin node subnet mask.

d. In the **Hostname** field, enter the admin node FQDN. HPE requires you to enter an FQDN, not the shorter hostname, into this field. For example, enter `admin.cm.clusterdomain.com`. Failure to supply an FQDN in this field causes the `configure-cluster` command to fail.

e. Select **Next**.

You can specify the default route, if needed, in a later step.

11. On the **Network Settings** screen, complete the following steps:

a. Select **Hostname/DNS**.

b. In the **Hostname** field, enter the admin node hostname.

c. In the **Name Servers and Domain Search List**, enter the IP addresses of the name servers for your house network.

d. In the **Domain Search** field, enter the domains for your site.

e. Back at the top of the screen, select **Routing**.

The **Network Settings > Routing** screen appears.

f. In the **Default IPV 4 Gateway** field, enter your site default gateway.

g. Select **Next**.

12. On the **Local User** screen, complete one of the following actions:

- Provide information for additional user accounts and select **Next**.

  or

- Select **Skip User Creation** and select **Next**.

13. On the **Authentication for the System Administrator "root"** screen, complete the following steps:

a. In the **Password for root User** field, enter the password you want to use for the root user.

This password becomes the root user password for all the system nodes.

b. In the **Confirm password** field, enter the root user password again.

c. In the **Test Keyboard Layout** field, enter a few characters.

For example, if you specified a language other than English, enter a few characters that are unique to that language. If these characters appear in this plain text field, you can use these characters in passwords safely.

d. Select **Next**.

e. (Conditional) Confirm the password on the popup that appears.

Complete this step if a password popup appears.

14. On the **Installation Completed** screen, select **Finish**.

15. Use a text editor to open file `/etc/hosts`.

16. For each physical admin node, add an entry in the `/etc/hosts` file that contains the network address and the FQDN for each node.

Use the following format:

```
admin_node_IP admin_node_FQDN admin_node_hostname
```

The variables in the line are as follows:

Variable	Specification
admin_node_IP	The IP address of an admin node.
[*admin_node_IP2*]	For a single-node admin node, specify the IP address.
[*admin_node_IP3*]	For system admin controller high availability (SAC HA) admin nodes, create an additional line for the second admin node. For quorum HA admin nodes, create additional lines for the second admin node and the third admin node.
admin_node_FQDN	The FQDN of the admin node.
admin_node_hostname	The hostname of the admin node(s).

Example 1. If you have one physical admin node, add a line similar to the following:

```
# physical node address
100.162.244.251 acme-admin.acme.usa.com acme-admin
```

Example 2. If you have two physical admin nodes as part of a SAC-HA configuration, add lines similar to the following:

```
# physical node addresses
   100.162.244.251   acme-admin1.acme.usa.com    acme-admin1
   100.162.244.252   acme-admin2.acme.usa.com    acme-admin2
```

Example 3. If you have three physical admin nodes as part of a quorum HA configuration, add lines similar to the following:

```
# physical node addresses
   100.162.244.251   acme-admin1.acme.usa.com    acme-admin1
   100.162.244.252   acme-admin2.acme.usa.com    acme-admin2
   100.162.244.253   acme-admin3.acme.usa.com    acme-admin3
```

17. Reboot the physical admin nodes, and watch the reboot on the console.

Wait for the physical admin nodes to come back up.

18. Log into the admin nodes as the root user.

19. Enter a `ping` command to another server at your site to make sure that the network is functioning.

Wait for the `ping` to return.

20. Log out of the console window, and use the `ssh` command to log into the admin node as the root user.

21. (Optional) Add `admin` to the No Proxy Domains line (`no_proxy=` line).

If using a proxy, ensure that `admin` is added to the `No Proxy Domains` line in the YaST2 proxy settings for the following:

- The admin node
- The virtual admin nodes of a high availability (HA) cluster
- The login nodes

For information about how to configure a proxy, see the SLES documentation.

(Conditional) Configuring the storage unit

About this task

Complete this procedure if you want to configure a system admin controller high availability (SAC HA) admin node.

For the storage unit, the typical configuration is a 2-LUN storage unit. For a cluster with two slots, you can use one LUN per slot.

The following procedure assumes the following about the storage unit:

- The unit is attached to the two admin nodes.
- It is known to be working properly.
- It hosts no content that you want to save. This procedure wipes the storage completely.

Procedure

1. Enter the `lsscsi` command on each physical node to determine the disk devices that each node can recognize.

   In the `lsscsi` output, the storage unit reports as `MSA 2050 SAS`.

   For example, the following output shows that the nodes can recognize disks `/dev/sdc` and `/dev/sdd`. The same disks also appear as `/dev/sde` and `/dev/sdf`, which are secondary paths.

   - On physical node 1, the following output shows that the MSA 2050 devices host `/dev/sdc` and `/dev/sdd`:

     ```
     # lsscsi
     [0:0:0:0]    disk    Generic- SD/MMC CRW       1.00  /dev/sdb
     [15:0:0:0]   enclosu HPE      Smart Adapter    1.04  -
     [15:1:0:0]   disk    HPE      LOGICAL VOLUME   1.04  /dev/sda
     [15:2:0:0]   storage HPE      P408i-a SR Gen10 1.04  -
     [16:0:0:0]   enclosu HP       MSA 2050 SAS     G22x  -
     [16:0:0:1]   disk    HP       MSA 2050 SAS     G22x  /dev/sdc
     [16:0:0:2]   disk    HP       MSA 2050 SAS     G22x  /dev/sdd
     [16:0:1:0]   enclosu HP       MSA 2050 SAS     G22x  -
     [16:0:1:2]   disk    HP       MSA 2050 SAS     G22x  /dev/sde
     [16:0:1:3]   disk    HP       MSA 2050 SAS     G22x  /dev/sdf
     ```

   - On the physical node 2, the following output shows that the MSA 2050 devices host `/dev/sdc` and `/dev/sdd`:

     ```
     # lsscsi
     [0:0:0:0]    disk    Generic- SD/MMC CRW       1.00  /dev/sdb
     [14:0:0:0]   enclosu HPE      Smart Adapter    1.04  -
     [14:1:0:0]   disk    HPE      LOGICAL VOLUME   1.04  /dev/sda
     ```

```
[14:2:0:0]   storage HPE      P408i-a SR Gen10 1.04  -
[15:0:0:0]   enclosu HP       MSA 2050 SAS     G22x  -
[15:0:0:1]   disk    HP       MSA 2050 SAS     G22x  /dev/sdc
[15:0:0:2]   disk    HP       MSA 2050 SAS     G22x  /dev/sdd
[15:0:1:0]   enclosu HP       MSA 2050 SAS     G22x  -
[15:0:1:2]   disk    HP       MSA 2050 SAS     G22x  /dev/sde
[15:0:1:3]   disk    HP       MSA 2050 SAS     G22x  /dev/sdf
```

The preceding output is an example. The device IDs associated with each disk vary by node and might be different for your configuration.

2. Enter the `pvscan` command on each physical node to determine the disk devices that are initialized and in use currently.

   In the `pvscan` output, the unused devices are **not** listed.

   For example:

   - On the first physical node, the following output shows that devices `/dev/sdc` and `/dev/sda2` are in use:

     ```
     # pvscan
     PV /dev/sdc    VG vgha1          lvm2 [100 GiB / 0    free]
     PV /dev/sda2   VG vg_host        lvm2 [200 GiB / 0    free]
     Total: 2 [300 GiB] / in use: 2 [0 GiB] / in no VG: 0 [0   ]
     ```

   - On the second physical node the following output shows that devices `/dev/sdc` and `/dev/sde` are in use:

     ```
     # pvscan
     PV /dev/sdc    VG vgha1          lvm2 [100 GiB / 0    free]
     PV /dev/sde    VG vg_host        lvm2 [200 GiB / 0    free]
     Total: 2 [300 GiB] / in use: 2 [0 GiB] / in no VG: 0 [0   ]
     ```

3. Based on your analysis of the `lsscsi` and `pvscan` commands, choose a disk that both nodes can recognize and that is not currently in use.

   A disk that appears in the `pvscan` output is initialized and might already contain data. Do not select a disk that appears in the `pvscan` output because it is likely that the disk already contains data. Any data currently stored on a disk that appears in `pvcsan` output is destroyed when the HA admin node begins to run. As an alternative, you can move the data to another disk. Proceed with caution.

   For example, the output in the preceding steps indicates the following:

   - `/dev/sdc` is recognized by both physical nodes.
   - `/dev/sdd` is not in use currently.

   In this example environment, `/dev/sdc` is a safe choice for the common disk.

4. Identify the world wide name (WWN) of the disk you want to use for the HA admin node.

   To identify the disk, use a combination of `ls` and `grep` commands. The following example shows the command that returns the WWN of the disk you chose:

   ```
   # ls -l /dev/disk/by-id/ | grep wwn
   lrwxrwxrwx 1 root root  9 Nov 10 13:26 wwn-0x60080e5000233c340000039f4d90ab57 -> ../../sdc
   lrwxrwxrwx 1 root root 10 Nov 10 13:26 wwn-0x60080e5000233c340000039f4d90ab57-part1 -> ../../sdc1
   lrwxrwxrwx 1 root root 10 Nov 10 13:26 wwn-0x60080e5000233c340000039f4d90ab57-part2 -> ../../sdc2
   ```

The preceding command returned information about the disk itself and two partitions. Use the WWN of the disk itself, not the disk partitions. In this example, the WWN for the disk is as follows:

`0x60080e5000233c340000039f4d90ab57`

Observe the ID. A later procedure requires you to specify this WWN in the `sac-ha-initial-setup.conf` file.

CAUTION: This procedure uses data from an example environment. Do not assume that your environment can yield the same results. In your environment, correct disk analysis is not likely to produce the same effect. Do not assume that the analysis of your environment will also lead you to select `/dev/sdc` as your HA admin node shared disk.

5. Erase the existing data on the shared disk.

NOTE: This step is destructive. If necessary, preserve the data now by moving the data from the shared disk to another disk at your site.

As the root user, enter the following commands from one of the physical admin nodes:

```
# parted /dev/sdX mklabel gpt
# dd if=/dev/zero of=/dev/sdX bs=512 count=16384
```

For *X*, specify the identifier for the disk you want to erase.

(Conditional) Enabling an input-output memory management unit (IOMMU)

About this task

Complete this procedure if the following are both true:

- You want to configure a system admin controller high availability (SAC HA) admin node.
- The physical admin nodes are Intel platform admin nodes such as HPE Proliant DL360 servers.

Procedure

1. Log into each of the physical admin nodes as the root user.
2. On each physical admin node, open the following file in a text editor:

   `/etc/default/grub`
3. Search for the following string in the file:

   `GRUB_CMDLINE_LINUX_DEFAULT`
4. Add `intel_iommu=on` to the end of the `GRUB_CMDLINE_LINUX_DEFAULT` line.
5. On each physical admin node, save and close the edited file.
6. On each physical admin node, enter one of the following commands:

   On RHEL systems, enter the following:

   ```
   # grub2-mkconfig -o /boot/efi/EFI/redhat/grub.cfg
   ```

On SLES systems, enter the following:

```
# grub2-mkconfig -o /boot/grub2/grub.cfg
```

Verifying the configuration

Procedure

1. Log into each physical admin node as the root user.

   Complete the steps in this procedure on all physical admin nodes.

2. Enter the following command to verify the time zone:

   ```
   # date
   ```

3. Enter the following command to verify the hostname and the IP address:

   ```
   # cat /etc/hosts
   ```

4. Enter the following command to verify the time:

   ```
   # chronyc sources -v
   ```

5. Enter the following command to verify the network configuration:

   ```
   # ip addr
   ```

6. Use the `hostnamectl` command to verify that the static host is set.

   For example, in the following output, the first line is `Static hostname: name`. Make sure that the hostname you specified for the physical admin node is the one that appears in the *name* field. The output shows the hostname set correctly on physical node `hikari`.

   ```
   # hostnamectl
      Static hostname: hikari
            Icon name: computer-server
              Chassis: server
           Machine ID: 68c22b359c3b486a8576088cc3538beb
              Boot ID: 1274c1d3b2884cacb90d368c616b2ed5
     Operating System: Red Hat Enterprise Linux 8.X (Ootpa)
          CPE OS Name: cpe:/o:redhat:enterprise_linux:8.X:GA
               Kernel: Linux 4.18.0-80.el8.x86_64
         Architecture: x86-64
   ```

7. (Conditional) Verify that IOMMU is enabled.

   Complete this step if you completed the following procedure:

   (Conditional) Enabling an input-output memory management unit (IOMMU)

   Enter the following command:

   ```
   # dmesg | grep -E "DMAR: IOMMU"
   ```

   The output is as follows on a correctly configured system:

   ```
   [    0.000000] DMAR: IOMMU enabled
   ```

8. (Conditional) Verify that the physical admin nodes can communicate with each other.

   Complete this step on physical admin nodes configured for high availability (HA).

For admin nodes configured as system admin controller high availability (SAC HA) nodes, enter a `ping` command from physical admin node 1 to physical admin node 2. Also enter a `ping` command from physical admin node 2 to physical admin node 1.

For admin nodes configured as quorum high availability HA nodes, enter commands as follows:

- Enter a `ping` command from physical admin node 1 to the following:
  - Physical admin node 2
  - Physical admin node 3
- Enter a `ping` command from physical admin node 2 to the following:
  - Physical admin node 1
  - Physical admin node 3
- Enter a `ping` command from physical admin node 3 to the following:
  - Physical admin node 1
  - Physical admin node 2

Slots

You can configure the cluster to boot from up to 10 slots. A **slot** consists of all the partitions related to a Linux installation.

On a factory-configured cluster, the default number of slots is one.

Multiple slots, especially on the admin node, can lead to a smoother update when it is time to upgrade the cluster manager or operating system software.

When the cluster is configured with two or more slots, you can clone a production slot to an alternative location, thus creating a fallback slot.

A multiple-slot disk layout creates the same disk layout on all nodes. Each slot includes the following:

- A `/boot` partition.
- A `/`, or root, partition.
- A `/boot/efi` partition. A slot includes this partition only if the node is an EFI node.

When you insert the cluster manager operating system installation disk and power on the admin node, you can select a boot method from the GNU GRUB menu. If you select **Install: Wipe Out and Start Over: Prompted**, the installer creates two slots and writes the initial installation to slot 1. After the system is installed, you cannot change the number of slots. If you attempt to change the number of slots, you destroy the data on the disks.

After you install a multislot cluster, you can boot the cluster with the operating system of your choice. This capability might be useful if you ever want to test an operating system or other software. When you have more than one slot, you can roll back an upgrade completely.

The following are some other characteristics of multiple-slot systems and single-slot systems:

Multiple-slot	Single-slot
You can install different operating systems, or different operating system versions, into different slots.	You can install only one operating system for the entire cluster.
As you increase the number of slots, you decrease the amount of disk space per slot. Hewlett Packard Enterprise recommends a minimum of 100 GB per slot.	A single slot uses all available disk space.

(Optional) Configuring a quorum high availability (quorum HA) admin node

About this task

The quorum HA solution eliminates the need for the HPE Modular Smart Array 2050 shared storage system that the system admin controller (SAC) HA solution requires.

A quorum HA admin node uses the Gluster file system in sharding mode to host a virtual machine image. When the cluster is running, the admin node resides in a virtual machine upon one of three servers configured as physical admin nodes. When a failover occurs, the virtual machine passes from the active node to one of the passive nodes. It uses Pacemaker to start and position the virtual machine as needed.

Step **4** in this quorum HA configuration procedure requires you to run the `/opt/clmgr/lib/q-ha/setup` script, which has the following effects:

- On each of the three servers, the installer creates the following:
  - Network bridges `br0` and `br1`.

    If the cluster has a separate BMC network, the installer also creates `br2`.
  - Bond `bond0`.
- The installer places the network interface cards (NICs) that are assigned to the management network and are listed as `phys[1-3]_mgt_nic[1-2]_ifname` in the table called **Table 3: Fields in the `hadb.conf` file** in an 802.3ad bond as `bond0`. The installer assigns this bond to `br1`.

  The 802.3ad link aggregation might require switch configuration.
- The installer creates a virtual machine disk named `adminvm.img` on the Gluster volume mounted at `/adminvm`. The configuration file for the virtual machine is located at `/adminvm/adminvm.xml`.
- The installer creates a virtual admin node on one of the servers in the quorum HA configuration. Enter one of the following commands to display the name of the host where the virtual admin is currently running:
  - On a RHEL cluster, enter the following command:

    ```
    # pcs resource status virt
    ```
  - On a SLES cluster, enter the following command:

    ```
    # crm resource status adminvm
    ```

The following procedure explains how to configure a quorum HA admin node.

Procedure

1. Verify the following:
   - Verify that the three physical servers that you want to configure into a quorum HA admin node are identical and are equipped with the following:

- An x86_64 architecture.
- Three or four NICs, dedicated as follows:

  a. One for the house (or site) network

  b. Two for the cluster management network

  c. If the cluster has separate BMC network, one for the BMC network

- Two storage devices dedicated as follows: one for the operating system and one for Gluster storage. Verify that the Gluster storage devices are of approximately the same size on each server.

- Verify that none of the NICs that reside in the physical admin servers are bonded.

  The quorum HA configuration scripts fail if any NICs in the physical admin nodes are bonded. Dismantle the bonding among any bonded NICs. You can reenable NIC bonding after the quorum HA configuration procedure is complete.

  For example, if you suspect that one or more NICs are bonded as `bond0`, display the contents of the `/bonding` directory. The following example shows that `ens2f0np0` and `ens2f1np1` are bonded at `bond0`:

  ```
  # cat /proc/net/bonding/bond0
  Bonding Mode: IEEE 802.3ad Dynamic link aggregation
  Slave Interface: ens2f0np0
  Slave Interface: ens2f1np1
  ```

- Verify that each of the three servers has an IP address on the house network and that all three servers can reach each other using an `ssh` connection. Passwordless `ssh` does not have to be configured.
- Verify that each of the three servers is equipped with a management card, which can be either an iLO device or a baseboard management controller (BMC). For each server node, verify that each server can reach the management cards in the other servers.
- Verify that each of the three servers are attached to the house network.

2. Through a console, log into the management card of one of the physical admin nodes.

The console can be a serial console or a virtual screen. For example, you can use KVM/Console redirection.

Because the configuration script used in this procedure restarts the network, you cannot log into one of these nodes through an `ssh` connection. The restart will break an `ssh` connection.

3. Copy the admin installer `.iso` file and the Linux distribution `.iso` file into the following directory on the physical admin node:

`/var/opt/sgi`

The installer synchronizes this directory to the other two physical admin nodes.

Each operating system requires at least one operating system `.iso` file, and some operating systems require additional `.iso` files. For example, for physical quorum HA admin nodes that run Rocky Linux, the high availability files are made available automatically in the following directory and are used by the quorum HA set-up tool:

`/var/opt/sgi/Rocky-HighAvailability-8.*.iso`

Review the tables in the following topic to make sure that you have all the software you need:

HPE Performance Cluster Manager operating system releases supported

4. Populate the following file with admin node information:

`/opt/clmgr/etc/hadb.conf`

NOTE: In the `hadb.conf` file, with the exception of the fields marked as optional, make sure to populate each field with a valid value.

Other configurations are possible, but contact your HPE representative before proceeding.

Use one of the following methods to complete this step:

- Method 1 - To populate the quorum HA configuration file during an interactive, question-and-answer session with the installer, complete the following steps:

  a. Open the following file in a text editor, delete any defaults or other prepopulated values, save the file, and close the file:

  `/opt/clmgr/etc/hadb.conf`

  b. Enter the following command and respond to each prompt:

  `# /opt/clmgr/lib/q-ha/setup`

  NOTE: Hewlett Packard Enterprise recommends interactive mode for inexperienced users.

- Method 2 - To populate the quorum HA configuration file by editing the configuration file directly, open the following file in a text editor and save the file after your editing session:

  `/opt/clmgr/etc/hadb.conf`

  NOTE: When editing this file, replace any default values with values appropriate to the cluster.

  Populate the fields in the `hadb.conf` file as follows:

Table 3: Fields in the `hadb.conf` file

Field name	Information to provide
`phys1_hostname`	The hostname that you specified when you installed the operating system on this node. For example: `phys-pub1`.
`phys1_house_ip`	The IP address that you specified when you installed the operating system on this node. Make sure that the network interface that you plan to assign to this IP address is not a bonded interface. The interface cannot be bonded at the time you run the `/opt/clmgr/lib/q-ha/setup` script. To bond the interface after the configuration task is complete, contact your HPE support representative.

Table Continued

Field name	Information to provide
phys2_hostname	The hostname that you specified when you installed the operating system on this node. For example: phys-pub2.
phys2_house_ip	The IP address that you specified when you installed the operating system on this node. Make sure that the network interface that you plan to assign to this IP address is not a bonded interface. The interface cannot be bonded at the time you run the /opt/clmgr/lib/q-ha/setup script. To bond the interface after the configuration task is complete, contact your HPE support representative.
phys3_hostname	The hostname that you specified when you installed the operating system on this node. For example: phys-pub3.
phys3_house_ip	The IP address that you specified when you installed the operating system on this node. Make sure that the network interface that you plan to assign to this IP address is not a bonded interface. The interface cannot be bonded at the time you run the /opt/clmgr/lib/q-ha/setup script. To bond the interface after the configuration task is complete, contact your HPE support representative.
phys1_head_ip	By default, this is set to 172.23.255.150. This is the IP address to be used on physical node 1 for the management network.
phys2_head_ip	By default, this is set to 172.23.255.151. This is the IP address to be used on physical node 2 for the management network.
phys3_head_ip	By default, this is set to 172.23.255.152. This is the IP address to be used on physical node 3 for the management network.
head_netmask	By default, this is set to 255.255.0.0.

Table Continued

Field name	Information to provide
predictable_network_support	By default, this is set to yes.
phys1_mgmt_nic1_ifname	Log into physical node 1, and enter the following command to retrieve this information: ip addr show This is the first NIC to be placed in bond0 on physical node 1.
phys1_mgmt_nic2_ifname	Log into physical node 1, and enter the following command to retrieve this information: ip addr show This is the second NIC to be placed in bond0 on physical node 1.
phys2_mgmt_nic1_ifname	Log into physical node 2, and enter the following command to retrieve this information: ip addr show This is the first NIC to be placed in bond0 on physical node 2.
phys2_mgmt_nic2_ifname	Log into physical node 2, and enter the following command to retrieve this information: ip addr show This is the second NIC to be placed in bond0 on physical node 2.
phys3_mgmt_nic1_ifname	Log into physical node 3, and enter the following command to retrieve this information: ip addr show This is the first NIC to be placed in bond0 on physical node 3.
phys3_mgmt_nic2_ifname	Log into physical node 3, and enter the following command to retrieve this information: ip addr show This is the second NIC to be placed in bond0 on physical node 3.
phys1_bmc_ip	Specify the management card IP address. Contact your network administrator regarding administrative and security requirements.

Table Continued

Field name	Information to provide
`phys1_bmc_user`	Specify the management card username. Contact your network administrator regarding administrative and security requirements.
`phys1_bmc_password`	Specify the management card password. Contact your network administrator regarding administrative and security requirements.
`phys1_bmc_hostname`	Specify the management card hostname. Contact your network administrator regarding administrative and security requirements. For example: `phys-admin1-bmc`.
`phys2_bmc_ip`	Specify the management card IP address. Contact your network administrator regarding administrative and security requirements.
`phys2_bmc_user`	Specify the management card username. Contact your network administrator regarding administrative and security requirements.
`phys2_bmc_password`	Specify the management card password. Contact your network administrator regarding administrative and security requirements.
`phys2_bmc_hostname`	Specify the management card hostname. Contact your network administrator regarding administrative and security requirements. For example: `phys-admin2-bmc`.
`phys3_bmc_ip`	Specify the management card IP address. Contact your network administrator regarding administrative and security requirements.
`phys3_bmc_user`	Specify the management card username. Contact your network administrator regarding administrative and security requirements.
`phys3_bmc_password`	Specify the management card password. Contact your network administrator regarding administrative and security requirements.
`phys3_bmc_hostname`	Specify the management card hostname. Contact your network administrator regarding administrative and security requirements. For example: `phys-admin3-bmc`.
`skip_firewall`	By default, this is set to `no`.
`phys1_head_hostname`	Specify a hostname that resolves the physical admin node on the management network. For example: `phys-admin1-head`.

Table Continued

Field name	Information to provide
phys2_head_hostname	Specify a hostname that resolves the physical admin node on the management network. For example: phys-admin2-head.
phys3_head_hostname	Specify a hostname that resolves the physical admin node on the management network. For example: phys-admin3-head.
admin_iso_path	Specify the path to the installation .iso file. The format is as follows: admin_iso_path=/var/opt/sgi/cm-admin-install-*version*-*op_sys*-x86_64.iso For example: admin_iso_path=/var/opt/sgi/cm-admin-install-1.10-sles15spX-x86_64.iso
preconfigured_house_bond_network	Specifies whether or not you created a house (or site) network with a bonded interface. Note the following: ◦ If you specify yes in this field, also specify the bonded network name in the house_bond_network_name field. ◦ If you specify no in this field, the configuration proceeds as if there were no bonded house network. By default, this is set to no.
house_bond_network_name	Specifies the name of the bonded house (or site) network. Specify a name in this field if you specified preconfigured_house_bond_network=yes. The name bond0 is reserved for internal use. Do not use the name bond0.
configurable_reboot	Specifies whether or not you can reboot the physical nodes manually. You might want to do this as part of the configuration process in order to check the network configuration on the physical nodes. Note the following: ◦ If you specify yes in this field, you can enter a reboot command to reboot the nodes manually. After the reboot, you can resume editing the hadb.conf file. ◦ If you specify no in this field, reboots during the configuration process are not supported. By default, this is set to no.

Table Continued

Field name	Information to provide
`sparse_vmimage`	Specifies whether or not the installer creates a sparse image for the admin node virtual machine (VM). Note the following: ◦ If you specify `yes` in this field, the installer creates the admin node VM image as a sparse file. The installer can create a sparse VM image file more quickly compared to when `sparse_vmimage=no`. If you specify `yes`, take care to not write any files in the admin node VM space because the cluster manager deletes them automatically. ◦ If you specify `no`, the installer does not create the admin node VM as a sparse file. On clusters with large disks, it can take over an hour for the installer to create the VM image when this field is set to `no`. By default, this is set to `no`.
`separate_bmc_network`	Specifies whether or not to create a dedicated BMC network. Note the following: ◦ If you specify `yes` in this field, specify custom values for all the remaining fields in this file. ◦ If you specify `no` in this field, and if a field has a default value, you can accept the default value. By default, this is set to `no`.
`phys1_bmc_network_ifname`	The network interface card (NIC) name for the dedicated BMC network on physical node 1. Log into physical node 1, and enter the following command to retrieve this information: `ip addr show`
`phys2_bmc_network_ifname`	The NIC name for the dedicated BMC network on physical node 2. Log into physical node 2, and enter the following command to retrieve this information: `ip addr show`
`phys3_bmc_network_ifname`	The NIC name for the dedicated BMC network on physical node 3. Log into physical node 3, and enter the following command to retrieve this information: `ip addr show`
`phys1_bmc_bridge_ip`	By default, this is set to `172.24.255.150`.

Table Continued

Field name	Information to provide
phys2_bmc_bridge_ip	By default, this is set to 172.24.255.151.
phys3_bmc_bridge_ip	By default, this is set to 172.24.255.152.
bmc_bridge_nw_netmask	By default, this is set to 255.255.0.0. Do not change this value.

5. Enter the following command and respond to the prompts:

   ```
   # /opt/clmgr/lib/q-ha/setup
   ```

   This step validates the quorum HA configuration information you supplied and creates the HA admin node.

6. Use the ssh command to log into the third physical admin node, and enter the following command:

   ```
   # virsh console adminvm
   ```

7. Use one of the following procedures to install an operating system on the virtual machine:

   - **Configuring RHEL 8 on the admin node**
   - **Configuring SLES 15 on the admin node**

8. (Optional) Enter the following command to monitor the configuration on physical admin node 2 and to make sure that the virt resource started:

```
# crm_mon
Cluster Summary:
  * Stack: corosync
  * Current DC: nano-3 (version 2.0.5+20201202.ba59be712-2.30-2.0.5+20201202.ba59be712) - partition with quorum
  * Last updated: Mon Oct 18 08:20:10 20XX
  * Last change:  Mon Sep 27 15:02:54 20XX by root via crm_resource on nano-2
  * 3 nodes configured
  * 4 resource instances configured

Node List:
  * Online: [ nano-1 nano-2 nano-3 ]

Active Resources:
  * p_ipmi_fencing_1    (stonith:external/ipmi):         Started nano-3
  * p_ipmi_fencing_2    (stonith:external/ipmi):         Started nano-1
  * p_ipmi_fencing_3    (stonith:external/ipmi):         Started nano-2
  * adminvm     (ocf::heartbeat:VirtualDomain):  Started nano-3
```

(Optional) Configuring a system admin controller high availability (SAC HA) admin node

About this task

A SAC HA admin node requires two physical admin nodes that use the x86_64 architecture.

When the cluster is running, the admin node resides in a virtual machine upon one of two physical admin nodes. When a failover occurs, the virtual machine passes from the active node to the passive node.

When you create a SAC HA admin node, you install the cluster manager software, operating system software, and supporting software on two physical admin nodes. After the installation and configuration is complete, the admin node operates within a virtual machine that can reside on either of the two physical hosts.

The following procedures explain how to configure a SAC HA admin node:

Procedure

1. **Creating and installing the high availability (HA) software repositories on the physical admin nodes**
2. **Preparing to run the HA admin node configuration script**
3. **Running the high availability (HA) admin node configuration script**
4. **Starting the HA virtual manager and installing the cluster manager on the virtual machine**

Creating and installing the high availability (HA) software repositories on the physical admin nodes

About this task

The following procedure explains how to install the software repositories on each node.

Procedure

1. Use the `ssh` command to log into one of the physical admin nodes.
2. Copy the installation files (the operating system `.iso` files) to `/var/opt/sgi` on the node.

   Each operating system requires at least one operating system `.iso` file, and some operating systems require additional `.iso` files. Review the tables in the following topic to make sure that you have all the software you need:

   HPE Performance Cluster Manager operating system releases supported
3. Use the `ssh` command to log into the other admin node.

   When prompted, provide the root user login and password credentials.
4. Use the `rsync` command to copy the files from this admin node to the other admin node.

   For example, assume that you used `ssh` to log into a node named `admin2`. To copy the files from the node named `admin1` to the node named `admin2`, enter the following command:

   ```
   # rsync -avz admin1:/var/opt/sgi/*.iso /var/opt/sgi/
   ```
5. Set the path to the `.iso` file for the admin node.

You need this information for the `admin_iso_path=` variable in the `sac-ha-initial-setup.conf` file.

Enter the following commands:

```
# mkdir /root/sw
# ssh phys_admin2
# mkdir /root/sw
# scp host_system:/path/cm-admin-install-1.10-os-x86_64.iso /root/sw
# rsync -avz /root/sw/ phys_admin:/root/sw/
# exit
```

The variables are as follows:

Variable	Specification
host_system	The name of the node that currently hosts the `.iso` file.
path	The path to the `.iso` file on the host node.
os	The name of the operating system.

For example, if you downloaded the `.iso` file to a Linux laptop, the `scp` command might look as follows:

```
# scp user1@desktop:/home/user1/iso/\
cm-admin-install-1.10-rhel8X-x86_64.iso /root/sw/
```

Preparing to run the HA admin node configuration script

About this task

The configuration setup script configures the two physical nodes to communicate with each other and the storage unit. Edit this script and provide information within the script before you run the script.

The following procedure explains how to edit the setup script and provide the information that the script requires.

Procedure

1. Decide which node you want to designate as physical node 1 and physical node 2.
2. Log into each of the physical nodes as the root user.

   Each physical node sees itself as the primary physical node. Each physical node sees the other node as the secondary physical node.
3. On physical node 1, enter the `ip addr` command.

   The command displays NIC and MAC addresses. For example:

```
linux:~ # ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: eno1: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether ec:eb:b8:89:f2:90 brd ff:ff:ff:ff:ff:ff
```

```
3: eno2: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether ec:eb:b8:89:f2:91 brd ff:ff:ff:ff:ff:ff
4: eno3: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether ec:eb:b8:89:f2:92 brd ff:ff:ff:ff:ff:ff
5: eno5: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether 48:df:37:66:c1:30 brd ff:ff:ff:ff:ff:ff
6: eno4: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether ec:eb:b8:89:f2:93 brd ff:ff:ff:ff:ff:ff
7: eno6: <BROADCAST,MULTICAST> mtu 1500 qdisc noop state DOWN group default qlen 1000
    link/ether 48:df:37:66:c1:38 brd ff:ff:ff:ff:ff:ff
linux:~ #
```

The interface names and the MAC addresses are highlighted in **bold** in the preceding output. In a subsequent step, you need this bolded information to specify the following:

- The physical MAC addresses for physical node 1
- Whether this node uses predictable network names

4. On physical node 2, enter the `ip addr` command.

   Again, you need the interface names and the MAC addresses in the command output to specify the following:

   - The physical MAC addresses for physical node 2
   - Whether this node uses predictable network names

5. On physical node 1, use a text editor to open the following file:

   `/etc/opt/sgi/sac-ha-initial-setup.conf`

   The installer automatically copies `sac-ha-initial-setup.conf` to `sac-ha-initial-setup.conf.example`. If you edit the file but subsequently discard your edits, reinstate the original file and remove the `.example` suffix.

6. On physical node 1, complete the lines in the `sac-ha-initial-setup.conf` file that the software requires to be edited for this cluster.

   This file pertains to the two physical nodes for this HA admin node.

   The cluster configuration file contains several lines that end in =" ". Some of these lines contain default settings that you must assess for your site. For the other lines that end in =" ", specify information for your HA cluster. The file contains comments that provide guidance regarding how to complete each line. **Table 4: Configuration file inputs** shows the lines that you must edit within the file.

   NOTE: The `sac-ha-initial-setup.conf` file contains many fields. You do not have to populate all the fields with information from your cluster.

   Table 4: Configuration file inputs contains information about the fields that you must edit. Do not edit the other fields.

Table 4: Configuration file inputs

Configuration file line	Specification
Physical node 1 - MAC addresses:	
`phys1_eth0=""`	The MAC address shown in the output you retrieved from the command in Step **3**.
`phys1_eth1=""`	The MAC address shown in the output you retrieved from the command in Step **3**.
`phys1_eth2=""`	The MAC address shown in the output you retrieved from the command in Step **3**.
`phys1_eth3=""`	The MAC address shown in the output you retrieved from the command in Step **3**.
Physical node 2 - MAC addresses:	
`phys2_eth0=""`	The MAC address shown in the output you retrieved from the command in Step **4**.
`phys2_eth1=""`	The MAC address shown in the output you retrieved from the command in Step **4**.
`phys2_eth2=""`	The MAC address shown in the output you retrieved from the command in Step **4**.
`phys2_eth3=""`	The MAC address shown in the output you retrieved from the command in Step **4**.
Predictable network names for the physical nodes:	

Table Continued

Configuration file line	Specification
`predictable_network_support="yes"`	If physical node 1 and physical node 2 use predictable network names, do not edit this line. For guidance, refer to the output from the following steps earlier in this procedure: • Step **3** • Step **4** The nodes use predictable network names if the output shows interfaces with `enoX`. In this case, do not edit this line. That is, retain `predictable_network_support="yes"` in the file. The nodes do not use predictable network names if the output shows interfaces with `ethX`. In this case, edit this line to appear as follows: `predictable_network_support="no"`
`phys1_initial_house="eth0"`	If physical node 1 uses predictable network names, edit this line, and set this value to `eno1`. If physical node 1 does not use predictable network names, do not edit this line. **NOTE:** The value `eno1` includes a lowercase `o`, not a zero (`0`).
`phys2_initial_house="eth0"`	If physical node 2 uses predictable network names, edit this line, and set this value to `eno1`. If physical node 2 does not use predictable network names, do not edit this line. **NOTE:** The value `eno1` includes a lowercase `o`, not a zero (`0`).
Physical node 2 - hostname and IP address:	
`phys2_house_hostname=""`	Obtain this hostname from your network administrator. Specify the hostname. Do not specify the FQDN.

Table Continued

Configuration file line	Specification
`phys2_house_ip=""`	Obtain this IP address from your network administrator. This is the IP address on your house network for access to the second physical admin node.

`bonding_method="active-backup"`
Or
`bonding_method="802.3ad"`
	Bonding method to use for the `bond0` interface. Valid choices are `active-backup` or `802.3ad`.

Physical node 1 - Node controller hostname and node controller IP address:	
`phys1_bmc_hostname=""`	Obtain the node controller hostname for physical node 1 from your network administrator.
`phys1_bmc_ipaddr=""`	Obtain the node controller IP address for physical node 1 from your network administrator.
Physical node 2 - Node controller hostname and node controller IP address:	
`phys2_bmc_hostname=""`	Obtain the node controller hostname for physical node 2 from your network administrator.
`phys2_bmc_ipaddr=""`	Obtain the node controller IP address for physical node 2 from your network administrator.
Additional information:	
`wwn=""`	Specify the information you retrieved in the following procedure: **(Conditional) Configuring the storage unit** The value you need consists of a number string. For example: `wwn="60080e5000233c340000039f4d90ab57"`
`volume_group_ID=""`	The number of the LUN on the storage device. Typically, this value is `0`. For information about how to derive this number, see your storage unit documentation.

Table Continued

Configuration file line	Specification
`admin_iso_path=""`	The full path to the installation `.iso` file. Specify the path that you configured in the following procedure: **Creating and installing the high availability (HA) software repositories on the physical admin nodes**
`cpus=""`	The number of virtual CPUs assigned to the virtual machine admin node that manages the cluster. The maximum number is *max-cpu_threads* - 4. The following command retrieves CPU information for the `cpus=` field in the configuration file: `# less /proc/cpuinfo` In the output, look for the values for the following fields: • `processor` • `cpu cores` To arrive at the correct specification for the `cpus=` field for your cluster, use the information in these fields. The comments in the file also contain information about how to specify this field. For example, set `cpus="6"` or `cpus="8"`.

`memory=""`	The amount of memory allocated to the virtual machine admin node that manages the cluster.
	
	The following command retrieves information for the `memory=` field in the configuration file:
	
	`# free -h`
	`total . . .`
	`Mem: 62G . . .`
	`Swap: 2.0G . . .`
	
	In the output, observe the values under the `total` column for the `Mem` field and the `Swap` field.
	
	Typically, you can allocate 4GB per virtual CPU that you specified on the `cpus=` line. The comments in the file also contain information about how to specify this field.
	
	For example, if you specified `cpus="6"`, specify `memory="24GB"`. If you specified `cpus="8"`, specify `memory="32GB"`.

Table Continued

Configuration file line	Specification

`rootsize=""`	Specify the amount of shared disk space on the storage unit that you want to allocate to the admin node virtual machine.
	
	The `fdisk` command retrieves information for the `rootsize=` field in the configuration file. This command has the following format:
	
	`fdisk -l` *`disk`*
	
	For *disk*, specify the disk identifier for the shared disk.
	
	For example:
	
	`# fdisk -l /dev/sdb`
	`Disk /dev/sdb: 4000.0 GB, . . .`
	`.`
	`.`
	`.`
	
	The `rootsize=` field requires a value in MB and recommends that you specify a value that is 80% of the LUN size. The minimum size is 94GB.
	
	For this example output, calculate 4000 X 1024 X 0.8, which yields 3276800. For the configuration file, specify `rootsize=3276800`.

`mail_to="root@localhost"`	Email address of the `root` user.

NOTE: At this time, edit only the fields shown in the preceding table. The comments in the `sac-ha-initial-setup.conf` file describe other fields that you can set in a troubleshooting situation.

7. Save and close the `sac-ha-initial-setup.conf` file on physical node 1.

Running the high availability (HA) admin node configuration script

About this task

The configuration script configures the two physical admin nodes to work together, but the script does not configure the shared storage.

When you run the configuration script, one or more steps might fail. If a step fails, you can stop the script, correct the problem, and restart the script.

NOTE: For information about HA admin node configuration troubleshooting, see the following:

Troubleshooting an HA admin node configuration

Procedure

1. Log into the first (or primary) physical admin node as the root user.
2. Enter the following command to navigate to the root directory:

   ```
   # cd ~
   ```

3. Enter the following command to configure the HA system:

   ```
   # sac-ha-initial-setup
   ```

   To obtain help information for the `sac-ha-initial-setup.conf` script, enter the following command:

   ```
   # sac-ha-initial-setup --help
   ```

4. As the script prompts, enter the following command on each physical admin node to reboot the nodes:

   ```
   # reboot
   ```

   Wait for the boot to complete.

5. Log back into both of physical admin nodes as the root user.
6. On each of the physical admin nodes, use the `df` command to verify that the `/images` file system is mounted.

   For example:

   ```
   # df -h /images
   Filesystem                              Size  Used  Avail  Use% Mounted on
   /dev/mapper/sac-ha-vol3_mpath-part2  3.7T  146G  3.5T   4%   /images
   ```

7. Run the configuration script again on physical admin node 1.

   For example, enter the following on physical admin node 1:

   ```
   # sac-ha-initial-setup
   ```

8. (Optional) Enter the following command to monitor the configuration on physical admin node 2 and to make sure that the `virt` resource started:

   ```
   # crm_mon
   Cluster Summary:
     * Stack: corosync
     * Current DC: hikari2 (version 2.0.5+20201202.ba59be712-2.30-2.0.5+20201202.ba59be712) - partition with quorum
     * Last updated: Wed Oct 20 10:15:39 20XX
     * Last change:  Fri Sep 24 12:25:17 20XX by hacluster via crmd on hikari
     * 2 nodes configured
     * 9 resource instances configured

   Node List:
     * Online: [ hikari hikari2 ]

   Active Resources:
     * p_ipmi_fencing_1    (stonith:external/ipmi):         Started hikari2
     * p_ipmi_fencing_2    (stonith:external/ipmi):         Started hikari
     * sbd_stonith (stonith:external/sbd):  Started hikari2
     * Clone Set: dlm-o2cb-fs-images-clone [dlm-o2cb-fs-images-group]:
       * Started: [ hikari hikari2 ]
     * virt        (ocf::heartbeat:VirtualDomain):  Started hikari2
     * mailTo      (ocf::heartbeat:MailTo):         Started hikari2
   ```

9. Verify that the admin node image was created with the size you specified.

   For example:

   ```
   # ls -lh /images/vms/
   total 135G
   -rw------- 1 qemu qemu 1.0T Sep 24 12:20 sac.img
   ```

Starting the HA virtual manager and installing the cluster manager on the virtual machine

About this task

The following procedure explains how to bring up the virtual machine manager. The HA installation and configuration process creates a virtual machine. One of the two physical admin nodes hosts the virtual machine at any given time.

Procedure

1. Log into physical node 1.

   None of the previous HA configuration steps required a graphics terminal. You could complete all the previous steps from a text-based terminal. Starting with this procedure, however, you are required to log in from a graphics terminal in one of the following ways:

   - Log into the physical console. Make sure that the cluster is booted at run level 5 (`init 5`).

     Or

   - Log in from a remote terminal through an `ssh` session with X11 forwarding. For example:

     ```
     # ssh -C -XY root@physical_node1_addr
     ```

     For *physical_node1_addr*, specify the IP address or hostname of physical node 1.

     Or

   - Log in through a VNC session. Make sure that the cluster is booted at run level 5 (`init 5`).

2. On physical node 1, enter the following command:

   ```
   # virt-manager &
   ```

3. In the **Virtual Machine Manager** window, click **File > Add Connection**.

4. On the **Add Connection** popup, complete the following steps:

   a. Click the **Connect to remote host** box.

   b. In the **Hostname** field, enter the hostname of physical node 2.

   c. Click the **Autoconnect** box.

   d. Click **Connect**.

5. In the **Virtual Machine Manager** window, double click the **sac running** icon.

   The window that appears is the interface to the virtual machine that runs on the physical nodes. This window is the interface to the admin node that you can use for system administration tasks.

6. Use the arrow keys to select **Install: Wipe Out and Start Over: Prompted**, and press Enter.

7. At the **Enter number of slots to allow space for: (1-10):** prompt, enter the integer number of slots you want on this cluster, and press Enter.

8. At the **Enter which slot to install to: (1-10):** prompt, enter the integer number that corresponds to the slot you want to install, and press Enter.

9. At the **Destructively bypass sanity checks? (y/n):** prompt, enter **y**, and press Enter.

10. At the **Is this a physical admin node in an SAC-HA configuration? (normally no) (y/n):** prompt, enter **n**, and press Enter.

11. At the **Use predictable network names for the admin node? (normally yes) (y/n):** prompt, enter **y**, and press Enter.

    For information about predictable network names, see the following:

    Predictable network interface card (NIC) names

12. At the **Additional parameters (like console=, etc):** prompt, enter `console=ttyS0,115200n8`, and press Enter.

    This step enables you to log into the virtual admin node from one of the physical admin nodes. For more information about how to log into the virtual admin node, see the following:

    Connecting to the virtual admin node in a cluster with a high availability (HA) admin node

13. At the **OK to Proceed? (y/n):** prompt, enter **y**, and press Enter.

14. Wait for the software to install on the virtual machine.

15. Configure an operating system and network for the virtual machine.

    Use the graphical connection to the virtual machine, and complete one of the following procedures:

    - **Configuring RHEL 8 on the admin node**
    - **Configuring SLES 15 on the admin node**

    After this step is complete, there is no need to log into the physical hosts. Also, there is no need to use the `virt-manager` tool. To connect to the HA admin node, use one of the following commands to log into the virtual machine:

    - `ssh -C -XY root@admin_vm_addr`

      Or

    - `ssh root@admin_vm_addr`

    For *admin_vm_addr*, enter the admin node virtual machine IP address or hostname.

Configuring the cluster software on the admin node

Procedure

1. **Preparing to configure the cluster software on the admin node**
2. **(Optional) Configuring the management network manually**
3. Configuring the cluster software on the admin node. Use one of the following methods:
   - **Using the cluster definition file to specify the cluster configuration**

     Or
   - **Using the menu-driven cluster configuration tool to specify the cluster configuration**
4. **Completing the admin node software installation**
5. **(Conditional) Configuring an unsupported Ethernet switch into the cluster**
6. **(Conditional) Renaming the HPE Slingshot interconnect hostname to have an `hsn` prefix**

Preparing to configure the cluster software on the admin node

Procedure

1. Locate the cluster manager USB device, or verify the path to the online software repository at your site.

   You can configure the software from either physical media or from an ISO on your network.
2. From a graphics screen or through an `ssh` connection, log into the admin node as the root user., as follows:

   This step differs depending on whether your admin node is a single node or is a two-node SAC HA admin node, as follows:

   - For a single-node admin node, Hewlett Packard Enterprise recommends that you run the cluster configuration tool as follows:
     - From the graphics screen

       Or
     - From an `ssh` session to the admin node

     Avoid running the `configure-cluster` command from a serial console.
   - For an HA admin node, create an `ssh` connection to the host that is running the `virt` resource, and enter the `virt-viewer` command. For example:

     ```
     # ssh -C -XY root@phys_admin1
     # virt-viewer
     ```

     If the Virtual Machine Manager interface does not appear, log into the physical node, and enter `virt-viewer sac` on that host.
3. (Conditional) Open the ports that the cluster manager requires.

   Complete this step if you configured a firewall on the admin node or anywhere else in the cluster.

To avoid monitoring failures, do not permit other software to use the cluster manager ports.

The following table shows the outward-facing TCP ports that are required to be open on the admin node firewall during installation.

Service	Port(s)
External port for SSH	TCP 22
External ports required for webpage and GUI. You can start the cluster manager web server on a different port. For more information, see the following: **Starting the cluster manager web server on a non-default port**	TCP 80, 443, 1099, and 49150
HPE Performance Cluster Manager REST API	8080

For more information about port requirements, see the following:

- `/opt/clmgr/etc/cmuserver.conf`
- **HPE Performance Cluster Manager Administration Guide**

(Optional) Configuring the management network manually

The cluster manager installation process typically includes using a cluster definition file or running the `configure-cluster` command to configure the cluster management networks. As an alternative, you can use operating system commands to configure the management network manually. If you configure the management network manually, you can still use a cluster definition file or run the `configure-cluster` command to configure the rest of the cluster.

If you choose to configure the management network manually, observe the following requirements:

- Configure `bond0` with at least one IP address.
- To manage the node controllers, configure `bond0` with an additional IP address from the `head-bmc` management network. This is often noted as `bond0:bmc`.
- Combine IPV4 and IPV6 routes if needed.
- In the cluster definition file, you can set the management network interface with the following configuration attribute:

  `admin_mgmt_interfaces=existing`

  For example:

```
.
.
.
[attributes]
admin_house_interface=eno1
admin_mgmt_interfaces="existing"
admin_mgmt_bmc_interfaces="existing"
.
.
.
```

In the cluster definition file, you can also set the node controller network interfaces with the following configuration attribute:

```
admin_mgmt_bmc_interfaces=existing
```

- Make sure the cluster definition file describes the management network you configured. Failure to do so can produce unexpected results.

If you want to run the `configure-cluster` command after you configure the management network, navigate to the **Management Network Interfaces Selection** menu. That menu lets you specify network interfaces for `bond0` and lets you specify **Use existing Settings for Management**. Alternatively, make sure that the cluster definition file is complete, and supply the name of the cluster definition file as input to the `configure-cluster` command.

Using the cluster definition file to specify the cluster configuration

Prerequisites

This method assumes that you have a cluster definition for the cluster.

Procedure

1. Use the `cm repo add` command, in the following format, to create a repository for the installation package:

   `cm repo add` *`path_to_iso`*

   For *path_to_iso*, specify the full path to installation ISO.

   If you have a USB device mounted in the admin node USB port, specify the path to that USB port. If operating system and cluster manager software reside in an ISO file on your network, specify the path to the files on your network.

   For example, enter the following commands to add a repository for a SLES ISO file that is required for SLES platforms and verify the repositories:

   ```
   # cm repo add /tmp/SLE-15-SPX-Full-x86_64-GM-Media1.iso
   # cm repo show
   ```

2. (Optional) Add updates and patches for the operating system software and for the cluster manager.

   Complete this step if updates are available and you want to update the software at this time.

   Cluster manager patch names and distribution update names can vary from these examples. The examples in this step add the repository as a custom repository and then select the patches. These commands assume the following:

   - The packages updates are at the following location:

     ```
     /opt/clmgr/repos/SLES15-SPX-Updates-x86_64
     ```

   - The cluster manager patches are in the following location:

     ```
     /opt/clmgr/repos/patch11627-x86_64
     ```

   Example 1. To add cluster manager `patch11627` on an x86_64 admin node, run the following commands:

   ```
   # cm repo add /opt/clmgr/repos/patch11627-x86_64 --custom patch11627-x86_64
   # cm repo select patch11627-x86_64
   ```

   Example 2. To add SLES 15 update repos, run the following commands:

```
# cm repo add /opt/clmgr/repos/SLES15-SPX-Updates-x86_64 --custom SLES15-SPX-Updates-x86_64
# cm repo select SLES15-SPX-Updates-x86_64
```

3. Enter the following command to define the cluster according to the content in the cluster definition file:

   ```
   # configure-cluster --configfile path
   ```

   For *path*, specify the path to the configuration file.

Using the menu-driven cluster configuration tool to specify the cluster configuration

About this task

The cluster configuration tool presents you with many default settings. Hewlett Packard Enterprise recommends that you keep the default settings if possible.

NOTE: The `bond0` network interface is reserved. The cluster manager installer creates and configures `bond0` as part of this procedure.

Procedure

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

   ```
   # configure-cluster
   ```

2. On the **House Network Interface Selection** screen, complete the following steps:

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

      Make sure 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.

   b. Click **OK**.

3. On the **Management Network Interfaces Selection** screen, complete the following steps:

   a. Use the space bar and arrow keys to select one or two NICs for the management network.

   b. Click **OK**.

4. 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 click **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 click **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**.

5. 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)**, 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.

NOTE: If you configured a high availability (HA) admin node, select the bonding mode that you configured on the two physical admin nodes.

b. On the **Main Menu** screen, click **OK** to select the **Initial Setup Menu**.

On a configured cluster, you can see the interfaces you specified in the following file:

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

6. On the **Cluster Configuration Tool: Initial Cluster Setup** screen, select **OK** on the screen.

   The message on the screen is as follows:

   ```
   All the steps in the following menu need to be
   completed in order.  Some settings are harder
   to change once the cluster has been deployed.
   ```

7. On the **Initial Cluster Setup Tasks** screen, select **R Repo Manager: Set Up Software Repos**, and click **OK**.

   The next few steps describe how to create repositories for the following:

   - The operating system software for compute nodes and for infrastructure nodes
   - The cluster manager software
   - (Optional) Additional software for HPE Message Passing Interface (MPI), AMD ROCm, SLURM, or other products

   Locate your system disks before you proceed. The menu system prompts you to insert physical media or specify a path for some of the preceding software.

8. On the **One or more ISOs were embedded on the ...** screen, select **Yes**.
9. Wait for the software repositories to configure.
10. At the `press ENTER to continue` prompt, press **Enter**.
11. On the **Would you like to create repos from media? ...** screen, select one of the following:

    - **Yes**. After you select **Yes**, proceed to the following:

      Step **12**

      Or

    - **No**. After you select **No**, proceed to the following:

      Step **14**

12. On the **Please enter the path to the media:** screen, enter the path to the installation media.

    a. Enter the path to the media as follows:

    - If the media is on a local file system, enter the full path to the mount point or the `.iso` file. Select **OK** after entering the path.
    - If the media is on an NFS share, enter the full path to the `.iso` file in *`server_name`*`:/`*`path_name/iso_file`* format. Select **OK** after entering the path.
    - If the media is on a remote server, enter the URL to the expanded media on the remote server. For example, the `.iso` file could be mounted on the loopback device on the remote server. Select **OK** after entering the URL.

    b. On the **Media registered successfully with crepo ...** screen, select **OK**.

    c. On the **Would you like to create repos from media? ...** screen, select **Yes** if you have more software that you to register.

    If you select **Yes**, repeat the preceding tasks in this sequence for the next media path.

    If you select **No**, proceed to the next step.

13. Repeat the following steps until all software is installed:

    - Step **11**
    - Step **12**

    If you plan to configure MPT and run MPT programs, make sure to install the HPE Message Passing Interface (MPI) software.

14. On the **Initial Cluster Setup Tasks** screen, select **I Install and Configure Admin Cluster Software**, and select **OK**.

    This step installs the cluster software that you wrote to the repositories.

15. On the **Initial Cluster Setup Tasks** screen, select **N Network Settings**, and select **OK**.

16. On the **About to create secrets ...** popup window, select **Yes**.

17. On the **Admin node network and database will now be initialized** popup, select **OK**.

18. Create one or more data networks.

    The cluster manager does not automatically create a data network.

    NOTE: The cluster manager requires each defined subnet address to be unique. That is, the head network and the BMC network cannot be the same.

    The following substeps show how to create a data network and an InfiniBand network.

    To configure a data network, complete the following steps:

a. On the **Cluster Network Settings** screen, select **A Add Subnet**, and select **OK**.

b. On the **Select network type** screen, press the space bar to move the asterisk (*) to the second line. This action selects the lower line, which now appears as follows:

```
(*) 4 Data Network
```

c. Select **OK**.

d. On the **Insert network name, subnet and netmask** screen, enter information to define the data network. Use the arrow keys to move from field to field on this screen. Enter the following information:

Field	Information needed
name	A unique name for this network. For example: `data10g`.
subnet	The network IP address (start of the range) for the nodes on the data network.
netmask	Subnet mask for the nodes on the data network.
gateway (optional)	An IP address within the subnet that can be used as a default gateway. (Optional)

e. On the **Network *name* ...** screen, verify the information that you specified for the routed management network, and select **OK**.

To configure an InfiniBand network, for any kind of cluster, complete the following steps:

a. On the **Cluster Network Settings** screen, select **A Add Subnet**, and select **OK**.

b. On the **Select network type** screen, press the space bar to move the asterisk (*) to the second line. This action selects the lower line, which now appears as follows:

```
(*) 5 IB Network
```

c. Select **OK**.

d. On the **Insert network name, subnet and netmask** screen, enter information to define the InfiniBand network. Use the arrow keys to move from field to field on this screen. Enter the following information:

Field	Information needed
name	A unique name for this InfiniBand network. For example: `ib0`.
subnet	The network IP address (start of the range) for the nodes on the data network.

Table Continued

Field	Information needed
netmask	Subnet mask for the nodes on the data network.
gateway (optional)	An IP address within the subnet that can be used as a default gateway. (Optional)

e. On the **Network *name* ...** screen, verify the information that you specified for the routed management network, and select **OK**.

19. On the **Cluster Network Settings** screen, select **S List and Adjust Subnet Addresses**, and select **OK**.

20. Verify the information on the **Caution: You can adjust ...** screen, and click **OK**.

21. Review the settings on the **Subnet Network Addresses - Select Network to Change** screen, and modify these settings only if necessary.

This screen displays the default networks and netmasks that reside within the cluster. Complete one of the following actions:

- To accept the defaults, select **Back**.

  Or

- To change the network settings, complete the following steps:

  a. Highlight the setting you want to change, and select **OK**.

  b. Enter a new subnet IP address, netmask, gateway, or VLAN, and select **OK**.

  c. Press Enter.

For example, it is possible that your site has existing networks or conflicting network requirements. For additional information about the IP address ranges, see the following:

Subnetwork information

On the **Update Subnet Addresses** screen, the **Head Network** field shows the admin node IP address. Hewlett Packard Enterprise recommends that you do not change the IP address of the admin node if at all possible. You can change the IP addresses of the InfiniBand network or the Omni-Path Express network. These networks are named **IB0** and **IB1**. You can change the `IB0` and `IB1` IP addresses to match the IP requirements of the house network, and then select **Back**.

22. On the **Cluster Network Settings** screen, select **D Configure Cluster Domain Name**, and select **OK**.

23. On the **Please enter the domain name for this cluster** pop-up window, enter the domain name, and select **OK**.

The domain you specify becomes a subdomain of your house network.

For example, enter `cm.clusterdomain.com`.

24. On the **Domain name configured** screen, click **OK**.

25. On the **Please adjust the domain_search_path as needed ...** screen, click **OK**.

The default search paths use *head* and *head*`-BMC` networks. You can adjust this as needed after the cluster is configured. For information, see the following:

Adjusting the domain name service (DNS) search order

26. Select **P Domain Search Path** to verify the domain search path.

27. (Optional) On the **Cluster Network Settings** screen, select **U Configure Udpcast Settings**, and select **OK**.

    On the **Udpcast Settings** screen, select one of the following, and select **OK**.

    The selections are as follows:

    - **U Admin Udpcast RDV Multicast Address**
    - **T Admin Udpcast TTL**
    - **G Global Udpcast RDV Multicast Address**

    For each of the preceding selections, enter a value, and click **OK**.

    For information about the actions available from the preceding settings, select the setting. An informational window appears. When finished, click **Back** until you get to the **Cluster Network Settings** screen.

28. On the **Cluster Network Settings** screen, adjust the VLAN settings.

    - If the cluster is configured to use multiple VLANs and it requires L3 routing to achieve end-to-end connectivity, you can adjust the settings. Use the following steps to change the VLAN numbers used by the supported routing protocols. The supported protocols are OSPF and routing information protocol (RIP).

      a. On the **Cluster Network Settings** screen, select **X Configure Management Network Routing Settings**, and click **OK**.

      b. On the **Management Network Routing Settings** screen, select **O OSPF VLAN Settings** , and click **OK**.

      c. On the **Change OSPF VLAN #, Network, Subnet Mask** screen, use the up and down arrows to highlight the field you want to specify, specify a value, and click **OK**.

        - **OSPF VLAN # [2~4094]**
        - **OSPF Base Network [X.X.X.X]**
        - **OSPF Base Netmask [X.X.X.X]**
        - When finished specifying new values, click **OK**.
        - When finished, click **Back**.

      d. On the **Management Network Routing Settings** screen, select **R RIP VLAN Settings**, and select **OK**.

      e. On the **Change RIP VLAN #, Network, Subnet Mask** screen, use the up and down arrows to highlight the field you want to specify, specify a value, and click **OK**.

        - **RIP VLAN # [2~4094]**
        - **RIP Base Network [X.X.X.X]**
        - **RIP Base Netmask [X.X.X.X]**
        - When finished specifying new values, click **OK**.
        - When finished, click **Back**.

      f. When finished, click **Back**.

29. On the **Cluster Network Settings** screen, select **Back**.

30. On the **Initial Cluster Setup Tasks -- all Required** popup, select **S Perform Initial Admin Node Infrastructure Setup**, and select **OK**.

31. On the following screen, select **OK**:

    ```
    A script will now perform the initial cluster
    set up including setting up the database and
    some network settings.
    ```

32. In the **Please enter the Domain Search Path for this cluster** box, verify the information, adjust if needed, and click **OK**.

33. On the **Domain Search Path Configured** screen, click **OK**.

34. On the **Enter up to three DNS resolvers IPs** screen, make adjustments if needed, and select **OK**.

35. On the **Setting DNS Forwarders to X.X.X.X** screen, review the display and take one of the following actions:

    - To change the display, select **No**, and make adjustments if needed.

      Or

    - If the display is correct, select **Yes**.

36. On the **Copy admin ssh configuration ...** screen, take one of the following actions:

    - To change the display, select **No**, and make adjustments if needed.

      Or

    - If the display is correct, select **Yes**.

37. On the **Create which images now?** screen, confirm the images that you want to create.

    The following shows a representation of this screen:

    ```
    Create which images now?
         [ ] default  Default flat compute image (Recommended)
         [ ] su-lead  SU Leader node image (Required for su-leader nodes)
         [ ] lead     ICE Leader node (RLC) image (Required for ICE)
         [ ] ice      ICE compute node image (Required for ICE)
         [ ] none     Skip image creation (only check this box)

                      <  OK  >              < Back >
    ```

    Use the arrow keys and the space bar to select the images that the cluster requires. Remove the asterisk (`*`) character from any unneeded images.

    On a cluster without leader nodes, select `default`.

    When the screen shows the images that you want to create, select **OK**. It can take up to 30 minutes to create the images.

    If you clear any fields, the installer does not create an image for that particular node type. If you do not want the installer to create any images, select **none**.

    Wait for the completion message. The script writes log output to the following log file:

    `/var/log/cinstallman`

38. (Conditional) On the **One or more ISOs were embedded on the admin install media and copied to ...**, screen, select **OK**.

Depending on what you have installed, this screen might not appear.

39. On the **Initial Cluster Setup Complete** screen, select **OK**.

    This action returns you to the cluster configuration tool main menu.

40. On the **Initial Cluster Setup Tasks -- All Required** screen, select **M Configure Switch Management Network**, and click **OK**.

41. On the **Default Switch Management Network setting for newly discovered ...** screen, select **Yes** and select **OK**.

42. On the **Initial Cluster Setup Tasks -- All Required** screen, select **O Configure Monitoring**, and click **OK**.

    The installation process installs and configures monitoring software on the cluster nodes. This step explains how to enable the monitoring software at installation time. You can enable various types of monitoring. By default, monitoring software is installed but not enabled.

    - To enable native monitoring, complete the following steps:

      a. On the **Cluster Monitoring Settings** screen, select **Native Monitoring**, and click **OK**.

      b. On the **Enable native monitoring?** screen, select **Y yes**, and click **OK**.

      c. On the **Native monitoring has been set to enable** screen, click **OK**, and wait while the system configures native monitoring.

      d. On the **Cluster Monitoring Settings** screen, click **Back**.

    - To enable Kafka, OpenSearch, and Alerta monitoring, complete the following steps:

      a. On the **Cluster Monitoring Settings** screen, select **Kafka/ELK/Alerta Monitoring**, and click **OK**.

      b. On the **Enable Kafka/ELK/Alerta Monitoring?** screen, select **Y yes**, click **OK**, and wait while the system configures Kafka, ELK, and Alerta services.

      c. On the **Kafka/ELK/Alerta monitoring has been set to enable** screen, click **OK**.

      d. On the **Cluster Monitoring Settings** screen, click **Back**.

    - To enable system infrastructure monitoring (SIM), complete the following steps:

      a. On the **Cluster Monitoring Settings** screen, select **SIM Monitoring**, and click **OK**.

      b. On the **Enable SIM Monitoring?** screen, select **Y yes**, click **OK**, and wait while the system configures SIM.

      c. On the **SIM monitoring has been set to enable** screen, click **OK**.

      d. On the **Cluster Monitoring Settings** screen, click **Back**.

43. On the **Initial Cluster Setup Tasks -- All Required** screen, select **P Predictable Network Names**, and select **OK**.

44. On the **Default Predictable Network Names ...** popup, select **Yes** or **No**. These selections have the following effect:

    - Select **Yes** and select **OK** to use predictable names on future equipment. For example, if you select **Yes** here, the cluster is configured to add new equipment with predictable names later.

      If the admin node is configured with predictable names, this popup has **Yes** highlighted because that is the cluster-wide default.

Or

- Select **No** and select **OK** to use legacy names on future equipment.

  If the admin node is configured with legacy names, this popup has **No** highlighted because that is the cluster-wide default.

NOTE: Hewlett Packard Enterprise recommends that you do not mix predictable names with legacy names in the same cluster. To change the naming scheme for a cluster component, run the node discovery commands (again) on that component. This action reconfigures the component into the cluster with the alternative naming scheme. For more information about predictable names and legacy names, see the following:

Predictable network interface card (NIC) names

---

45. Select **Back**.
46. Select **Quit**.

Completing the admin node software installation

About this task

The following procedure completes the admin node software installation.

Procedure

1. Enter the `cattr list -g` command and examine the output to verify the features you configured with the cluster configuration tool.

   The `cattr` output differs from cluster to cluster depending on configuration choices and hardware. To respecify any global values, start the cluster configuration tool again, and correct your specifications. To start the cluster configuration tool, enter the following command:

   ```
   # configure-cluster
   ```

2. Correct any aspect of the installation that is incorrect.

   For example, the installation process typically creates a default compute node image for you. If that process fails, the `cattr` output does not display a default image name. In this case, create a default compute node image manually. When the `cmcinventory` service runs during the installation, it searches for a default image with a name that adheres to a specific format. Name the image so that it includes the distribution name, `rhel` or `sles`, plus the operating system distribution release level. For example, `rhel8.X`, `sles15spX`, or any other image name that includes only the distribution name and the release level. At a minimum, include the cluster manager repository and the distribution repository in the default image you create. You can include additional repositories. For more information about how to create an image, see the following:

   HPE Performance Cluster Manager Administration Guide

(Conditional) Allocating IP addresses for physical quorum high availability (HA) admin nodes

About this task

Complete this procedure if the cluster has three physical admin nodes for a quorum HA configuration.

The procedure in this topic allocates the IP addresses used by the physical admin nodes for the private network within the cluster.

Procedure

1. Obtain the following values from the `/opt/clmgr/etc/hadb.conf` file:

   - `phys1_head_ip=`
   - `phys2_head_ip=`
   - `phys3_head_ip=`
   - `phys1_bmc_ip=`
   - `phys2_bmc_ip=`
   - `phys3_bmc_ip=`

2. Determine whether the management controllers for the physical admin nodes are on the house network or on the management network.

   The management controllers are the iLO or BMC devices in each node. These devices can reside on the house network, also called the *site network* or the *public network*, or on the management network. The following steps use the `cm node add` command to reserve IP addresses for both the physical nodes and the management controllers. Include the management controllers on the `cm node add` commands if the management controllers are also on the management network.

3. Open a file in a text editor and create a small cluster definition file.

   For example, create a file called `phys-admins.config` with the following contents:

   ```
   [discover]
   internal_name=service100000, hostname1=phys-admin1-head,
   mgmt_net_name=head, mgmt_net_ip=phys1_head_ip_value,
   mgmt_net_macs=phys1_eth1_value, generic
   internal_name=service100001, hostname1=phys-admin2-head,
   mgmt_net_name=head, mgmt_net_ip=phys2_head_ip_value,
   mgmt_net_macs=phys2_eth1_value, generic
   internal_name=service100002, hostname1=phys-admin3-head,
   mgmt_net_name=head, mgmt_net_ip=phys3_head_ip_value,
   mgmt_net_macs=phys3_eth1_value, generic
   ```

   NOTE: The values of `100000`, `100001`, and `100002` in the preceding file can be any values. The values are the node numbers. For these values, specify a large value that is greater than the number of physical compute nodes you ever expect to have in the cluster.

4. Use a `cm node add` command that adds nodes from the cluster definition file you created.

   For example:

   ```
   # cm node add -c phys-admins.config --skip-switch-config --skip-refresh-netboot
   ```

(Conditional) Allocating IP addresses for physical system admin controller high availability (SAC HA) admin nodes

About this task

Complete this procedure if the cluster has two physical admin nodes for a SAC HA configuration.

The procedure in this topic allocates the IP addresses used by the physical admin nodes for the private network within the cluster.

Procedure

1. Obtain the following values from the `sac-ha-initial-setup.conf` file:

   - `phys1_head_ip=`
   - `phys1_eth1=`
   - `phys2_head_ip=`
   - `phys2_eth1=`

2. Open a file in a text editor and create a small cluster definition file.

   For example, create a file called `phys-admins.config` with the following contents:

   ```
   [discover]
   internal_name=service100000, hostname1=phys-admin1, mgmt_net_name=head,
   mgmt_net_ip=phys1_head_ip_value, mgmt_net_macs=phys1_eth1_value, generic
   internal_name=service100001, hostname1=phys-admin2, mgmt_net_name=head,
   mgmt_net_ip=phys2_head_ip_value, mgmt_net_macs=phys2_eth1_value, generic
   ```

   NOTE: The values of `100000` and `100001` in the preceding file can be any values. The values are the node numbers. For these values, specify a large value that is greater than the number of physical compute nodes you ever expect to have in the cluster.

3. Use a `cm node add` command that adds nodes from the cluster definition file you created.

   For example:

   ```
   # cm node add -c phys-admins.config --skip-switch-config --skip-refresh-netboot
   ```

4. Use a `cat` command to verify these values in the `/etc/hosts` file.

   For example:

   ```
   # cat /etc/hosts | grep phys-admin
   172.23.200.1  phys-admin1.head.cm.cluster.net phys-admin1 service100000
   172.23.200.2  phys-admin2.head.cm.cluster.net phys-admin2 service100001
   ```

(Conditional) Configuring an unsupported Ethernet switch into the cluster

About this task

Complete this procedure if the cluster includes any unsupported Ethernet switches.

The cluster manager supports the Ethernet switches as described in the cluster manager release notes. An advantage to using supported Ethernet switches is that you can use cluster manager tools, such as `switchconfig`, to manage them.

If the cluster includes switches that are not supported, modify the installation procedure according to the steps in this topic. Use commands specific to that switch to complete some configuration steps manually.

Unsupported switches are included in the cluster as unmanaged switches. For these switches, the cluster manager does not attempt to automatically configure any switch settings.

Procedure

1. Enter the following command:

   ```
   # cadmin --enable-discover-skip-switchconfig
   ```

   This command accomplishes the following:

   - It prevents the cluster manager from logging into management switches at a global level.
   - It allows you to configure the unsupported switches later in the installation.

2. Configure the switches for multicast, or configure the cluster manager to use unicast.

   This step ensures that each node receives its image in an efficient manner. Do one of the following:

   - Verify whether the unsupported switch is configured for **IGMP** and **IGMP Snooping**. Configure those two settings if they are not in effect at this time. The cluster manager uses a multicast protocol called UDPcast to image compute nodes during the boot process. For multicast to be successful, the management switches must support IGMP and IGMP Snooping. For information, see the switch configuration documentation.

     Or

   - Configure the cluster manager to use BitTorrent when it images the compute nodes. BitTorrent is not a multicast method. It is unicast.

     For information about how to change the method by which the compute nodes receive images, see the following:

     Node provisioning takes too long or fails to complete

3. (Optional) Create entries for the unsupported switch in the cluster definition file.

   When switch entries appear in the cluster definition file, the admin node assigns an IP address to a DHCP request from the switch. These entries also enable the admin node to match a static IP address for the switch to the hostname for the switch.

   For an example entry, see the following:

   Cluster definition file example - Entries for an unsupported switch

NOTE: After the cluster manager installation is complete, consider one of the following:

- Enabling DHCP on the unsupported switch
- Configuring a static IP address on the unsupported switch

For information, see the documentation for the unsupported switch. DHCP enables the cluster manager to assign an IP address to the switch. To manage these switches remotely, do the following for the switch:

- Enable either Telnet or SSH.
- Create a remote username and strong password.

Because you ran the `cadmin --enable-discover-skip-switchconfig` command before you run the node discovery commands, DHCP assigns supported switches an IP address. In this way, you can use the `ssh` command or Telnet to connect to the supported switches if necessary. Assigning a static IP achieves the same outcome. That is, the management switch has an entry in `/etc/hosts`, but the cluster manager does not remotely log into the switch automatically.

(Conditional) Renaming the HPE Slingshot interconnect hostname to have an `hsn` prefix

About this task

Complete the procedure in this topic if the cluster nodes use Mellanox network interface cards (NICs), which are generally found on HPE Slingshot 10 systems.

The post-install script that resides in the following directory generates userspace device manager rules:

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

The rules themselves reside in the following file on the node:

`/usr/lib/udev/rules.d/94-cm-slingshot.rules`

By default, the script generates rules for HPE Slingshot 200Gbps NIC devices found on the system. The procedure in this topic explains how to modify the post-install script to search for the Mellanox NICs.

Procedure

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

   `/opt/clmgr/image/scripts/post-install/50all.create_hsn_udev`
3. Search for `DEV_TYPE`, and replace the current value with `mellanox`.
4. Save and close the file.
5. (Conditional) Synchronize the script to shared storage.

   Complete this step if the cluster has scalable unit (SU) leader nodes.

   Enter the following command:

   ```
   # cm image sync --scripts
   ```

(Optional) Configuring software RAID on cluster nodes

You can use the cluster manager to configure standard software RAID levels for admin nodes, leader nodes, and compute nodes. The RAID levels include 0, 1, 4, 5, 6, and 10.

Each RAID scheme has its own distinct metadata. For any given RAID level, the RAID can be one of the following types:

- BIOS-assisted software RAID metadata
- Native metadata

The cluster manager supports MD RAID and BIOS-assisted software RAID.

If you want to configure nodes to boot from disk with no network or miniroot help, then your options are limited to either BIOS assisted SW RAID, at any RAID level, or MD metadata with RAID 1 only.

In general, configure RAID for a system disk as a one-time action for the life of the hardware.

NOTE: At this point in the installation process, the admin node is configured. To change the admin node RAID configuration during a reinstallation, use one of the following procedures:

- **(Optional) Configuring BIOS-assisted RAID (BAR) on the `root` partition of a leader node or a compute node**
- **(Optional) Configuring software MD RAID 1 on the `root` partition of a leader node or a compute node**
- **(Optional) Configuring software MD RAID on leader nodes and on compute nodes**

(Optional) Configuring BIOS-assisted RAID (BAR) on the `root` partition of a leader node or a compute node

About this task

The procedure in this topic uses the cluster manager to configure BAR on a leader node or on a compute node.

NOTE: At this point in the installation process, the admin node is configured. To change the admin node RAID configuration during a reinstallation, complete this procedure before you install an operating system on the admin node.

Procedure

1. Log into the admin node as the root user.
2. Determine which leader nodes and compute nodes you want to configure with BAR.
3. Configure the BIOS on the leader nodes or compute nodes for software RAID.

   For example, change the mode from `AHCI` to `RAID`.
4. Add the following parameters to the list of parameters in the cluster definition file:

   - `md_metadata=imsm`
   - `md_raidlevel=`*`n`*

For *n*, specify an integer that represents the RAID level you want. The default is 1.

- (Optional) `force_disk="`*`device1`*`,`*`device2`*`,...,`*`deviceN`*`"`

  By default, the cluster manager configures RAID 1 on the first two empty disks.

  If you specify this parameter, specify the path to each disk device required for the RAID level you choose. For example:

  ```
  force_disk="/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:1,/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:2 "
  ```

  NOTE: Disk identifiers specified in the `force_disk` field, such as `/dev/sda`, are used once and are automatically assembled by MD thereafter. These identifiers are not persistent. They could unexpectedly change from what you assumed.

  Even though the identifiers are used only once, Hewlett Packard Enterprise recommends that you use `/dev/disk/by-id` or `/dev/disk/by-path` names for the disks instead. In this way, the following occur:

  - The cluster manager operates on the exact disks you target.
  - You do not rely on device names that might not point where you expect.

5. Verify the new RAID volume.

   To verify that the new RAID volume is the new boot device, stop at BIOS on a reboot and navigate to the **Boot** menu.

(Optional) Configuring software MD RAID 1 on the `root` partition of a leader node or a compute node

About this task

The cluster manager enables booting a leader node or a compute node from its disk only when the disk is configured as MD RAID 1.

Procedure

1. Log into the admin node as the root user.
2. Determine which leader nodes and compute nodes you want to configure with software MD RAID 1.
3. Add the following parameters to the list of parameters in the cluster definition file:

   - `md_metadata=md`
   - `md_raidlevel=1`
   - (Optional) `force_disk="`*`device1`*`,`*`device2`*`,...,`*`deviceN`*`"`

     By default, the cluster manager configures RAID 1 on the first two empty disks.

     If you specify this parameter, specify the path to each disk device required for the RAID level you choose. For example:

     ```
     force_disk="/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:1,/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:2 "
     ```

NOTE: Disk identifiers listed with `force_disk`, such as `/dev/sda`, are used once and are automatically assembled by MD thereafter. These identifiers are not persistent. They could unexpectedly change from what you assumed.

Even though the identifiers are used only once, Hewlett Packard Enterprise recommends that you use `/dev/disk/by-id` or `/dev/disk/by-path` names for the disks instead. In this way, the following occur:

- The cluster manager operates on the exact disks you target.
- You do not rely on device names that might not point where you expect.

(Optional) Configuring software MD RAID on leader nodes and on compute nodes

About this task

The following procedure explains how to use the `cm node provision` command to configure the following:

- Software RAID on leader nodes.
- Software RAID on compute nodes.

Procedure

1. Log into the admin node as the root user.
2. Use the `cm node provision` command in the following format to provide the primary specifications, reboot, and reprovision:

   ```
   cm node provision -n nodes --md-metadata value --md-raidlevel integer \
   --wipe-disk --force-disk devices
   ```

   The variables are as follows:

Variable	Specification
nodes	The affected nodes. For example, use `n[1-5]` for hostnames `n1` through `n5`.
value	The metadata type. Specify one of the following: • `imsm`, which specifies BIOS software RAID. The node must have the Intel Storage Manager or its equivalent for BIOS support. • `md`, which specifies native metadata.

Table Continued

Variable	Specification
integer	An integer that signifies the RAID level you want to configure. The default is `1`.
devices	The disk device names. Use the following format: "*device1*,*device2*,...,*deviceN*" For example: "/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:1,/dev/disk/by-path/pci-0000:5c:00.0-scsi-0:1:0:2" By default, the cluster manager configures RAID 1 on the first two empty disks. **NOTE:** Disk identifiers listed with `--force_disk`, such as `/dev/sda`, are used once and are automatically assembled by MD thereafter. These identifiers are not persistent; they could unexpectedly change from what you assumed. Even though the identifiers are used only once, Hewlett Packard Enterprise recommends that you use `/dev/disk/by-id` or `/dev/disk/by-path` names for the disks instead. In this way, the following occur: • The cluster manager operates on the exact disks you target. • You do not rely on device names that might not point where you expect.

Verifying and splitting the cluster definition file

About this task

A **cluster definition file** contains the following:

- A list of cluster components
- Component-specific characteristics that need to be specified

Complete the following procedure to verify whether you have a cluster definition file, whether that cluster definition file is formatted correctly, and whether the file contains all the information required for the cluster nodes.

Procedure

1. Retrieve a copy of the cluster definition file.

   For clusters that are configured with at least one working slot, enter the following command to generate a cluster definition file:

   ```
   cm system show configfile --all > filename
   ```

   For *filename*, specify any file name. You can write the cluster definition file to any directory.

   The following are additional notes regarding the cluster definition file:

   - The `cm system show configfile` command shown in this step writes one cluster definition file to *filename*. This file lists all the cluster components.
   - If necessary, you can obtain the cluster definition file used in the manufacturing process from your technical support representative.

2. Split the cluster definition file into additional files.

Cluster type	Content of the split files
HPE Cray XD or HPE Apollo cluster without SU leader nodes. For example, an HPE Apollo 80 or an HPE Apollo 2000.	For these clusters, the original cluster definition file contains the information necessary to configure the management switches and the compute nodes. Split the cluster definition file into two files, one for the management switches and one for the compute nodes. If you have components such as power distribution units (PDUs) or additional compute nodes deployed as service nodes, create an additional file for these additional components.
Any cluster that requires a granular configuration approach.	Using more than one cluster definition file lets you take a step-by-step approach to the order in which components are configured into the cluster. Always create a cluster definition file for the management switches and use that file to configure the management switches into the cluster ahead of any other components. Some clusters have compute nodes that serve as service nodes or login nodes and are attached directly to the admin node. Create an additional cluster definition file for those compute nodes, too.

The following are example cluster definition file names and content:

Example file name	Content
`mgmtsw.config`	Management switches only

`compute.config`
 	
`pdu.config`
	Compute nodes or components that the node discovery commands and configuration process do not configure automatically.

The following is one method for splitting a single configuration file into multiple files:

a. Use the `cp` command to copy the original cluster definition file to another one or two files.

b. Name the files according to the components they describe. For example, name one file for switches, one file for compute nodes.

c. Open the file(s) you just created, and search for the `[discover]` section. Use an editor such as `vim`.

d. Retain the lines that pertain to the components for which the file is named. Delete the other lines. For example, in a file for management switches, delete the lines that pertain to all other components. The file should contain only lines for management switches.

e. Review these files carefully before proceeding.

The following examples show the contents of example files. The files show parts of cluster definition files for various components. The ellipsis (`...`) indicates that the lines can be longer and include more information.

Example 1. To create a cluster definition file for management switches only, include the following types of lines:

```
[discover]
internal_name=mgmtsw0, type=spine, ...
internal_name=mgmtsw1, type=leaf, ...
```

Example 2:

If a switch is not defined in the cluster definition file, enter information into the cluster definition file manually. To find the switch MAC address, either open a console to the switch or visually inspect the outside of the switch to find its label. If the switch does not support DHCP, configure a static IP address for the switch that matches the `mgmt_net_ip=` attribute in the configuration file. For example:

```
[discover]
# Aruba VSX Dual Control Plane Spine Management Switches
internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:d4:43:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, mgmt_net_ip=172.23.255.252, hostname1=sw-spine01,
mgmtsw_partner=sw-spine02
internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:d3:07:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, mgmt_net_ip=172.23.255.253, hostname1=sw-spine02,
mgmtsw_partner=sw-spine01
# Aruba VSX Dual Control Plane Leaf Management Switches
internal_name=mgmtsw2, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:ab:44:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, mgmt_net_ip=172.23.255.100, hostname1=sw-leaf01,
mgmtsw_partner=sw-leaf02
internal_name=mgmtsw3, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:cd:07:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, mgmt_net_ip=172.23.255.101, hostname1=sw-leaf02,
mgmtsw_partner=sw-leaf01
```

Example 3. To create a cluster definition file for compute nodes only, include the following types of lines:

```
[discover]
internal_name=service0, mgmt_net_name=head, mgmt_bmc_net_name=head-bmc, ...
internal_name=service1, mgmt_net_name=head, mgmt_bmc_net_name=head-bmc, ...
```

For more information, see the following:

Cluster definition file contents

Cluster definition file examples with node templates, network interface card (NIC) templates, and predictable names

Cluster definition file contents

Hewlett Packard Enterprise recommends that you use a cluster definition file when you configure and bring up the cluster system. When you use a cluster definition file, all the cluster configuration data resides in files that are easy to maintain and easy to edit. The cluster definition file also removes uncertainly when you configure the cluster. When you use the node discovery commands, specify the cluster definition file that includes the nodes and components that you want to configure.

The node discovery commands include the `cm node add` command, the `cm node discover add` command, and for some purposes, also the `configure-cluster` command. All these node discovery commands can accept a cluster definition file as input.

In the cluster definition file, each component is defined with several configuration attributes. For example, these configuration attributes can include MAC addresses, IP addresses, component roles, hostnames, management network details, the node image assignment, and much more.

For information about configuration attributes, enter the following command:

```
# man cluster-configfile
```

If you no longer have the cluster definition file for the cluster, you can obtain the original cluster definition file from the HPE factory. Another way to obtain a cluster definition file is to enter the following command and build a file from the resulting file:

```
# cm system show configfile --all
```

The following table shows the types of cluster components in the cluster definition file:

Component	Notes
Management switches	If you use management switches that HPE does not support, see the following before you include information about the switches in the cluster definition file: **(Conditional) Configuring an unsupported Ethernet switch into the cluster** **Configuring a new switch**
Power distribution units (PDUs)	The cluster manager does not configure PDUs into the cluster automatically on clusters without leader nodes. Define the PDUs in the cluster definition file. PDUs are numbered starting with `0`. For example, `pdu0`, `pdu1`, `pdu2`, and so on.
Compute nodes	If you do not have a cluster definition file, use the `cm node discover` command to configure the compute nodes into the cluster. If you have a cluster definition file, verify the entries for each compute node under the `[discover]` heading, and use the `cm node add` command to add the compute nodes into the cluster. For information about required fields for compute nodes, enter the following command: `# man cluster-configfile`

By default, HPE configures nodes with hostnames that correspond to their default number, as follows:

Graphic processing units (GPUs) are numbered starting with 1. For example, the factory configures graphical compute nodes with names such as `r01g01`.

Cluster definition file examples with node templates, network interface card (NIC) templates, and predictable names

Contemporary cluster definition files contain node template sections and use predictable NIC names. Use the following keywords at the start of sections in the file that pertain to node templates and NIC templates:

- `[templates]`

  The cluster manager assumes that the lines following the `[templates]` keyword define the characteristics for a specific node type.

  For example, you can define templates for the compute nodes.

  Templates are useful when they pertain to multiple nodes, for example, many identical compute nodes. You can describe the nodes once, in the template section of the cluster definition file. The node template definitions can describe kernel names, image names, node controller authentication info, and other node characteristics.

For more information, see the `node-templates`(8) manpage.

- `[nic_templates]`

  NIC templates pertain to the NIC devices in specific nodes. Each node template can have one or more NIC templates. The NIC templates explain how to tie networks to interfaces. There can be one NIC template per network. The NIC template definitions can describe the network interfaces for the network, the network name, bonding settings, and so on.

  If you want to have a `[nic_templates]` section in the cluster definition file, also create a `[templates]` section.

  Predictable names pertain to the NICs within each node. These NIC names are the same across like hardware.

  If you have an HA admin node, the two physical admin nodes use legacy names. The HA admin node, which is a virtual machine, uses predictable names.

  InfiniBand devices do not use predictable names.

  For more information about predictable names, see the following:
  Predictable network interface card (NIC) names

By default, the cluster manager reads in templates from the following file when you run the cluster configuration tool:

`/etc/opt/sgi/default-node-templates.conf`

Cluster definition file example - Cluster with HPE Apollo Moonshot system cartridges

About this task

The following procedure explains how to configure HPE Moonshot system cartridges into a cluster.

Procedure

1. Obtain the IP address of the HPE Moonshot chassis.

   If the chassis is configured with a static IP address, connect to the chassis console and determine the iLOCM IP address.

   If the chassis is configured to use DHCP, complete the following steps:

   a. Power on (plug in) the chassis.

      You do not need to power-on the individual cartridges.

      For cabling information, see the following:

      HPE Moonshot 1500 Chassis Setup and Installation Guide

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

   c. Monitor the `/var/log/messages` file.

      Use a command such as `tail -f`.

   d. Wait for an entry that shows the `DHCPDISCOVER` line that includes the MAC address of the iLOCM, and observe the chassis IP address in the lines that follow.

      For example:

      ```
      May 19 15:36:38 cmutay1 dhcpd: DHCPDISCOVER from 9c:b6:54:8a:28:72 via eth0
      May 19 15:36:38 cmutay1 dhcpd: DHCPOFFER on 10.117.23.6 to 9c:b6:54:8a:28:72 via eth0
      May 19 15:36:42 cmutay1 dhcpd: DHCPREQUEST for 10.117.23.6 (10.117.20.74) from 9c:b6:54:8a:28:72 via eth0
      May 19 15:36:42 cmutay1 dhcpd: DHCPACK on 10.117.23.6 to 9c:b6:54:8a:28:72 via eth0
      ```

The IP address is `10.117.23.6`.

2. From the admin node, use the `cm_scan_moonshot` command to generate information that you can include in the cluster definition file.

   This step generates node definitions for all the cartridges in the HPE Moonshot system chassis.

   The `cm_scan_moonshot` command has several parameters. The following command line shows the basic parameters needed to generate information for the cluster definition file:

   `cm_scan_moonshot -L` *`ilocm_ip(s)`* `-G ["`*`string`*`"] -n` *`name_syntax`* `-o` *`outfile`*

   The variables are as follows:

Variable	Specification
ilocm_ip(s)	The IP address of one or more iLOCMs, that is the iLO chassis controllers. If you specify more than one IP address, use a comma (`,`) to separate each address.
string	Optional. A string of node information that you want the cluster manager to write to the output file. Enclose the string in quotation marks (`" "`).
name_syntax	A pattern for the generated node names. You can include the wildcard characters that this command supports. For a list of these characters, enter the following: `# cm_scan_moonshot -h` For example, if you specify `-n node%2i`, the command generates node names that start with `node` and have a 2-integer suffix. That is, in the cluster definition file, the nodes are numbered as `node01`, `node02`, `node03`, … `node99`.
outfile	An output file name.

   For example:

   - Example 1. Assume that you have one chassis, and you want to configure the 10 cartridges in that chassis into an HPE Apollo cluster. Enter the following command to generate node definitions:

     ```
     # cm_scan_moonshot -L 172.24.5.5 \
     -G "tpm_boot=no, predictable_net_names=yes, force_disk=/dev/sda, destroy_disk_label=yes" \
     -n node%2i -o /tmp/moonshot.txt
     ```

     This command scans the HPE Moonshot system chassis at `172.24.5.5` and generates a file called `moonshot.txt`. The file contains a series of 10 node definitions suitable for appending to a cluster definition file and is as follows:

     ```
     internal_name=service01, hostname1=node01, mgmt_net_macs=38:ea:a7:0f:48:08, mgmt_bmc_net_macs=38:ea:a7:0f:66:fe,
     mgmt_bmc_net_ip=172.24.5.5, card_type=ILOCM, architecture=x86_64, console_device=ttyS0, baud_rate=9600,
     bmc_username=admin, bmc_password=admin123, tpm_boot=no, predictable_net_names=yes, force_disk=/dev/sda,
     destroy_disk_label=yes
     internal_name=service02, hostname1=node02, mgmt_net_macs=38:ea:a7:0f:3d:b6, mgmt_bmc_net_macs=38:ea:a7:0f:66:fe,
     mgmt_bmc_net_ip=172.24.5.5, card_type=ILOCM, architecture=x86_64, console_device=ttyS0, baud_rate=9600,
     bmc_username=admin, bmc_password=admin123, tpm_boot=no, predictable_net_names=yes, force_disk=/dev/sda,
     destroy_disk_label=yes
     .
     .
     .
     ```

The `-G` parameter appends the additional configuration attributes, in a comma-separated list, to the lines for each compute node.

- Example 2. Assume that you have two chassis and that you want to generate node definitions in the output file that include the configuration attribute `predictable_net_names=yes`. Enter the following command:

```
# cm_scan_moonshot -L 172.24.5.4,172.24.5.5 \
-G "predictable_net_names=yes" -n node%3i -o /tmp/moonshot.txt
INFO: It looks like StrictHostKeyChecking is set to 'no' in /root/.ssh/config...
Make sure you can ssh to all client nodes without providing a password or answering
(yes/no) to a registration question or various CMU commands/systems will fail to run.
45 nodes scanned from ILOCM 172.24.5.4
45 nodes scanned from ILOCM 172.24.5.5

Scanning complete. 90 node(s) written to file /opt/clmgr/tmp/tmp_scan_file-20077
Final scan results written to file: /tmp/moonshot.txt
```

3. Open the output file and the cluster definition file in a text editor.
4. Find the `[discover]` section in the cluster definition file, and add the lines from the output file at the end of the `[discover]` section.

   The following shows a cluster definition file that contains lines for the HPE Apollo Moonshot system cartridges:

```
[templates]
.
.
.
[discover]
internal_name=service01, hostname1=node01, mgmt_net_macs=38:ea:a7:0f:48:08, mgmt_bmc_net_macs=38:ea:a7:0f:66:fe,
mgmt_bmc_net_ip=172.24.5.5, card_type=ILOCM, architecture=x86_64, console_device=ttyS0, baud_rate=9600,
bmc_username=admin, bmc_password=admin123, tpm_boot=no, predictable_net_names=yes, force_disk=/dev/sda,
destroy_disk_label=yes
internal_name=service02, hostname1=node02, mgmt_net_macs=38:ea:a7:0f:3d:b6, mgmt_bmc_net_macs=38:ea:a7:0f:66:fe,
mgmt_bmc_net_ip=172.24.5.5, card_type=ILOCM, architecture=x86_64, console_device=ttyS0, baud_rate=9600,
bmc_username=admin, bmc_password=admin123, tpm_boot=no, predictable_net_names=yes, force_disk=/dev/sda,
destroy_disk_label=yes
.
.
.
```

5. Use the `cm node add` command in the following format to configure the cartridges into the cluster:

   `cm node add --allow-duplicate-macs-and-ips -c` *`config.file`*

   For *config.file*, specify the name of the cluster definition file you edited in this procedure.

6. Use the following command to scan the chassis:

   `cm_scan_moonshot -L` *`ilocm_ip(s)`*

   For *ilocm_ip(s)*, specify the same IP address(es) that you specified in the following step:

   Step **2**

   These IP address(es) are for one or more iLOCMs, that is the iLO chassis controllers. If you specify more than one IP address, use a comma (`,`) to separate each address.

   When you use the `cm_scan_moonshot` command in this format, the command updates the cluster database with cartridge and node location information. This command is essential for proper power operations.

7. Enter the following commands:

```
# cm node update config --sync -n '*'
# cm node refresh netboot -n '*'
```

8. Use the `cm node provision` command to provision each node with an image.
9. Proceed to the following:

Cluster definition file example - Cluster with 100 compute nodes and no leader nodes

This example cluster definition file is for a cluster with 100 compute nodes and no leader nodes. For simplicity, the example file shows only two compute nodes and the management switches. The following information highlights some characteristics of this cluster:

- The information in the `internal_name` field defines the role for each compute node in this cluster.

  The format of the `internal_name` field for each service node, compute node or scalable unit (SU) leader node must be `service`*n*, where *n* is a number from `1` through `101`.

- The `hostname1` field defines the hostname that users must specify when they want to log into a node. The `hostname1` field contains the text that appears in the output for most cluster manager commands.

  You can specify any name, in any format, for the hostname. You can use the `hostname` of the node as its `internal_name`.

- The cluster definition file specifies a multicast installation that uses `udpcast` transport for the compute service nodes. The compute service nodes are `service1` and `service101`.

- The top-level switch, `mgmtsw0`, is defined as spine switch. This switch is always connected to the admin node. Switch `mgmtsw1` is defined as a leaf switch. Switch `mgmtsw1` is connected to the spine switch, `mgmtsw0`.

- The definition for both switches includes `ice=no` because this cluster does not have leader nodes or ICE compute nodes.

The file is as follows:

```
# File compute.config
# Cluster definition file for regular compute 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, custom_groups="comp"
[templates]
# service 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, custom_groups="pbs,login"
[nic_templates]
template=compute, network=head, bonding_master=bond0, bonding_mode=active-backup,
net_ifs="eno1"
template=compute, network=head-bmc, net_ifs="bmc0"
template=compute, network=ib0, net_ifs="ib0"
template=compute, network=ib1, net_ifs="ib1"
[discover]
internal_name=mgmtsw0, mgmt_net_macs="40:b9:3c:a2:54:50", mgmt_net_name=head,
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=spine,
mgmt_net_ip=172.23.255.254
internal_name=mgmtsw1, mgmt_net_macs="40:b9:3c:a4:6c:a7", mgmt_net_name=head,
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=leaf, mgmt_net_ip=172.23.100.1
internal_name=service1, mgmt_bmc_net_macs="20:67:7c:e4:9a:12",
mgmt_net_macs="00:0f:53:21:98:13", template_name=compute
```

```
internal_name=service2, mgmt_bmc_net_macs="20:67:7c:e4:9a:23",
mgmt_net_macs="00:0f:53:21:98:24", template_name=compute
internal_name=service3, mgmt_bmc_net_macs="20:67:7c:e4:9a:34",
mgmt_net_macs="00:0f:53:21:98:35", template_name=compute
internal_name=service4, mgmt_bmc_net_macs="20:67:7c:e4:9a:45",
mgmt_net_macs="00:0f:53:21:98:46", template_name=compute
internal_name=service5, mgmt_bmc_net_macs="20:67:7c:e4:9a:56",
mgmt_net_macs="00:0f:53:21:98:57", template_name=compute
internal_name=service6, mgmt_bmc_net_macs="20:67:7c:e4:9a:67",
mgmt_net_macs="00:0f:53:21:98:68", template_name=compute
.
.
.
```

Cluster definition file example - Compute nodes with an Arm (AArch64) architecture type

If any compute nodes in the cluster are of the Arm (AArch64) architecture type, specify additional information in the cluster definition file for the nodes. For these nodes, specify the following keywords:

- `image=`*`image_name`*
- `kernel=`*`kernel_name`*
- `architecture=`*`arch`*

The following file defines compute nodes with an Arm (AArch64) architecture:

```
# File aarch64_compute.config
# Cluster definition file for AArch64 architecture compute nodes
[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=ipxe-direct, disk_bootloader=no,
predictable_net_names=yes, console_device=ttyS0, conserver_ondemand=no, conserver_logging=yes,
rootfs=disk, card_type=iLO, baud_rate=115200, bmc_username=ADMIN, bmc_password=ADMIN,
image=sles15sp5-arm64, kernel=5.14.21-150500.48-default, architecture=aarch64

[nic_templates]
template=compute, network=head, bonding_master=bond0, bonding_mode=active-backup, net_ifs="eno1"
template=compute, network=head-bmc, net_ifs="bmc0"
template=compute, network=ib0, net_ifs="ib0"
template=compute, network=ib1, net_ifs="ib1"

[discover]
internal_name=mgmtsw0, mgmt_net_macs="40:b9:3c:a2:54:50", mgmt_net_name=head,
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=spine, mgmt_net_ip=172.23.255.254
internal_name=mgmtsw1, mgmt_net_macs="40:b9:3c:a4:6c:a7", mgmt_net_name=head,
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=leaf, mgmt_net_ip=172.23.100.1,
internal_name=service1, mgmt_bmc_net_macs="20:67:7c:e4:9a:12", mgmt_net_macs="00:0f:53:21:98:13",
template_name=compute
internal_name=service2, mgmt_bmc_net_macs="20:67:7c:e4:9a:23", mgmt_net_macs="00:0f:53:21:98:24",
template_name=compute
internal_name=service3, mgmt_bmc_net_macs="20:67:7c:e4:9a:34", mgmt_net_macs="00:0f:53:21:98:35",
template_name=compute
internal_name=service4, mgmt_bmc_net_macs="20:67:7c:e4:9a:45", mgmt_net_macs="00:0f:53:21:98:46",
template_name=compute
internal_name=service5, mgmt_bmc_net_macs="20:67:7c:e4:9a:56", mgmt_net_macs="00:0f:53:21:98:57",
template_name=compute
internal_name=service6, mgmt_bmc_net_macs="20:67:7c:e4:9a:67", mgmt_net_macs="00:0f:53:21:98:68",
template_name=compute
```

Cluster definition file example - Virtual admin node on an HA admin cluster

The example in this topic is a cluster definition file fragment that assigns an IP address to the storage unit. When the storage unit has an IP address, the virtual admin node can access the storage unit whenever the need arises. In addition,

the file assigns IP addresses to the physical admin nodes. The presence of these IP addresses enables access to the physical admin nodes from the virtual admin node.

The file fragment is as follows:

```
# File generic_components.config
# Cluster definition file for components in the cluster that only need an IP address
[discover]
internal_name=mgmtsw0, mgmt_net_macs="40:b9:3c:a2:54:50", mgmt_net_name=head,
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=spine, mgmt_net_ip=172.23.255.254
internal_name=service50, mgmt_net_name=head, mgmt_net_macs="00:0f:45:ac:93:13",
hostname1=is5110a, discover_skip_switchconfig=yes, generic
internal_name=service51, mgmt_net_name=head, mgmt_net_macs="00:0f:45:ac:93:aa",
hostname1=is5110b, discover_skip_switchconfig=yes, generic
internal_name=service52, mgmt_net_name=head, mgmt_net_macs="00:02:aa:ac:9a:ff",
hostname1=genericnode1, discover_skip_switchconfig=yes, generic
internal_name=service53, mgmt_net_name=head, mgmt_net_macs="00:ca:31:a3:9c:b9",
hostname1=othernode1, discover_skip_switchconfig=yes, other
```

In this example, notice that the storage unit is configured.

Cluster definition file example - Specifying a specific IP address

When you run the node discovery commands for a specific component, you can specify an IP address for that component on any of the networks.

For example, the following node definition shows the parameters that you can use to define network IP address specifications for node `service0`:

```
# File specific_ip.config
# Cluster definition file for compute nodes with specific IP addresses for various networks
[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,
data1_net_interfaces="ens1f0,ens1f1", data1_net_name="tengignet", data1_net_bonding_mode=802.3ad,
data1_net_bonding_master=bond1, mgmt_bmc_net_if=yes

[nic_templates]
template=compute, network=head, bonding_master=bond0, bonding_mode=active-backup, net_ifs="eno1"
template=compute, network=head-bmc, net_ifs="bmc0"
template=compute, network=ib0, net_ifs="ib0"
template=compute, network=ib1, net_ifs="ib1"

[discover]
internal_name=service101, mgmt_bmc_net_macs="20:67:7c:e4:9a:10", mgmt_net_macs="00:0f:53:21:98:11",
data1_net_macs="00:03:80:aa:bb:ca,00:03:80:aa:bb:cb", mgmt_bmc_net_ip=172.24.1.1,
mgmt_net_ip=172.23.1.1, data1_net_ip=10.10.1.1, ib_0_ip=10.148.1.1, ib_1_ip=10.149.1.1,
template_name=compute, mgmt_bmc_net_if_ip=172.24.1.4

internal_name=service102, mgmt_bmc_net_macs="20:67:7c:e4:9a:21", mgmt_net_macs="00:0f:53:21:98:22",
data1_net_macs="00:03:80:aa:bb:ab,00:03:80:aa:bb:ac", mgmt_bmc_net_ip=172.24.1.2,
mgmt_net_ip=172.23.1.2, data1_net_ip=10.10.1.2, ib_0_ip=10.148.1.2, ib_1_ip=10.149.1.2,
template_name=compute, mgmt_bmc_net_if_ip=172.24.1.5

internal_name=service103, mgmt_bmc_net_macs="20:67:7c:e4:9a:32", mgmt_net_macs="00:0f:53:21:98:33",
data1_net_macs="00:03:80:aa:bb:ea,00:03:80:aa:bb:eb", mgmt_bmc_net_ip=172.24.1.3,
mgmt_net_ip=172.23.1.3, data1_net_ip=10.10.1.3, ib_0_ip=10.148.1.3, ib_1_ip=10.149.1.3,
template_name=compute, mgmt_bmc_net_if_ip=172.24.1.6
```

After installation, you can use the `cm node set --update-ip` command to change the IP address setting as needed. For more information, see the following:

HPE Performance Cluster Manager Administration Guide

Cluster definition file example - HPE Apollo 20 nodes

If the cluster includes any HPE Apollo 20 compute nodes, set the `dhcp_bootfile=ipxe-direct` configuration attribute for these nodes in the cluster definition file. This attribute is required on HPE Apollo 20 compute nodes.

For example:

```
internal_name=service222, hostname1=apollo222, mgmt_bmc_net_name=head-bmc, mgmt_bmc_net_macs="a4:bf:01:6a:08:73",
mgmt_net_bonding_master=bond0, mgmt_net_bonding_mode=active-backup, mgmt_net_macs="a4:bf:01:6a:08:72", disk_bootloader=no,
geolocation="apollo222", predictable_net_names=yes, mgmt_bmc_net_name=head-bmc, mgmt_net_name=head,
mgmt_net_bonding_master=bond0, transport=bt, redundant_mgmt_network=no, switch_mgmt_network=yes, dhcp_bootfile=ipxe-direct,
conserver_logging=yes, conserver_ondemand=no, tpm_boot=no, disk_bootloader=no, mgmtsw=mgmtsw0, console_device=ttyS0,
mgmt_net_bonding_mode=active-backup, rootfs=tmpfs, mgmt_net_interfaces="eno1", card_type=IPMI, bmc_username=admin123,
bmc_password=admin123, baud_rate=115200
```

Cluster definition file example - HPE Apollo 80 nodes

About this task

This topic explains the following:

- How to check the cluster definition file for an HPE Apollo 80 node.
- How to complete the configuration for an HPE Apollo 80 node.

Procedure

1. Create a cluster definition file that includes the chassis controller and the nodes.

   Key information in this file includes the `generic` keyword and the `mgmt_bmc_net_ip=` address. For example, assume that you create file `cmc.computes.config` with the following lines:

```
[templates]
name=compute-1G-A80, console_device=ttyAMA0, conserver_logging=yes, mgmt_net_name=head,
mgmt_net_interfaces="eno1", mgmt_bmc_net_name=head-bmc, rootfs=disk, mgmt_net_bonding_master=bond0,
dhcp_bootfile=grub2, disk_bootloader=no, transport=bt, switch_mgmt_network=yes, tpm_boot=no,
conserver_ondemand=no, mgmt_net_bonding_mode=802.3ad, redundant_mgmt_network=no,
predictable_net_names=yes, baud_rate=115200, data1_net_name=ib0, data1_net_interfaces="ib0",
card_type=bmx, architecture=aarch64, mgmt_bmc_net_ip=172.24.0.14,
mgmt_bmc_net_macs="2c:d4:44:ce:c9:51"

# chassis controller info:
[discover]
internal_name=service90, hostname1=a80cmc, generic, mgmt_bmc_net_name=head-bmc,
mgmt_bmc_net_macs="2c:d4:44:ce:c9:51", mgmt_net_macs="2c:d4:44:ce:c9:52",  mgmt_net_name=head,
switch_mgmt_network=yes, conserver_logging=no, conserver_ondemand=no, mgmt_bmc_net_ip=172.24.0.14,
mgmt_net_ip=172.23.0.14

# compute node info:
internal_name=service200, hostname1=a80-0, rack_nr=1, chassis=1, node_nr=0,
mgmt_net_macs="2c:d4:44:ce:8a:4e", template_name=compute-1G-A80

internal_name=service201, hostname1=a80-1, rack_nr=1, chassis=1, node_nr=1,
mgmt_net_macs="2c:d4:44:ce:8a:53", template_name=compute-1G-A80

internal_name=service202, hostname1=a80-2, rack_nr=1, chassis=1, node_nr=2,
mgmt_net_macs="2c:d4:44:ce:8a:6b", template_name=compute-1G-A80

internal_name=service203, hostname1=a80-3, rack_nr=1, chassis=1, node_nr=3,
mgmt_net_macs="2c:d4:44:ce:8a:6c", template_name=compute-1G-A80

internal_name=service204, hostname1=a80-4, rack_nr=1, chassis=1, node_nr=4,
mgmt_net_macs="2c:d4:44:ce:8a:6d", template_name=compute-1G-A80

internal_name=service205, hostname1=a80-5, rack_nr=1, chassis=1, node_nr=5,
mgmt_net_macs="2c:d4:44:ce:8a:6e", template_name=compute-1G-A80
```

```
internal_name=service206, hostname1=a80-6, rack_nr=1, chassis=1, node_nr=6,
mgmt_net_macs="2c:d4:44:ce:8a:50", template_name=compute-1G-A80

internal_name=service207, hostname1=a80-7, rack_nr=1, chassis=1, node_nr=7
mgmt_net_macs="2c:d4:44:ce:8a:65", template_name=compute-1G-A80
```

2. Use the `cm node add` command in the following format to configure the components into the cluster:

   `cm node add -c config_file --allow-duplicate-macs-and-ips`

   For *config_file*, specify the name of the first cluster definition file you edited in this procedure. This file includes information about the chassis controllers and switches. For example, specify `cmc.computes.config`.

3. Use the `cm node provision` command to provision each node with an image.

4. Proceed to the following:

   Backing up the cluster

Cluster definition file example - Entries for service nodes with NICs for a data network

About this task

If you used the menu-driven cluster configuration tool to create a data network, create a cluster definition file for the service nodes that host the data network.

Specify this cluster definition file to the `cm node add` command to configure these nodes into the cluster. Run the `cm node add` against this file before you configure the compute nodes into the cluster. Alternatively, you could include these nodes in the cluster definition file.

Procedure

1. On the admin node, create a new cluster definition file, and add the node specifications for the two service nodes to the file.

   The following shows a completed example cluster definition file for the two services nodes. The two data networks' attributes are shown in **bold**:

```
[discover]
hostname1=toki-1-srv0, internal_name=service0, mgmt_bmc_net_name=head-bmc, mgmt_bmc_net_macs="0c:c4:7a:1b:45:93",
mgmt_net_name=head, mgmt_net_bonding_master=bond0, mgmt_net_bonding_mode=active-backup,
mgmt_net_macs="0c:c4:7a:14:04:6e", mgmt_net_interfaces="ens1f0", mgmt_net_interface_name="toki-1-srv0",
data1_net_name=ib0, data1_net_interfaces="ib0", data1_net_interface_name="toki-1-srv0-ib0", data2_net_name=ib1,
data2_net_interfaces="ib1", data2_net_interface_name="toki-1-srv0-ib1", rootfs=disk, transport=udpcast,
conserver_logging=yes, conserver_ondemand=no, dhcp_bootfile=grub2, disk_bootloader=no, mgmtsw=mgmtsw0,
predictable_net_names=yes, redundant_mgmt_network=no, switch_mgmt_network=yes, tpm_boot=no, console_device=ttyS1,
architecture=x86_64, card_type="IPMI"
```

   NOTE: To configure service nodes for a high-speed network or a 10G network, create a cluster definition file similar to the one in this step.

2. Use the `cm node provision` command to provision each node with an image.

Cluster definition file example - Attributes for a management switch

The cluster manager supports different types of redundancy protocol switches. Traditionally, the terminology for redundancy has been **stacking**, which means two or more physical switches act as a single logical switch. This is known as **single control plane**. When two physical switches each act as independent logical switches, this is known as **dual control plane**.

When you define a dual control plane spine or leaf type management switch, specify the `mgmtsw_partner=`*`hostname`* attribute in the cluster definition file to define the dual control plane partner switch.

The following example defines a dual control plane spine switch in the cluster definition file:

```
[discover]
internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:d4:43:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, mgmt_net_ip=172.23.255.252, hostname1=sw-spine01,
mgmtsw_partner=sw-spine02
internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:d3:07:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, mgmt_net_ip=172.23.255.253, hostname1=sw-spine02,
mgmtsw_partner=sw-spine01
```

NOTE: When you define the IP addresses of the dual control plane spine switches, do not specify the IP address of the head network gateway. The dual control plane switches use an Active Gateway protocol to emulate the head network gateway. This protocol virtualizes the IP address in case of partial switch failure.

The following command shows how to identify this IP address:

```
# cadmin --show-head-gateway
172.23.255.254
```

The following example defines a dual control-plane leaf switch in the cluster definition file:

```
[discover]
internal_name=mgmtsw2, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:ab:44:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, mgmt_net_ip=172.23.255.100, hostname1=sw-leaf01,
mgmtsw_partner=sw-leaf02
internal_name=mgmtsw3, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:cd:07:00", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, mgmt_net_ip=172.23.255.101, hostname1=sw-leaf02,
mgmtsw_partner=sw-leaf01
```

The following example defines a single control plane spine switch in the cluster definition file:

```
internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:aa:32:12",
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=spine, mgmt_net_ip=172.23.255.254
```

The following example defines a single control plane leaf switch in the cluster definition file:

```
internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:ba:56:12",
redundant_mgmt_network=yes, net=head/head-bmc, ice=no, type=leaf, mgmt_net_ip=172.23.255.100
```

Cluster definition file example - Entries for an unsupported switch

The following entries define an unsupported switch in the cluster definition file:

```
### Example of a config file with unsupported switches and a defined IP address
[discover]
internal_name=service50, hostname1=dell-sw1, mgmt_net_name=head, mgmt_net_macs="0a:cc:99:98:e5:af", generic,
mgmt_net_ip=172.23.255.240
internal_name=service51, hostname1=dell-sw2, mgmt_net_name=head, mgmt_net_macs="0a:cc:99:98:e7:aa", generic,
mgmt_net_ip=172.23.255.241
```

(Optional) Creating a custom partitions configuration file

Prerequisites

- **(Optional) Configuring custom partitions on the admin node**
- Respond to the installation dialog prompts in a way that facilitates custom partitions as described in the following topic:

  Inserting the installation USB device and booting the admin node

About this task

The procedure in this topic explains how to create a configuration file for custom partitions on one or more compute nodes.

Procedure

1. Change to the following directory:

   `/opt/clmgr/image/scripts/pre-install`

2. Open file `custom_partitions.cfg`.

   This name is the default name for the custom partition configuration file, but you can rename this file as needed. You can create multiple files. If you create multiple files, you can use any names for the files.

   NOTE: The order in which you list filesystems 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 filesystem for `/var` and a filesystem for `/var/log`, list `/var` before `/var/log`.

   Many versions of Linux require that the root filesystem (`/`) 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 the guidelines in the custom partition configuration file to describe the custom partitions you want to create.
4. Save and close the custom partition configuration file.
5. Open the cluster definition file for the compute nodes.
6. Add information about compute node custom partitions to the cluster definition file.

   Decide which compute nodes require custom partitions. Locate the node definition lines for those nodes in the compute node cluster definition file. For each line, add the following configuration attribute, which points to the custom partition configuration file:

   `custom_partitions=`*`file`*`.cfg`

   For example, assume that node `service1` uses the partition layout specified in the default custom partition file `custom_partitions.cfg`. You could have the following specification in the cluster definition file for compute nodes:

   ```
   internal_name=service1, mgmt_bmc_net_name=head-bmc,
   mgmt_bmc_net_macs=0c:c4:7a:c0:77:fc, mgmt_net_name=head,
   mgmt_net_macs="0c:c4:7a:c0:7a:00,0c:c4:7a:c0:7a:01", hostname1=r01n02,
   ```

```
rootfs=disk, transport=udpcast, redundant_mgmt_network=no,
switch_mgmt_network=yes, conserver_logging=yes,
conserver_ondemand=no, console_device=ttyS1,
custom_partitions=custom_partitions.cfg
```

7. Save and close the cluster definition file.

Configuring the management switches into the cluster

About this task

The procedure in this topic adds the management switches into the cluster.

Procedure

1. Log into the admin node as the root user.
2. Verify the cluster definition files you created.
3. Configure the management switches into the cluster.

   Use the `cm node add` command in the following format:

   `cm node add -c management_switch_file`

   For *management_switch_file*, specify the name of the file that includes information about the management switches.

   For example:

   ```
   # cm node add -c mgmtsw.config
   ```

4. (Optional) Monitor the switch configuration process.

   If management switches or components that require management switch configuration were configured, enter the following command to monitor the progress of the switch configuration:

   ```
   # tail -f /opt/clmgr/log/switchconfig.log
   ```

5. Use the `cm mgmtswitch set` command to change the management switch password for the `admin` account.

   The format for this command is as follows:

   `cm mgmtswitch set -s hostname -p new_password --update-switch --skip-update-config`

   The variables are as follows:

Variable	Specification
hostname	The hostname of the management switch
new_password	A strong, new password for the switch

   For example:

   ```
   # cm mgmtswitch set -s mgmtsw0 -p Hp3@dm!n2o20 --update-switch --skip-update-config
   ```

   NOTE: Hewlett Packard Enterprise strongly recommends that you implement standard and secure practices to store all passwords at your site. Do not lose this information.

6. Enter the following command to save the changed configuration to the nonvolatile memory (NVM) on the switches:

   ```
   # switchconfig config -s all --save
   ```

(Conditional) Creating a compute node image for a fabric management node (FMN) and assigning the image to an FMN

About this task

The HPE Performance Cluster Manager 1.10 release was tested with HPE Slingshot interconnect version 2.1.0. The FMN must host one of the following operating systems:

- RHEL 8.7
- SLES 15 SP4

Notice that the cluster manager does not support the preceding operating systems on admin nodes or on leader nodes. The cluster manager supports the preceding operating systems on compute nodes, service nodes, and FMNs.

For more HPE Slingshot information, see the HPE Slingshot documentation.

Complete the procedure in this topic if the cluster has HPE Slingshot interconnect fabric.

Procedure

1. Verify the operating system on the FMN and on the cluster manager management network.

   If the operating system levels are different, you might have to create a separate repository environment as part of the process to create an FMN image. In addition, review the HPE Slingshot documentation and be aware of the HPE Slingshot requirements.

2. **Creating an image for a fabric management node (FMN)**
3. Complete one of the following procedures to provision the fabric management node (FMN):

   - **Method 1 - Configuring a new fabric management node (FMN) into the cluster and assigning an image to that new FMN node**
   - **Method 2 - Assigning a new fabric management node (FMN) image to an existing FMN in the cluster**

4. **Verifying that the new fabric management node (FMN) compute node image is hosted on the FMN**

Creating an image for a fabric management node (FMN)

Procedure

1. Download the FMN software from the HPE Support Center.

   You can download the FMN software to a host at your site and move the packages to the admin node.

   To create a RHEL FMN image, do one of the following:

   - Make both the RHEL repository and the extra packages for enterprise Linux (EPEL) repository available as remote repositories on the admin node.

Or

- Create local mirrors of the required RHEL and EPEL software.

2. Log into the admin node as the root user.
3. Add the cluster manager repository.

   This step differs depending on the operating system, as follows:

   - On RHEL 8, enter the following command:

     ```
     # cm repo add cm-1.10-cd1-media-rhel8X-x86_64.iso
     ```

   - On SLES 15, enter the following command:

     ```
     # cm repo add cm-1.10-cd1-media-sles15spX-x86_64.iso
     ```

4. Add the operating system distribution repository.

   Include any operating system updates. This step differs depending on the operating system, as follows:

   - On RHEL 8, enter the following command:

     ```
     # cm repo add RHEL-8.X.X-20211013.2-x86_64-dvd1.iso
     ```

   - On SLES 15, enter the following command to install the base distribution and the updates:

     ```
     # cm repo add SLE-15-SPX-Full-x86_64-GM-Media1.iso
     ```

5. Add the HPE Slingshot repositories.

   Complete the following steps:

   a. Create a directory to host the repositories:

      ```
      # mkdir -p /opt/clmgr/repos/other/fmn
      ```

   b. Enter a `tar` command and a `cm repo add` command in the following formats to add the HPE Slingshot repository:

      - ```
        tar -zxvf slingshot_tar_file_name.tar.gz \
        -C /opt/clmgr/repos/other/fmn/

        cm repo add --custom slingshot-fmn-packages-ss_version-os \
        /opt/clmgr/repos/other/fmn/slingshot-fmn-packages-ss_version-os
        ```

        The variables are as follows:
        ```

Variable	Specification
slingshot_tar_file_name	The name of the HPE Slingshot interconnect `tar` file name. For example: `slingshot-installer-rpms-sles15sp4-2.1.0-548.tar.gz`
ss_version	The string of numbers that identifies the HPE Slingshot interconnect software version. For example, `2.1.0-548`.
os	The operating system. For example, one of the following: ◦ `redhat-8update7`, which is RHEL 8.7 ◦ `sles15sp4` ◦ `rocky`

c. Use the `mkdir` command in the following format to create a directory for the HPE Slingshot repository:

```
mkdir /opt/clmgr/repos/other/fmn/slingshot-fmn-packages-ss_version
```

d. Use the `tar` and `cm repo add` commands in the following format to add the HPE Slingshot repository:

```
tar -zxvf \
sc-firmware-2.X.X.XXX-release.rpm.tar.gz \
-C /opt/clmgr/repos/other/fmn/slingshot-fmn-packages-ss_version/

cm repo add --custom slingshot-fmn-packages-ss_version \
/opt/clmgr/repos/other/fmn/slingshot-fmn-packages-ss_version
```

The variables are as follows:

Variable	Specification
X.X.XXX	The last digits of the HPE Slingshot version number.
release	Identifies the date and the revision number. For example, `20230808061517_ad411cf92b7c.x86_64`
ss_version	The string of numbers that identifies the HPE Slingshot interconnect software version. For example, `2.1.0-548`.

e. (Conditional) Add the RHEL EPEL repository.

Complete this step on RHEL platforms.

Enter the following commands:

```
# cm repo add \
--custom fmn-epel8 \
https://dl.fedoraproject.org/pub/epel/8/Everything/x86_64/
```

6. Create a repository group for the fabric management node repository.

   This step differs depending on the operating system, as follows:

   - On RHEL 8, enter the following commands:

     ```
     # cm repo group add fmn-rhel8.X \
     --repos Cluster-Manager-1.10-rhel8.X-x86_64 \
     Red-Hat-Enterprise-Linux-8.X.X-x86_64 fmn-epel8
     ```

   - On SLES 15, enter the following commands:

     ```
     # cm repo group add fmn-sles15spX \
     --repos Cluster-Manager-1.10-sles15spX-x86_64 \
     SLE-15-SPX-Full-x86_64
     ```

7. Build the initial FMN image.

   Hewlett Packard Enterprise recommends against including the HPE Slingshot revision level in the image name. When you omit the revision level, you allow the image name to be relevant now and in the future.

   This step differs depending on the operating system, as follows:

   - On RHEL 8, enter the following command:

     ```
     # cm image create -i fmn-rhel8.X \
     --pkglist /opt/clmgr/image/rpmlists/generated\
     /generated-group-fmn-rhel8.X.rpmlist \
     --repo-group fmn-rhel8.X
     # cm image set -i fmn-rhel8.X --repo-group fmn-rhel8.X
     ```

   - On SLES 15, enter the following commands:

     ```
     # cm image create -i fmn-sles15spX \
     --pkglist /opt/clmgr/image/rpmlists/generated\
     /generated-group-fmn-sles15spX.rpmlist \
     --repo-group fmn-sles15spX
     # cm image set -i fmn-sles15spX --repo-group fmn-sles15spX
     ```

8. Add the HPE Slingshot repositories to the HPE Slingshot repository group.

   This step differs depending on the operating system, as follows:

- On RHEL 8, enter the following command:

  ```
  # cm repo group add fmn-rhel8.X \
  --repos slingshot-fmn-packages-ss_version-redhat \
  slingshot-fmn-packages-ss_version fmn-epel8
  ```

- On SLES 15, enter the following command:

  ```
  # cm repo group add fmn-sles15spX \
  --repos slingshot-fmn-packages-ss_version-sles \
  slingshot-fmn-packages-ss_version
  ```

9. Install the HPE Slingshot packages.

This step differs depending on the operating system, as follows:

- On RHEL 8, enter the following commands to reset the modules needed to build the FMN image:

  ```
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  module reset container-tools
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  module enable container-tools
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  module reset nginx
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  module enable nginx:1.16
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  install slingshot-fmn-redhat
  ```

- On SLES 15, enter the following command:

  ```
  # cm image zypper -i fmn-sles15spX --duk --repo-group fmn-sles15spX \
  install slingshot-fmn-sles15spX
  ```

10. Add the package that contains the HPE Slingshot fabric check monitoring services.

Enter one of the following commands:

- On RHEL 8 admin nodes, enter the following command:

  ```
  # cm image dnf -i fmn-rhel8.X --duk --repo-group fmn-rhel8.X \
  install slingshot-fabric-check
  ```

- On SLES 15 admin nodes, enter the following command:

  ```
  # cm image zypper -i fmn-sles15spX --duk --repo-group fmn-sles15spX \
  install slingshot-fabric-check
  ```

11. Apply recommended HPE Slingshot settings and remove certificate files.

This step differs depending on the operating system, as follows:

- On RHEL 8, enter the following command:

  ```
  # (
  echo "sysctl -w net.ipv4.neigh.default.gc_thresh1=512"
  ```

```
echo "sysctl -w net.ipv4.neigh.default.gc_thresh2=8000"
echo "sysctl -w net.ipv4.neigh.default.gc_thresh3=10000"
) >> /opt/clmgr/image/images/fmn-rhel8.X/etc/sysctl.conf
# rm -f /opt/clmgr/image/images/fmn-rhel8.X/opt/slingshot/config/ssl\
/fabric-manager.crt
# rm -f /opt/clmgr/image/images/fmn-rhel8.X/opt/slingshot/config/ssl\
/fabric-manager.key
```

- On SLES 15, enter the following command:

```
# (
echo "sysctl -w net.ipv4.neigh.default.gc_thresh1=512"
echo "sysctl -w net.ipv4.neigh.default.gc_thresh2=8000"
echo "sysctl -w net.ipv4.neigh.default.gc_thresh3=10000"
) >> /opt/clmgr/image/images/fmn-sles15spX/etc/sysctl.conf
# rm -f /opt/clmgr/image/images/fmn-sles15spX/opt/slingshot/config/ssl\
/fabric-manager.crt
# rm -f /opt/clmgr/image/images/fmn-sles15spX/opt/slingshot/config/ssl\
/fabric-manager.key
```

12. Create a software revision that documents the image you created.

This step differs depending on the operating system, as follows:

- On RHEL 8, enter the following command:

```
# cm image revision commit -i fmn-rhel8.X \
-m "Installed Slingshot 2.X.X packages"
```

- On SLES 15, enter the following command:

```
# cm image revision commit -i fmn-sles15spX \
-m "Installed Slingshot 2.X.X packages"
```

13. Install the admin node `ssh` keys into the image.

This step differs depending on the operating system, as follows:

- On RHEL 8, enter the following commands:

```
# mkdir /opt/clmgr/image/images/fmn-rhel8.X/root/.ssh
# cp /root/.ssh/id_rsa /opt/clmgr/image/images/fmn-rhel8.X/root/.ssh/
# cp /root/.ssh/id_rsa.pub /opt/clmgr/image/images/fmn-rhel8.X\
/root/.ssh/
# ssh-keygen -p -i -f \
/opt/clmgr/image/images/fmn-rhel8.X/root/.ssh/id_rsa -m pem -N ""
```

- On SLES 15, enter the following commands:

```
# mkdir /opt/clmgr/image/images/fmn-sles15spX/root/.ssh
# cp /root/.ssh/id_rsa /opt/clmgr/image/images/fmn-sles15spX\
/root/.ssh/
# cp /root/.ssh/id_rsa.pub /opt/clmgr/image/images/fmn-sles15spX\
/root/.ssh/
# ssh-keygen -p -i -f \
/opt/clmgr/image/images/fmn-sles15spX/root/.ssh/id_rsa -m pem -N ""
```

Method 1 - Configuring a new fabric management node (FMN) into the cluster and assigning an image to that new FMN node

Prerequisites

Creating an image for a fabric management node (FMN)

Procedure

1. Log into the admin node as the root user.
2. Obtain the management MAC addresses for the FMN.

   You can obtain this information from the FMN node BIOS system. Alternatively, complete the following steps:

   a. Visually inspect the FMN and its cabling. Figure out the switch to which the FMN node is connected.

   b. Use the `switchconfig` command in the following format to display the management MAC addresses associated with the components connected to the switch:

   ```
   switchconfig info -s switch_hostname --fdb
   ```

   For *switch_hostname*, specify the hostname of the switch to which the FMN is connected.

   For example:

   ```
   # switchconfig info -s mgmtsw0 --fdb
   ==== L2 FDB (mac-address-table) Information on mgmtsw0 ====
   Running command - `show fdb`...
   Mac               Vlan cat /etc/*releaseAge Flags Port / Virtual Port List
   ------------------------------------------------------------------------------
   98:f2:b3:21:23:f4 Default(0001) 0012 d m L 1:2
   98:f2:b3:21:53:14 Default(0001) 0004 d m L 1:4
   d0:67:26:d7:5a:b8 Default(0001) 0024 d m L 1:10
   d0:67:26:d7:5a:ba Default(0001) 0000 d mi L 2:10
   ec:eb:b8:94:38:8a Default(0001) 0008 d m L 1:8
   ec:eb:b8:9b:84:28 Default(0001) 0000 d mi L 1:7
   f4:03:43:49:2b:b9 Default(0001) 0000 d m L 1:23
   f4:03:43:49:2b:ba Default(0001) 0013 d m L 2:23
   f4:03:43:49:ca:c8 Default(0001) 0000 d mi L 1:3
   .
   .
   .
   ```

   In the command output, look for the port to which the FMN is connected on each switch.

   In the preceding output example, the MAC address you need is the MAC address on the row that matches the port number where the fabric manager is plugged into the switch. Refer to the cluster system configuration to find which port(s) to which the fabric manager network cables are connected.

   The preceding output example is for a specific switch. Each switch generates unique output.

3. Use a text editor to create a cluster definition file for the FMN.

   NOTE: This step includes example lines for cluster definition files. When you create the file for your cluster, make sure to use values that are appropriate for the cluster.

Example 1. The following file, `fmn.conf`, is appropriate for FMNs with bonded management interfaces.

```
[templates]
name=fmn, tpm_boot=no, mgmt_net_bonding_mode=802.3ad, rootfs=disk, baud_rate=115200, mgmt_net_bonding_master=bond0,
switch_mgmt_network=yes, force_disk=/dev/sda, bmc_username=XXXX, transport=udpcast, mgmt_net_name=head,
card_type=ILO, console_device=ttyS0, bmc_password=XXXX, conserver_ondemand=no, conserver_logging=yes,
mgmt_bmc_net_name=head-bmc, redundant_mgmt_network=yes, predictable_net_names=yes, disk_bootloader=no,
dhcp_bootfile=ipxe-direct, mgmt_bmc_net_if=yes

[discover]
internal_name=fmn, hostname1=fmn1, mgmt_bmc_net_macs="ec:eb:b8:94:38:8a",
 mgmt_net_macs="ec:eb:b8:9b:84:28", mgmt_net_interfaces="ens9998",
template_name=fmn, image=fmn-sles15spX, extra_routes=yes
```

Example 2. The following cluster definition file, `fmn2.conf`, is appropriate for FMNs with single management interfaces.

```
[templates]
name=fmn, tpm_boot=no, rootfs=disk, baud_rate=115200, switch_mgmt_network=yes, force_disk=/dev/sda,
bmc_username=XXXX, transport=udpcast, mgmt_net_name=head, card_type=ILO, console_device=ttyS0, bmc_password=XXXX,
conserver_ondemand=no, conserver_logging=yes, mgmt_bmc_net_name=head-bmc, predictable_net_names=yes, disk_bootloader=no,
dhcp_bootfile=ipxe-direct, mgmt_bmc_net_if=yes

[discover]
internal_name=fmn, hostname1=fmn1, mgmt_bmc_net_macs="94:40:c9:47:2f:12", mgmt_net_macs="14:02:ec:db:d5:c1",
mgmt_net_interfaces="ens9998", template_name=fmn, image=fmn-sles15spX, extra_routes=yes
```

The preceding examples are for guidance only. Make sure to replace the following fields with values that are specific to this cluster:

- `bmc_username` - obtain from the FMN specification for this cluster
- `bmc_password` - obtain from the FMN specification for this cluster
- `card_type` - obtain from the FMN specification for this cluster
- `console_device` - obtain from the FMN specification for this cluster
- `force_disk` - obtain from the FMN specification for this cluster
- `image` - obtain from the FMN specification for this cluster
- `mgmt_bmc_net_macs` - obtain from `switchconfig` output
- `mgmt_bmc_net_interfaces` - obtain from the FMN specification for this cluster
- `mgmt_net_macs` - obtain from `switchconfig` output

To configure more than one FMN into the cluster, add another line of configuration attributes for the additional FMN under the `[discover]` heading in the cluster definition file.

For more cluster definition file examples, see the following:

Verifying and splitting the cluster definition file

4. Use the `cm node add` command in the following format to configure the FMN into the cluster:

   `cm node add --update-templates --skip-existing-nodes -c conf_file`

   For *conf_file*, specify the name of the cluster definition file for the FMN.

5. Wait a few moments, and enter the following command to determine the node status:

   `# cm power status -n "fm*"`

   If the node is up, the command displays either `Off` or `On`. Wait for the node to be up before you proceed to the next step.

6. Provision the FMN to use the new compute image.

   `# cm node provision -n "fm*" --rootfs disk --force-disk disk_name --wipe-disk`

For *disk_name*, specify a `by-id` or `by-path` persistent disk name. Alternatively, specify a disk name in the format `/dev/sda`, but be aware that this disk name style is not persistent.

Method 2 - Assigning a new fabric management node (FMN) image to an existing FMN in the cluster

Prerequisites

Creating an image for a fabric management node (FMN)

Procedure

1. Log into the admin node as the root user.
2. Configure the servers that function as FMNs to network boot.

   When configured to PXE boot, the nodes boot over the network. Enter the following command to configure the FMN to network boot using the `efiboot` settings:

   ```
   # clush -b -w "fm*" "efibootmgr -n \$(efibootmgr | awk '/PXE Boot/ {gsub(/Boot/,\"\"); print \$1}') \
   | egrep 'BootNext|(PXE Boot)'"
   ---------------
   fmn,fmn2
   ---------------
   BootNext: 0008
   Boot0008  PXE Boot
   ```

   The output from the `clush` command includes the following two lines:

   - The first line is `BootNext: (`*`hexadecimal_number`*`)`. The *hexadecimal_number* value indicates the boot option for the server to use the next time it boots.
   - The second line includes the boot device number, `Boot` *`hexadecimal_number`*, assigned to the network boot option.

   When the command in this step is successful, as the preceding output shows, the hexadecimal numbers match.
3. Reprovision the FMN to use the new compute image.

   ```
   # cm node provision -n "fm*" -i image --rootfs disk --force-disk disk_name --wipe-disk
   ```

   The variables are as follows:

Variable	Specification
image	The name of the new FMN node image. For example, `"fmn*"`, `fmn-rhel8.X`, or `fmn-sles15spX`.
disk_name	A disk name. Specify a `by-id` or `by-path` persistent disk name. Alternatively, specify a disk name in the format `/dev/sda`, but be aware that this disk name style is not persistent.

Verifying that the new fabric management node (FMN) compute node image is hosted on the FMN

Prerequisites

One of the following procedures has been completed:

- **Method 1 - Configuring a new fabric management node (FMN) into the cluster and assigning an image to that new FMN node**

  Or

- **Method 2 - Assigning a new fabric management node (FMN) image to an existing FMN in the cluster**

Procedure

1. Log into the admin node as the root user.
2. Verify that the FMNs show as `BOOTED` and that they are using the FMN compute image:

   ```
   # cm power status -n "fm*"
   ```

3. Verify that the compute image you created for the FMN resides on the FMN:

   ```
   # cm node show -I -n "fm*"
   ```

4. Verify that the FMN is running the `fabric-manager` service:

   ```
   # cm node run -d -n "fm*" "systemctl is-active fabric-manager"
   ```

5. (Conditional) Verify that the FMNs are all running the same HPE Slingshot release version.

   Complete this step if you installed a new image on two or more FMNs.

   Enter the following command on each FMN:

   ```
   # /usr/bin/fmn-show-version
   FMN base OS installed version: SUSE Linux Enterprise Server 15 SP4
   FMN Scripts : 2.1.0
   FMN CLI : 2.1.0
   Slingshot Fabric Manager : 2.1.0
   Slingshot Certificate Manager : 2.1.0
   Slingshot Tools : 2.1.0
   Slingshot Document : 2.1.0
   Slingshot UI : 2.1.0
   Slingshot Web Server : 2.1.0
   Slingshot PKI Engine : 1.3.0
   Slingshot SDU RDA : not installed
   Slingshot Switch Firmware (Downloadable) : 2.1.0.59
   Rosetta Development Library : not installed
   ```

   Examine the output, and make sure all the release levels are identical.

6. Configure the FMN software.

   For information about how to configure the FMN software, see the HPE Slingshot documentation.

Running the `cm node add` command on clusters without leader nodes

About this task

The following topics use the `cm node add` command to add nodes to the cluster.

Procedure

1. Use the following procedure to run the `cm node add` command:

   Running the `cm node add` command on a cluster without scalable unit (SU) leader nodes

Running the `cm node add` command on a cluster without scalable unit (SU) leader nodes

Procedure

1. Through an `ssh` connection, log into the admin node as the root user.
2. (Conditional) Activate the NFS compute node image.

   Complete this step if the diskless compute nodes are configured in the cluster definition file with `rootfs=nfs`.

   When you complete this step, the cluster manager assumes that you want to configure the cluster manager to provide NFS services for diskless compute nodes. In other words, you enable the cluster manager to configure administrative nodes to act as an NFS servers for the diskless compute nodes in the cluster. When each compute node boots, it mounts a copy of the NFS root file system to use as the compute node root file system.

   Enter the following command:

   ```
   # cm image activate -i image
   ```

   For *image*, specify the image name.
3. Configure the management switches into the cluster.

   a. First, enter the `cm node add` command in the following format to update the cluster database with relevant templates:

      ```
      cm node add -c mgmtsw_file --update-templates
      ```

      For *mgmtsw_file*, specify the name of the cluster definition file that defines the management switches.

      For example:

      ```
      # cm node add -c mgmtsw.config --update-templates
      ```

   b. Second, enter the `cm node add` command in the following format to configure all the management switches defined in the cluster definition file:

      ```
      cm node add --config-file mgmtsw_file
      ```

      For *mgmtsw_file*, specify the name of the cluster definition file that defines the management switches.

For example:

```
# cm node add --config-file mgmtsw.config
```

4. (Optional) Monitor the switch configuration process.

   If management switches or components that require management switch configuration were configured, enter the following command to monitor the progress of the switch configuration:

   ```
   # tail -f /opt/clmgr/log/switchconfig.log
   ```

5. Use the `cm mgmtswitch set` command to change the management switch password for the `admin` account.

   The format for this command is as follows:

   ```
   cm mgmtswitch set -s hostname -p new_password --update-switch --skip-update-config
   ```

   The variables are as follows:

Variable	Specification
hostname	The hostname of the management switch
new_password	A strong, new password for the switch

   For example:

   ```
   # cm mgmtswitch set -s mgmtsw0 -p Hp3@dm!n2o20 --update-switch --skip-update-config
   ```

   NOTE: Hewlett Packard Enterprise strongly recommends that you implement standard and secure practices to store all passwords at your site. Do not lose this information.

6. Enter the following command to save the changed configuration to the nonvolatile memory (NVM) on the switches:

   ```
   # switchconfig config -s all --save
   ```

7. Run the `cm node add` command twice to configure the compute nodes into the cluster.

   a. First, enter the `cm node add` command in the following format to update the cluster database with relevant templates:

      ```
      cm node add --config-file computes_file --update-templates
      ```

   b. Second, enter the `cm node add` command in the following format to configure the compute nodes defined in the cluster definition file:

      ```
      cm node add --config-file computes_file
      ```

   For example:

   ```
   # cm node add --config-file compute.config --update-templates
   # cm node add --config-file compute.config
   ```

8. (Optional) Use the `cm node console` command to monitor the PXE boot process on one or more compute nodes.

   This command has the following format:

   ```
   cm node console -n hostname
   ```

For *hostname*, specify the hostname of one of the nodes in the cluster definition file. For example, `service1`.

9. Verify that all nodes booted.

   Enter one or more `cm power status` commands. For example, enter the following command to verify the boot status of the service nodes:

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

`cm node add` command examples that use a cluster definition file

The following topics show how to use the `cm node add` command with a cluster definition file.

`cm node add` command example - updating templates in the cluster database

The `[templates]` section of the cluster definition file lets you define node characteristics for a group of nodes. If you edit the `[templates]` section or the `[nic_templates]` sections, enter the `cm node add` command in the following format to update the cluster database:

`cm node add -c` *`config_file_name`* `--update-templates`

For *config_file_name*, specify the name of the configuration file you need to update.

For example:

```
# cm node add -c compute.config --update-templates
```

`cm node add` command examples - configuring one, several, or all components

You can use a single `cm node add` command to configure one component, multiple components, or all cluster components. In the examples in this topic, the format is as follows:

`cm node add -c` *`cluster_definition_file`* `[--arg1` *`value`*`] [--arg2` *`value`*`] ...`

In this format, the command reads the *cluster_definition_file* and adds one, several, or all cluster components defined in the file to the cluster database. During an initial installation, if you run the `cm node add` command multiple times, the required discovery order is as follows:

- Management switches
- All other node types and component types

The following examples show these methods:

- Configuring all management switches

  The following command adds all management switches named in `mgmtsw.config` to the cluster:

  ```
  # cm node add --config-file mgmtsw.config
  ```

- Configuring one management switch

  The following command adds a single management switch, named `mgmtsw0`, to the cluster. The switch has an entry in the cluster definition file called `mgmtsw.config`. The command is as follows:

  ```
  # cm node add --config-file mgmtsw.config -n mgmtsw0
  ```

- Configuring one compute node

The following command adds one compute node, named `n1`, to the cluster. The node has an entry in the cluster definition file called `compute.config`. The command is as follows:

```
# cm node add --config-file compute.config -n n1
```

- Configuring multiple compute nodes

  The following command adds ten compute nodes, named `n1` through `n10`, to the cluster. The hostnames for these nodes are `n1` through `n10`. The nodes have entries in the cluster definition file called `compute.config`. Within file `compute.config`, make sure that nodes 1 through 10 are listed sequentially. The command is as follows:

  ```
  # cm node add --config-file compute.config -n n[1-10]
  ```

- Configuring one power distribution unit (PDU)

  The following command adds one PDU, named `pdu1`, to the cluster. The node has an entry in the cluster definition file called `pdu.config`. The command is as follows:

  ```
  # cm node add --config-file pdu.config -n pdu1
  ```

NOTE: After you add the nodes to the cluster, you can provision the nodes with an image.

(Conditional) Configuring cooling components

About this task

HPE Cray XD clusters and HPE Apollo clusters without leader nodes can use HPE Adaptive Rack Cooling Systems (ARCS) components. With these types of clusters, you can use cluster manager tools to view cooling component alerts. Complete the following procedure to enable viewing of cooling component alerts:

Procedure

Configuring an HPE Adaptive Rack Cooling System (ARCS) component

Configuring an HPE Adaptive Rack Cooling System (ARCS) component

About this task

After this procedure is complete, the ARCS component is enabled in the power and cooling infrastructure manager (PCIM). You can use PCIM to monitor the cooling components. For more information about PCIM, see the following:

HPE Performance Cluster Manager Administration Guide

Procedure

1. Log in as the root user to the admin node.
2. Obtain the MAC address of the ARCS component.

   If necessary, complete the procedure in the following topic, and return here when you have the MAC address:

   Using the `switchconfig` command to determine the MAC address for a cooling component
3. Enable the ARCS component.

   Use the `cm cooldev arcs add` command in one of the following formats to enable the ARCS component:

   - Format 1 - Adds the ARCS component to the cluster based on its MAC address:

     `cm cooldev arcs add -m component_mac_addr -n hostname [-i ip_addr]`

     Use this command format the first time an ARCS component is added to the cluster. This command requires you to provide the MAC address and a hostname.
   - Format 2 - Adds the ARCS component to the cluster using a previously assigned IP address:

     `cm cooldev arcs add -n hostname -i ip_addr`

     Use this format, if the IP address was statically configured, is reachable, and is active on the ARCS component.

   The variables are as follows:

Variable	Specification
component_mac_addr	The MAC address of the component. If the command fails to configure the MAC address you specify, see the `cm cooldev cdu add` help output for information about specifying the `--Interface` *`NIC`* parameter.
hostname	The hostname that you want to assign to the cooling component, or the hostname that is active on the component.
ip_addr	In Format 1, you can specify an IP address, as follows: • If you specify an IP address, make sure it is an active IP address. Such an IP address might have been assigned statically. • If you do not specify an IP address, the cluster manager assigns an IP address, configures that IP address in DHCP, and enables the ARCS component to obtain that IP address. In Format 2, you do not specify the cooling component MAC address, so specify a statically assigned *ip_addr* address. This IP address is required to be active. In this case, it is assumed that the MAC address is already in the cluster database. You might use this format for a reinstallation or if you need to add the ARCS component to the cluster again after a maintenance period or outage.

For more information about the commands to add, delete, or display ARCS components, see the manpages for these commands or enter one or more of the following:

```
# cm cooldev arcs -h
# cm cooldev arcs add -h
# cm cooldev arcs delete -h
# cm cooldev arcs show -h
```

4. Repeat the preceding steps for each additional ARCS component as needed.

Using the `switchconfig` command to determine the MAC address for a cooling component

Procedure

1. Log into the admin node as the root user.
2. Obtain network information for the cluster or plan to visually inspect the components and cabling.

   Proceed as follows:

   • If you have network information, such as the spreadsheet used for the cluster when it was manufactured at the factory, proceed to Step **3**.
   • If you do not have network information, you need to visually inspect the cluster. Proceed to Step **4**.

3. Examine the network information for the cluster.

If the cluster was assembled at the factory, a network spreadsheet is available. If necessary, contact your HPE representative to obtain a copy. From the spreadsheet, determine the following:

- The hostname of the switch into which the cooling component is plugged.
- The switch port for the cable that attaches the cooling component to the cluster.

Proceed to Step **7**.

4. Enter the following command to retrieve the hostnames for all the switches in the cluster:

```
# cm group system show mgmt_switch
mgmtsw0
mgmtsw1
mgmtsw100
mgmtsw101
mgmtsw102
mgmtsw103
mgmtsw104
mgmtsw105
mgmtsw2
```

This command shows you how many switches are in the cluster and the hostnames of the switches. You might find this information useful when completing the rest of the steps in this procedure.

5. Check the labels on the cables going into each switch.

Example labels are in the **Cable label** column of the following table:

Cable label	Orientation	Derived hostname
SW0A	Top switch, ports 1/0/X	`mgmtsw0`
SW0B	Bottom switch, ports 2/0/X	`mgmtsw0`
SW1A	Top switch, ports 1/0/X	`mgmtsw1`
SW1B	Bottom switch, ports 2/0/X	`mgmtsw1`

As you can see, the you can derive the hostname for each switch by examining the labels on the cables.

6. Find the cable that connects the switch and the cooling unit.

Note the port number on the switch that the cable plugs into.

7. Enter the `switchconfig` command in the following format:

```
switchconfig info -s mgmtsw --fdb
```

For *mgmtsw*, specify the hostname of the management switch that the cooling component is plugged into.

For example:

```
# switchconfig info -s mgmtsw1 --fdb
```

8. Analyze the output from the `switchconfig` command.

In the `switchconfig` command output, find the line for the cooling component port in the switch.

For example, assume that the cooling component is plugged into switch port 12. In the following output, the line for port 12 is highlighted. The information for the MAC address is in column 1. Properly formatted, the MAC address is `78:04:73:2f:a7:13`.

```
# switchconfig info -s mgmtsw1 --fdb
==== L2 FDB(mac-address-table) Table Information on mgmtsw1 ====

Running command - `display mac-address`...

        MAC Address       VLAN ID    State           Port/NickName           Aging
        2067-7ce4-f31c    1          Learned         GE1/0/7                 Y
        2067-7ce4-f336    1          Learned         GE1/0/3                 Y
        2067-7ce4-f34c    1          Learned         GE1/0/5                 Y
        48df-3787-a820    1          Learned         BAGG125                 Y
        48df-3787-d080    1          Learned         BAGG125                 Y
        48df-3789-4590    1          Learned         BAGG125                 Y
        7804-732f-a713    1          Learned         GE1/0/12                Y
        98f2-b3ea-244f    1          Learned         BAGG111                 Y
        d4c9-efcf-b186    1          Learned         BAGG111                 Y
        ec9b-8b60-7ea6    1          Learned         BAGG125                 Y
        ec9b-8b60-7eb0    1          Learned         BAGG125                 Y
        ec9b-8b60-7ea6    1998       Learned         BAGG125                 Y
        ec9b-8b60-7ebd    1998       Learned         BAGG125                 Y
```

(Conditional) Configuring power distribution units (PDUs) into the cluster

About this task

PDUs distribute AC power to the cluster components. PDUs are optional. The cluster manager requires you to configure the PDUs as a separate task. Use the information in this procedure to configure the PDUs into the cluster.

On HPE Cray XD clusters, on HPE Apollo clusters, and on SGI Rackable clusters, the PDUs reside in each rack. For these clusters, and for all other clusters with PDUs that reside in racks, include the PDUs in the cluster definition file.

For example, assume that you need to include a definition for `pdu0`. A line such as the following in the cluster definition file configures the PDU numbered `pdu0`:

```
internal_name=pdu0, mgmt_bmc_net_name=head-bmc,
geolocation="cold aisle 4 rack 1 B power",
mgmt_bmc_net_macs=99:99:99:99:99:99,
hostname1=testpdu0
```

To enable PDU monitoring, configure the `pdu-collect` service. For information about how to configure the `pdu-collect` service, see the following:

HPE Performance Cluster Manager System Monitoring Guide

Procedure

1. Use a text editor to create a file for the PDUs.

   For example, create file `pdu.config`.

   If you have a cluster definition file that includes PDU information, copy the PDU information from the cluster definition file into the PDU-specific file, and proceed to the following step:

   Step **4**

2. Include the following information in this file:

   - Specify the network upon which the PDU resides. For example, `head-bmc`, which specifies the head BMC network.
   - You can specify a geographic location setting. To add a text string that points to the physical location of a PDU, use the `geolocation=` parameter. For example:
     - `hot aisle 3 rack1 A power`
     - `cold aisle 4 rack 1 B power`

     The text string can include spaces and special characters. If you include spaces, enclose the string in quotation marks (").

     If you have multiple PDUs, multiple clusters, or multiple racks, this setting can be helpful. The geolocation setting is optional.

   For example, create a file that includes information similar to the following:

   ```
   internal_name=pdu0, mgmt_bmc_net_name=head-bmc,
   geolocation="cold aisle 4 rack 1 B power",
   ```

```
mgmt_bmc_net_macs=99:99:99:99:99:99,
hostname1=testpdu0
```

3. Save and close the file.
4. Use the `cm node add` command to configure the PDUs into the cluster.

   The format is as follows:

   ```
   cm node add -c cluster_definition_file_for_PDUs
   ```

   For *cluster_definition_file_for_PDUs*, specify the name of your cluster definition file.

   For example:

   ```
   # cm node add -c pdu.config
   ```

Configuring compute nodes that are not under the control of a leader node

About this task

Use the commands in this chapter to add compute nodes that do not reside in a chassis. These might be extra compute nodes deployed with user services. If a compute node resides in a chassis, use the `cmcinventory` command to add them to the cluster.

You can use the procedures in this chapter later if you add nodes or components to the cluster.

Procedure

1. Enter the following command, examine the output, and verify that all compute nodes have been added to the cluster:

   ```
   # cm node show
   ```

   If a compute node resides in a chassis, it should appear in the command output. If a node that resides in a chassis does not appear in the command output use the `cmcinventory` service to add the node into the cluster.

   If a compute node does not appear in the command output because it is not yet configured into the cluster, continue with this procedure. This is the case for nodes that are not under the control of a leader node. For example, this is the case for compute nodes deployed as login nodes.

2. Use one or both of the following procedures to configure compute nodes into the cluster:

   - **Configuring compute nodes with a cluster definition file and the `cm node add` command**. Use this command if you have a cluster definition file that includes the compute nodes.
   - **Configuring compute nodes without a cluster definition file by using the `cm node discover` command**. Use this procedure if you do not have a cluster definition file that includes the compute nodes.

Configuring compute nodes with a cluster definition file and the `cm node add` command

About this task

The `cm node add` command adds components, such as compute nodes or racks of multiple compute nodes, to a cluster. You can use this command to add many types of cluster components, but this topic specifically addresses compute nodes.

The command in this topic assumes that you have a cluster definition file that includes the following information for each compute node:

- The MAC address for the NIC
- The MAC address for the node controller
- The node controller credentials

For more information about the parameters to this command, enter the following:

```
cm node add -h
```

Procedure

1. Obtain or create a cluster definition file that includes compute node information.

   Include configuration attributes for the MAC addresses, IP addresses, and other information.

   For example, assume that `computes.config` is a cluster definition file with the following contents:

   ```
   hostname1=n1,mgmt_bmc_net_macs=00:11:22:33:44:44,mgmt_net_macs=00:11:22:33:44:45,\
   mgmt_net_ip=172.23.1.1,mgmt_bmc_net_ip=172.24.1.1,mgmt_net_name=head,mgmt_bmc_net_name=head-bmc,card_type=iLO,\
   bmc_username=admin,bmc_password=admin,baud_rate=115200,mgmt_net_bonding_mode=active-backup,mgmt_net_interfaces=eno1,\
   redundant_mgmt_network=no,rootfs=disk,conserver_logging=yes,console_device=ttyS0,dhcp_bootfile=grub2,transport=udpcast,\
   switch_mgmt_network=yes
   ```

2. Enter the following command:

   `cm node add -c` *`cluster_definition_file_for_new_nodes`*

   For *cluster_definition_file_for_new_nodes*, specify the name of your cluster definition file.

   For example:

   ```
   # cm node add -c computes.config
   ```

3. Use the `cm node provision` command to provision the new compute nodes with an image and (optionally) to power cycle the new compute nodes.

Configuring compute nodes without a cluster definition file by using the `cm node discover` command

About this task

The `cm node discover` command can configure compute nodes into the cluster without the use of a cluster definition file.

This command assumes the following:

- You do not have a cluster definition file that includes the nodes you want to add.
- The compute nodes are capable of being PXE booted.
- For the nodes you want to add, you do not know the MAC addresses of the node controllers or the MAC addresses of the NICs . If you know the MAC address information for the nodes you want to add, use the `cm node add` command to add the node.

Whether you have the MAC addresses or not, you can use `cm node discover` to set the node controller credentials. This command PXE boots a small operating system on the node to gather node information and (optionally) set credentials.

The `cm node discover` command guides you through an automated, incremental process for building a cluster definition file for adding new nodes to the cluster.

For more information about the parameters to this command, enter the following:

```
# cm node discover -h
```

To display help for the steps in this process, enter the following command:

```
# cm node discover help
```

Procedure

1. Verify that the new compute nodes are cabled and plugged in.
2. Log into the admin node as the root user.
3. Enter the following command to create a pool of IP addresses, with a short lease time, in the DHCP service:

   ```
   # cm node discover enable
   ```

   If necessary, specify additional parameters. For example, you can specify the following:

   - A specific subnet for the pool of IP addresses.
   - A specific miniroot for operating system discovery.

4. Manually press the power-on button for each of the new compute nodes.

   As each compute node powers up, the cluster manager grants a leased IP address from the pool, and the miniroot environment boots.

5. Enter the following command and observe the leased IP address information:

   ```
   # cm node discover status
   ```

   This command lists all the leased IP addresses and uses `ssh` to connect to each of these leased IP addresses. The command is trying to detect whether the nodes have PXE booted the cluster manager miniroot operating system. When the `ssh` attempt is successful, the cluster is in contact with the new compute node.

6. Make sure that the `cm node discover status` command shows all the nodes you want to add.

   Do not proceed to the next step until all nodes are shown in the output.

7. Enter the `cm node discover mkconfig` command, in a format similar to the following, to generate a cluster definition file for the new nodes:

   ```
   cm node discover mkconfig -o "bmc_username=uname, bmc_password=pwd"
   cluster_definition_file
   ```

   The variables are as follows:

Variable	Specification
uname	The BMC username you want to assign to the node controllers.
pwd	The BMC password you want to assign to the node controllers.
cluster_definition_file	The name for the output file, which becomes the cluster definition file for these nodes. For example, `computes.config`.

   The BMC credentials are required. This command creates a cluster definition file with very minimal entries for each new node. To add other common settings per node, expand the content in the `-o` option. For example, to configure the console to be `ttyS1`, change the `-o` option to the following:

   ```
   -o “bmc_username=username, bmc_password=password, console_device=ttyS1”
   ```

   For more information about the settings you can include on the `-o` option, see the following:

   Specifying configuration attributes

8. (Optional) Add node-specific settings in the cluster definition file.

At this point, you have a cluster definition file. If you want to specify node-specific settings, edit the cluster definition file now.

9. Enter the `cm node discover add` command, in the following format, to add the new compute node to the cluster manager database:

   `cm node discover add [-s] [-i image] [-d disk] cluster_definition_file`

   This command adds the new nodes and resets the node controllers so that they pick up appropriately configured IP addresses.

   The parameters and variables are as follows:

Parameter or variable	Specification
`-s`	Specify the `-s` parameter if the BMC credentials in the cluster definition file need to be configured in the BMC. If the BMC credentials are not configured in the BMC, this option is not needed.
`-i image`	The image you want to assign to the compute nodes. If you specify an image, the command reboots the nodes and provisions the nodes with the specified image. Otherwise, by default, this command powers off the nodes, which postpones provisioning. If you do not specify the `-i` option, the cluster manager powers down the nodes. You can use the `cm node provision` command to deploy an image to the nodes.
`-d disk`	Specify the `-d disk` parameter if you also specify the `-i image` parameter. For *disk*, specify the disk to install the *image*. The default is `/dev/sda`.
cluster_definition_file	The name of the cluster definition file for these nodes, which you created in the following step: Step **7**

10. Enter the following command to delete the pool of IP addresses from the DHCP service:

    `# cm node discover disable`

(Conditional) Adding controllers manually

About this task

The cluster manager adds most types of controllers to the cluster database automatically. However, the cluster manager does not add the following controllers or components to the database automatically:

- An external HPE Slingshot interconnect switch controller
- A Gigabyte chassis controller

As a troubleshooting tactic, you can also use the `cm controller` command to delete and then to add a misconfigured controller. Use `cm controller delete` to delete the misconfigured controller and then `cm controller add` to add the controller back in correctly.

If your cluster contains any of the preceding controller types, complete the procedure in this topic to add the controllers manually.

Procedure

1. Use the `cm controller add` command to configure the controller into the cluster database.

   The format of this command is as follows:

   ```
   cm controller add -c hostname -t controller_type -m mac_address -u username -p password
   ```

   The variables are as follows:

Variable	Specification
hostname	The hostname you want to assign to the controller.
controller_type	Enter one of the following keywords depending on the type of controller you want to add: • `external_switch`. Use this keyword for an external HPE Slingshot interconnect switch controller. • `gigabyte`. Use this keyword for a Gigabyte chassis controller.
mac_address	The MAC address of the controller.
username	The username used to log into the controller.
password	The password used to log into the controller.

2. Enter the `cm controller show` command to display the information for the controller you just added.

   The format for this command is as follows:

   ```
   cm controller show -c hostname
   ```

   For *hostname*, enter the hostname of the controller you just added.

For example:

```
# cm controller show -c x9000c1r3b0
NAME        TYPE                 ADMINISTRATIVESTATUS PROTOCOL CHANNEL MACADDRESS        IPADDRESS  IPV6ADDRESS
x9000c1r3b0 cmm_switch_controller online               None     None    XX:XX:XX:XX:XX:XX XX.XXX.X.X None
```

3. Repeat the preceding steps to configure all controllers into the cluster.

NOTE: If you have many controllers, you can create a file with controller information and specify that file as an argument to the following command:

`cm node add -c input_file`

This single command adds multiple controllers. For more information, enter the following command:

`# cm node add -h`

Using the `cm controller add` command

The `cm controller add` command adds an external switch controller or a Gigabyte controller to the cluster database. For more information, enter the following command:

`# cm controller add -h`

Using the `cm controller show` command

The `cm controller show` command displays information for all controllers of all types.

If you enter the command without any arguments, it displays all the controllers in the cluster. For example::

```
# cm controller show

NAME        TYPE                          ADMINISTRATIVESTATUS PROTOCOL                  CHANNEL MACADDRESS
x9000c1r3b0 cmm_switch_controller         online               None                      None    XX:XX:XX:XX:XX:XX
x9000c1r7b0 cmm_switch_controller         online               None                      None    XX:XX:XX:XX:XX:XX
x9000c1s0b0 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s0b1 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s1b0 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s1b1 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s2b0 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s2b1 cmm_node_controller           online               None                      None    XX:XX:XX:XX:XX:XX
x9000c1s3b0 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c1s3b1 cmm_node_controller           online               Cray,NO_IPMI,None,redfish None    XX:XX:XX:XX:XX:XX
x9000c3r3b0 cmm_switch_controller         online               None                      None    XX:XX:XX:XX:XX:XX
x9000c3r7b0 cmm_switch_controller         online               None                      None    XX:XX:XX:XX:XX:XX
x9000cec0   cabinet_environment_controller online               None                      None    None
x9000cec1   cabinet_environment_controller online               None                      None    None
```

NOTE: The preceding output was truncated from the right for inclusion in this documentation.

Using the `cm controller delete` command

The `cm controller delete` command deletes a controller from the cluster database. For more information about this command, enter the following:

`# cm controller delete -h`

Backing up the cluster

About this task

For information about how to back up the cluster, see the following:

HPE Performance Cluster Manager Administration Guide

At this time, make sure to back up the admin node and the cluster configuration files.

Whenever you make significant changes to the cluster configuration, back up the cluster.

Configuring additional features

The cluster manager includes features that you might have to configure depending on your components. Additionally, there are features that are not required but might be of use on your system.

NOTE: If you add or change anything on your cluster, remember to back up the cluster again.

For information about how to back up the cluster, see the following:

HPE Performance Cluster Manager Administration Guide

Configuring monitoring

About this task

For information about how to configure cluster monitoring, see the following:

HPE Performance Cluster Manager System Monitoring Guide

Configuring the GUI on a client system

About this task

You can configure the GUI on a client computer outside of the cluster system. For example, you can install the client software on a laptop computer.

For information about the client software required and how to start the GUI, see the following:

HPE Performance Cluster Manager Administration Guide

Starting the cluster manager web server on a non-default port

Procedure

1. On the admin node, use a text editor to adjust the settings in the following file:

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

2. Open the corresponding ports in the firewall.

Customizing nodes

You can use post-installation scripts to customize operations on compute nodes. The scripts can enable additional software, append data to configuration files, configure supplemental network interfaces, and perform other operations. For information about these scripts, see the following file:

`/opt/clmgr/image/scripts/post-install/README`

Naming the storage controllers for clusters with a system admin controller high availability (SAC HA) admin node

About this task

Complete the procedure in this topic if the cluster has a SAC HA admin node.

The following procedure configures names for the storage controllers. The names enable you to manage them from the admin node.

Procedure

1. Log into the admin node as the root user.
2. From the admin node, enter the following commands:

   ```
   # cm node add --node-def='hostname1=unita,internal_name=service100,mgmt_net_macs=00:50:B0:AB:F6:EE,
   generic' --skip-switch-config --skip-refresh-netboot
   # cm node add --node-def='hostname1=unitb,internal_name=service101,mgmt_net_macs=00:50:B0:AB:F6:EF,
   generic' --skip-switch-config --skip-refresh-netboot
   ```

   The commands in this step accomplish the following:

   - The commands configure hostnames and IP addresses for the storage controllers. These host names are `unita` and `unitb`.
   - The commands configure DHCP so that the storage devices automatically receive an IP address.

Adjusting the domain name service (DNS) search order

A DNS search path lists the order of subdomains to try when you (or a program) need to translate a hostname into an IP address.

If you use DNS as the method to convert hostnames into IP addresses, you can configure the following:

- A specific subdomain is the first IP address to be resolved. In addition, you can specify more than one subdomain and the order in which each subdomain is to be searched.
- A DNS resolution specification that applies to the cluster globally or only for a specific node.

The following are examples of subdomains that you can specify:

- HPE Slingshot interconnect IP addresses. For example, `hsn0.cm.clusterdomain.com` or `hsn1.cm.clusterdomain.com`.
- InfiniBand fabric IP addresses. For example, `ib0.cm.clusterdomain.com` or `ib1.cm.clusterdomain.com`.
- Management fabric IP addresses. For example, `head.cm.clusterdomain.com`, `hostmgmt.cm.clusterdomain.com`, or `gbe.cm.clusterdomain.com`.
- Public or external IP addresses. For example, `cm.clusterdomain.com` or `public.clusterdomain.com`.

The cluster manager sets the DNS search order after you run the cluster configuration tool. However, you can change the domain search order at any time after the cluster is installed and configured.

For more information, see the `resolv.conf` manpage.

The following topics include information about how to analyze, view, or configure search order:

- **Analyzing your environment**
- **Configuring the DNS search order**
- **Retrieving the DNS search order**

Analyzing your environment

Sometimes a host includes multiple network interfaces.

A command that does not specify the subdomain of `.gbe` or `.ib0` uses the DNS search path to determine the IP address to return, as follows:

- The host lookup command returns the `ib0` IP address when the DNS search path is one of the following:
  - `ib0.cm.clusterdomain.com cm.clusterdomain.com`

    or
  - `ib0.cm.clusterdomain.com gbe.cm.clusterdomain.com cm.clusterdomain.com`
- The host lookup command returns the `gbe` IP address when the search path is one of the following:
  - `gbe.cm.clusterdomain.com cm.clusterdomain.com`

    or
  - `gbe.cm.clusterdomain.com ib0.cm.clusterdomain.com cm.clusterdomain.com`
- If neither `ib0` nor `gbe` are in the DNS search path, the host lookup command returns the first entry in the DNS configuration file.

  When searching, specify the subdomains in the same search order as the domains are defined.

The DNS search order is more important when nodes with different interfaces try to reach each other. For example, if the admin node does not have an `ib0` interface, `gbe` needs to be first in the DNS search path for the admin node itself.

If IP address information for a node is in the `hosts` file, the system ignores the DNS search path.

The following topics explain how to view or configure the global or per-node search order:

- **Configuring the DNS search order**
- **Retrieving the DNS search order**

Configuring the DNS search order

Procedure

1. Log into the admin node as the root user.
2. Use the following `cm node set` command to set the DNS resolution order:

   `cm node set [-g] [-n` *`node`*`] --domain-search-path` *`new_domain_search_path`*

   The variables are as follows:

Variable or parameter	Specification
`-g`	Conditional. Use when you want to configure the global search order.
node	Conditional. Use when you want to configure the search order for one node. Specify the node hostname.
new_domain_search_path	One or more domains to search. If you specify more than one domain, the cluster manager searches the domains in the order specified. Use a comma (`,`) character to separate domains.

Example 1. The following command sets a global domain search path:

```
admin:~ # cm node set -g --domain-search-path ib0.cm.clusterdomain.com.com,head.cm.clusterdomian.com
```

Example 2. The following command sets the domain search path for `n0`:

```
admin:~ # cm node set -n n0 --domain-search-path head.cm.clusterdomain.com,ib0.cm.clusterdomain.com
```

Retrieving the DNS search order

Procedure

1. Log into the admin node as the root user.
2. Use the following `cadmin` command to show the DNS search order:

   ```
   cm node show --domain-search-path [-n node]
   ```

   For *node*, specify a node hostname. Specify this optional parameter when you want to retrieve the search path for a specific node. Do not specify this parameter if you want to retrieve the global domain search path.

   Example 1. The following command retrieves the global domain search path:

   ```
   # cm node show -g --domain-search-path
   ib0.cm.clusterdomain.com,head.cm.clusterdomain.com
   ```

   Example 2. The following command retrieves the domain search path for one node, `n0`:

   ```
   # cm node show --domain-search-path -n n0
   head.cm.clusterdomain.com,ib0.cm.clusterdomain.com
   ```

Configuring a back-up domain name service (DNS) server

About this task

Typically, the DNS on the admin node provides name services for the cluster. If you configure a backup DNS, the cluster can use a compute node as a secondary DNS server when the admin node is unavailable. You can configure a backup DNS only after the cluster is configured completely. This feature is optional.

The following procedure explains how to configure a compute node to act as a DNS.

Procedure

1. Through an `ssh` connection, log into the admin node as the root user.
2. Enter the following command to retrieve a list of available compute nodes:

   ```
   # cnodes --compute
   ```

   The preceding command lists all nodes that are classified as compute nodes, so the list includes fabric management nodes. Select a compute node for use as the backup DNS. Do not select a fabric management node for the backup DNS.
3. Enter the following command to start the cluster configuration tool:

   ```
   # /opt/sgi/sbin/configure-cluster
   ```
4. On the **Main Menu** screen, select **D Configure Domain Name System (DNS)**, and select **OK**.
5. On the **Domain Name System (DNS) Menu** screen, select **B Configure Backup DNS Server (optional)**, and select **OK**.
6. On screen that appears, enter the identifier for the compute node that you want to designate as the backup DNS, and select **OK**.

   For example, you could configure compute node `n101` as the host for the backup DNS server.

   To disable this feature, select **Disable Backup DNS** from the same menu and select **Yes** to confirm your choice.

Setting a static IP address for the node controller in the admin node

About this task

Complete the procedure in this topic if one or both of the following are true:

- Your site practices require a static IP address for the node controller.
- You want to configure a high availability (HA) admin node. In this case, perform this procedure on the node controllers on each of the two admin nodes.

When you set the IP address for the node controller on the admin node, you ensure access to the admin node when the site DHCP server is inaccessible.

The following procedures explain how to set a static IP address.

Method 1 -- To change from the BIOS

Use the BIOS documentation for the admin node.

Method 2 -- To change the IP address from the admin node

Procedure

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

   ```
   # ipmitool lan print 1
   ```
3. In the output from the preceding command, look for the `IP Address Source` line and the `IP Address` line.

For example:

```
IP Address Source       : DHCP Address
IP Address              : 192.168.2.59
```

Note the IP address in this step and decide whether this IP address is acceptable. The rest of this procedure explains how to keep this IP address or to set a different static IP address.

4. Enter the following command to specify that you want the node controller to have a static IP address:

```
# ipmitool lan set 1 ipsrc static
```

The command in this step has the following effect:

- The command specifies that the IP address on the node controller is a static IP address.
- The command sets the IP address to the IP address that is currently assigned to the node controller.

To set the IP address to a different IP address, proceed to the following step. If the current IP address is acceptable, you do not need to perform the next step.

5. (Conditional) Reset the static IP address.

Complete this step to set the static IP address differently from the current IP address. Enter `ipmitool` commands in the following format:

```
ipmitool lan set 1 ipaddr ip_addr
ipmitool lan set 1 netmask netmask
ipmitool lan set 1 defgw gateway
```

The variables are as follows:

Variable	Specification
ip_addr	The IP address you want to assign to the node controller.
netmask	The netmask you want to assign to the node controller.
gateway	The gateway you want to assign to the node controller.

For example, to set the IP address to 100.100.100.100, enter the following commands:

```
# ipmitool lan set 1 ipaddr 100.100.100.100
# ipmitool lan set 1 netmask 255.255.255.0
# ipmitool lan set 1 defgw 192.168.8.255
```

6. (Conditional) Repeat the preceding steps on the second admin node.

Complete this procedure again only if you want to configure a second admin node for a two-node high availability cluster.

Configuring Array Services for HPE Message Passing Interface (MPI) programs

About this task

You can configure compute nodes into an array. After you configure a set of nodes into an array, the Array Services software can perform authentication and coordination functions when HPE Message Passing Interface (MPI) programs are running. For more information, see the following:

HPE Message Passing Interface (MPI) User Guide

You can include the admin node in an array.

For general Array Services configuration information, see the manpages. The Array Services manpages reside on the admin node. If the HPE Message Passing Interface (MPI) software is installed on the admin node, you can retrieve the following manpages:

- `arrayconfig`(1M), which describes how to use the `arrayconfig` command to configure Array Services.
- `arrayconfig_smc`(1M), which describes Array Services configuration characteristics that are specific to clusters.

The procedures in the following topics assume the following:

- You want to create new a master image for the compute nodes configured for computing.

  And

- You want to configure a new master image for the compute nodes configured for user services.

After you create the images, you can push out the new images.

The alternative is to configure Array Services directly on the nodes themselves. This method, however, leaves you with an Array Services configuration that is overwritten the next time someone pushes new software images to the cluster nodes.

Procedure

1. **Planning the configuration**
2. **Preparing the Array Services images**
3. Complete one of the following:
   - **(Conditional) Permitting remote access to the service node**

     Or

   - **(Conditional) Preventing remote access to the service node**
4. **Distributing images to all the nodes in the array**
5. **Power cycling the nodes and pushing out the new images**

Planning the configuration

About this task

The following procedure explains how to plan your array and how to select a security level.

Procedure

1. Log into the admin node as the root user.
2. Verify that the HPE MPI is installed on the cluster.

   If HPE MPI is not installed on the admin node, complete the following steps:

   - On RHEL systems, enter the following command:

     ```
     # cm node dnf -n admin 'groupinstall HPE*MPI'
     ```

   - On SLES systems, enter the following command:

     ```
     # cm node zypper -n admin 'groupinstall HPE*MPI'
     ```

3. Use the `cm node show` command to display a list of available nodes, and decide which nodes you want to include in the array.

   For example:

   To display information about compute nodes, enter the following command:

   ```
   # cm node show -t system compute
   ```

   The command output includes information about nodes that might be configured as service nodes at this time.

4. Display a list of the available system images, and decide which images you want to edit.

   For example, the following output is for an example cluster running in production mode:

   ```
   # cm image show
   sles15spX                    # original, factory-shipped system image
   sles15spX.prod1              # customized image for this cluster
   ```

   The output includes image `sles15spX.prod1`. The `sles15spX.prod1` image is installed on a compute node that is configured as a service node. Image `sles15spX.prod1` is based on image `sles15spX`, but it can include software to support user logins and a backup DNS server.

   All system images are stored in the following directory:

   `/opt/clmgr/image/images`

   For each of these images, the associated kernel is `3.0.101-94-default`.

   The examples in this Array Services configuration procedure add the Array Services information to the customized, production images with the `.prod1` suffix.

5. Decide what kind of security you want to enable.

   Array Services includes its own authentication and security. If your site requires additional security, you can configure MUNGE security, which the installation includes. Your security choices are as follows:

   - `munge` on all the nodes you want to include in the array. Configures additional security provided by MUNGE. The installation process installs MUNGE by default. If you decide to use MUNGE, the MPI from HPE configuration process explains how to enable MUNGE at the appropriate time.
   - `none` on the service nodes and `none` on the compute nodes

     or

     `noremote` on the service nodes and `none` on the compute nodes

These specifications have the following effects:

- When you specify `none` on all the nodes you want to include in the array, all authentication is disabled.
- When you specify the following, users must run their jobs directly from the service nodes:
  - `noremote` on the service nodes

    And
  - `none` on the compute nodes

  In this case, users cannot submit MPI from HPE jobs remotely.

- `simple` (default). Generates hostname/key pairs by using either the OpenSSL, `rand` command, 64-bit values (if available) or by using `$RANDOM` Bash facilities.

Preparing the Array Services images

About this task

Before you create images that include Array Services, copy the production system images that your system is using now.

The following procedure explains how to prepare the images.

Procedure

1. Log into the admin node as the root user.
2. Use two `cm image copy` commands to clone the following:

   - One of the images that resides on a service node

     And
   - One of the images that resides on a compute node

   The format is as follows:

   `cm image copy -o existing_image -i new_image`

   The variables are as follows:

Variable	Specification
existing_image	The name of one of the existing images.
new_image	The new name for that to want to give to the image.

   For example, the following command copies the first-generation compute node production image to a new, second-generation production image:

   `# cm image copy -o sles15spX.prod1 -i sles15spX.prod2`
3. Enter the following command to change to the system images directory:

   `# cd /opt/clmgr/image/images`
4. (Optional) Use the `cp` command to copy the MUNGE key from the new service node image to the new compute node image.

Complete this step if you want to configure the additional security that MUNGE provides.

The MUNGE key resides in `/etc/munge/munge.key` and must be identical on all the nodes that you want to include in the array. The copy command is as follows:

```
cp /opt/clmgr/image/images/new_service_image/etc/munge/munge.key \
/opt/clmgr/image/images/new_compute_image/etc/munge/munge.key
```

The variables are as follows:

Variable	Specification
new_service_image	The name of the new service node image you created.
new_compute_image	The name of the new compute node image you created.

For example:

```
# cp /opt/clmgr/image/images/sles15spX.prod2/etc/munge/munge.key \
/opt/clmgr/image/images/ice-sles15spX.prod2/etc/munge/munge.key
```

5. Use the following command to install the new image on the service node:

   ```
   cm node provision -n hostname(s) -i new_service_image -s
   ```

   The variables are as follows:

Variable	Specification
hostname(s)	The hostname or hostnames of the service node. This node is the node that you want users to log into when they log into the array.
new_service_image	The name of the new image you created.

   For example, the following command installs the new image on node `n1`:

   ```
   # cm node provision -n n1 -i sles15spX.prod2 -s
   ```

6. Use the `ssh` command to log into the service node from which you expect users to run MPI from HPE programs.

   For example, log into `n1`.

7. Use the `arrayconfig` command to configure the service node and compute nodes into an array.

   You can specify more than one service node.

   The `arrayconfig` command creates the following files on the compute service node to which you are logged in:

   - `/etc/array/arrayd.conf`
   - `/etc/array/arrayd.auth`

   Enter the `arrayconfig` command in the following format:

   ```
   /usr/sbin/arrayconfig -a arrayname -f -m -A method nodes ...
   ```

   The variables are as follows:

Variable	Specification
arrayname	A name for the array. The default is `default`.
method	`munge`, `none`, or `simple`. A later step explains how to specify `noremote` for a service node.
nodes	A list of node IDs.

(Conditional) Permitting remote access to the service node

About this task

Complete this procedure in the following circumstances:

- If you specified `-A munge` or `-A simple` for authentication

  Or

- If you specified `-A none` for authentication, and you want to permit users to log into a service node remotely to submit MPI from HPE programs. The service node is assumed to be a compute node.

The following procedure assumes that you want to permit job queries and commands on the service node. It explains how to copy the array daemon files to the admin node.

Procedure

1. Log into one of the service nodes as the root user.
2. Copy the `arrayd.auth` file and the `arrayd.conf` files from the service node to the new service node image on the admin node.

   Enter the following command:

   ```
   # scp /etc/array/arrayd.* \
   admin:/opt/clmgr/image/images/service_image/etc/arrayd.*
   ```

   For *service_image*, specify the service node image on the admin node.

   Enter this command all on one line. The command in this step uses a backslash (`\`) character to continue the command to the following line.

   For example:

   ```
   # scp /etc/array/arrayd.* \
   admin:/opt/clmgr/image/images/sles15spX.prod2/etc/arrayd.*
   ```

3. Copy the `arrayd.auth` file and the `arrayd.conf` files from the service node to the new compute node image on the admin node.

   Enter the following command:

   ```
   # scp /etc/array/arrayd.* \
   admin:/opt/clmgr/image/images/compute_image/etc/arrayd.*
   ```

   For *compute_image*, specify the compute node image on the admin node. This is a compute node image.

   Enter this command all on one line.

(Conditional) Preventing remote access to the service node

About this task

Complete the procedure in this topic if you specified `-A none` for authentication, and you want to prevent users from logging into a service node remotely to submit MPI from HPE programs

This procedure explains how to prevent a service node from receiving any requests from other computers on the network. In this case, the service node can send requests to all remote nodes, but it does not listen on TCP port 5434 for any incoming requests. Complete the procedure in this topic if this behavior is required at your site.

The following procedure explains how to accomplish the following:

- How to configure the Array Services files to prevent remote access
- How to copy the array daemon files to the admin node

Procedure

1. Log into one of the service nodes as the root user.
2. Open the following file with a text editor:

   `/etc/array/arrayd.auth`
3. Enter the following, all on one line:

   `AUTHENTICATION NOREMOTE`
4. Save and close the file.

   Make sure that the file contains only the one line.
5. Enter the following command to copy `/etc/array/arrayd.auth` and `/etc/array/arrayd.conf` from the service node to the new service node image on the admin node:

   **# scp /etc/array/arrayd.* **
   admin:/opt/clmgr/image/images/*service_image*/etc/arrayd.*

   For example:

   **# scp /etc/array/arrayd.* **
   admin:/opt/clmgr/image/images/sles15spX.prod2/etc/arrayd.*
6. Log into the admin node as the root user.
7. Create file `/opt/clmgr/image/images/`*`compute_image`*`/etc/array/arrayd.auth`.

   For example:

   # vi /opt/clmgr/image/images/sles15spX.prod2/etc/array/arrayd.auth
8. Add the following all on one line:

   `AUTHENTICATION NONE`
9. Save and close the file.

   Make sure that the file contains only the one line.
10. Enter the following command to copy the `/etc/array/arrayd.conf` file to the compute nodes:

    **# scp /etc/array/arrayd.conf **
    admin:/opt/clmgr/image/images/*compute_image*/etc/arrayd.conf
11. Use the `cm image revision commit` command to back up the images.

Distributing images to all the nodes in the array

About this task

The following procedure explains how to complete the following tasks:

- Assign the new service node image to the service nodes
- Assign the new compute node image to the compute nodes

Procedure

1. Log into the admin node as the root user.
2. Assign the new service node image to the service nodes.

   Use one or more `cm node provision` commands in the following format:

   `cm node provision -n` *`hostname`* `-i` *`new_service_image`* `-s`

   The variables are as follows:

Variable	Specification
hostname	Specify the hostname for one or more of the service nodes. In the cluster definition file, this name appears in the `hostname1` field. You can specify the `*` wildcard character to represent a string of identical characters in this field. Use wildcard characters in the following situation: • If you want to specify more than one hostname and • If your nodes have names that are similar For example, if your hostnames are `n1`, `n2`, `n3`, and `n57`, specify `n*` in this field if you want to specify all service nodes.
new_service_image	Specify the name of the new service node image you created.

   Example 1. The following command assigns the new service node image to all service nodes:

   ```
   # cm node provision -n n* -i sles15spX.prod2 -s
   ```

   Example 2. The following command assigns the new service node image to `n101`:

   ```
   # cm node provision -n n101 -i sles15spX.prod2 -s
   ```

Power cycling the nodes and pushing out the new images

Procedure

1. Enter the following command to reboot the service nodes and the compute nodes:

   ```
   # cm power reboot -t node '*'
   ```

2. Use one or more `cm power` commands in the following format to power off the compute nodes that you want to reimage:

   ```
   cm power off -t node 'hostname'
   ```

   For *hostname*, specify one or more compute node hostnames.

   If you have many compute nodes, you can use wildcard characters.

   Issue as many `cm power` commands as needed.

3. (Conditional) Assign the new compute node image to the compute nodes.

   Complete this step if the cluster has compute nodes that use an NFS root file system.

   The following tables contain the instructions you need to complete this step.

   Complete the following steps on clusters with compute nodes that have NFS root file systems:

Step	Task

a.	Run the `cm image activate` command in the following format to activate the NFS image:
	
	`cm image activate -i new_compute_image`
	
	For *new_compute_image*, specify the name of the new compute node image you created.
	
	For example:
	
	`# cm image activate -i sles15spX.prod2`
	`Activate image - Syncing image to RO NFS path`
	`.`
	`.`
	`.`

b.	Assign the new compute node images to the compute nodes. For example, use the following `cimage` command: `# cimage --set sles15spX.prod2 3.0.101-108.38-default "r*i*n*"`

4. Use one or more `cm power` commands in the following format to start the compute nodes:

   ```
   cm power on -t node 'hostname'
   ```

   For *hostname*, specify the hostnames of the compute nodes.

   If you have many compute nodes, you can use wildcard characters.

   For example, the following command powers on all compute nodes:

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

Issue as many `cm power on` commands as needed.

Creating security certificates from a site-specific certificate authority (CA)

About this task

The cluster manager includes a security certificate. By default, the REST API and web server use the security certificate that the cluster manager provides.

If your site has a trusted security certificate from a CA, you can complete the procedure in this topic to regenerate the cluster security certificates based on your site certificate.

Procedure

1. Log into the admin node as the root user.
2. Use a text editor to open the following file:

   `/opt/clmgr/tools/gen-custom-rest-certs`
3. Edit the following fields in the file:

Field	Specification for this cluster
`caCert`	The full path to the CA certificate
`serverKey`	The full path to the server key
`serverCert`	The full path to the server certification

4. Save and close file `gen-custom-rest-certs`.
5. Enter the following command to create new security certificates:

   ```
   # /opt/clmgr/tools/gen-custom-rest-certs
   ```
6. Change to the `custom-certs` directory:

   ```
   # cd custom-certs
   ```
7. Enter the following commands to copy the newly generated certificates, the Java keystore, and the private keys to the appropriate directory:

   ```
   # cp *.pem /opt/sgi/secrets/CA/cert/
   # cp *.p12 /opt/sgi/secrets/CA/private/
   # cp *.key /opt/sgi/secrets/CA/private/
   ```
8. Use a text editor to open the following file:

   `/opt/clmgr/etc/cmuserver.conf`
9. Locate the line that begins with `CMU_JAVA_SERVER_ARGS`, and add the following string within the quotation marks (`" "`):

   `-Djdk.security.allowNonCaAnchor=true`

For example, after adding the new string, the line might look as follows:

```
CMU_JAVA_SERVER_ARGS="-Djdk.security.allowNonCaAnchor=true"
```

If other strings reside within the quotation marks, put this new string at the end.

10. Save and close the `cmuserver.conf` file.
11. Restart the cluster manager:

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

12. Enter the following command to refresh the bootstrap `tar` files on the admin node:

```
# cm node refresh secrets -n admin
```

13. Use the `cm node refresh` command in the following format to refresh the secrets on compute nodes and service nodes:

```
cm node refresh secrets -n nodes
```

For *nodes*, specify the hostnames of the compute nodes and service nodes that need refreshed certificates.

14. Use the `cm power command` in the following format to reboot the compute nodes and service nodes:

```
cm power reboot -n nodes
```

For *nodes*, specify the hostnames of the compute nodes or service nodes upon which you refreshed the secrets.

Troubleshooting cluster manager installations

Troubleshooting configuration changes

If a configuration change does not affect the cluster in the intended manner, try one of the following approaches:

- Edit the node image on the admin node. For example, you can try the following:
  1. On the admin node, reconfigure the image for the compute nodes that you use for user services
  2. Reimage the service nodes with the new, reconfigured image.
- Edit the node customization scripts.

Verifying the switch cabling

About this task

If the switches are not working, the first troubleshooting step is to verify the switch cabling.

The following figure shows a switch stack with two switches. In this switch stack, the two switches constitute the spine switch stack. One is the master switch and the other is the slave switch.

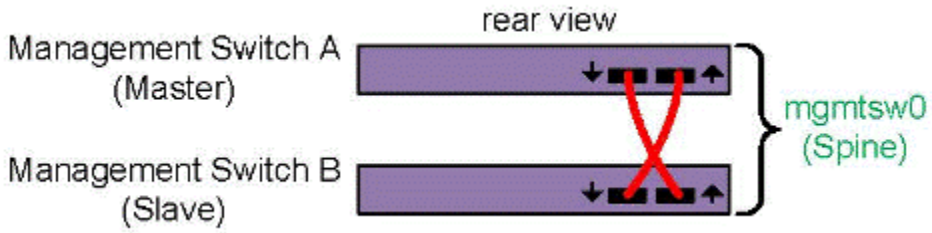

Figure 3: Spine switch stack with two switches

The following figure shows a switch stack with multiple switches. The first two switches constitute the spine switch stack, and the other switches constitute the secondary switch stack.

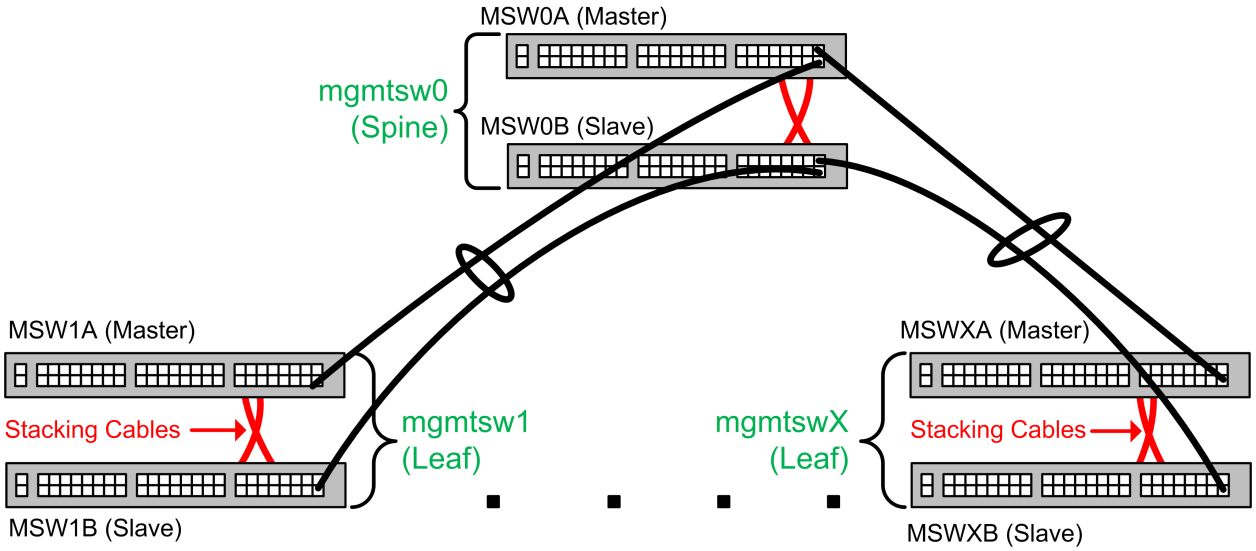

Figure 4: Switch stack with multiple switches

The following procedure explains how to inspect your switches and prepare for the configuration procedure.

Procedure

1. Visually inspect your system.

   Note the types of switches you have and their identifiers. At a minimum, you have at least one stacked (or non-stacked) management switch. The management switch that is connected directly to admin node is almost always considered the **spine** switch. Additional stacked or non-stacked switches connect to the spine switch. These additional switches are almost always considered to be **leaf** switches. In some configurations, leaf switches can connect to other leaf switches, and this creates a **multi-tiered** management network topology.

   When multiple physical switches are in a stacked configuration, those multiple physical switches can be thought of as a single **logical** switch. This means that the logical switch is assigned one IP address for remote management, and the `switchconfig` command can to configure the entire switch stack.

   NOTE: Determine whether your system contains management switches from Arista Networks, Inc. and whether the switches are using Multi-Chassis Link Aggregation (MLAG). In this case, each switch in an MLAG pair is independent and cannot be considered a stacked switch. Each switch in an MLAG pair receives an IP address separately and is managed separately.

   ---

2. Determine whether or not you have a cluster definition file.

   If you have a cluster definition file that contains the MAC addresses of the cluster components, you can safely have all nodes and all switches powered on when you run the node discovery commands. During the node discovery process, a cluster definition file ensures that a node with a MAC address is assigned an IP address that matches the node MAC address.

   If you do not have a cluster definition file, all nodes other than the admin node must begin in a powered off state. Console into the switch, and configure the switch manually.

3. (Conditional) Disconnect the secondary, redundant cables that connect switches together.

   Complete this step if you have not yet configured the management switches into the cluster. Or, complete this step if you plan to reset the management switches back to factory default settings. This action prevents networking loops.

   Use Method 1 or Method 2 to disconnect the switches. The instructions for both methods include an example that assumes that `mgmtsw0` is connected to `mgmtsw1` with the following cable mappings:

- mgmtsw0 port 1/48 ---- mgmtsw1 port 1/48
- mgmtsw0 port 2/48 ---- mgmtsw1 port 2/48

Also assume that mgmtsw1 needs to be reset to factory settings. The reset could be needed to obtain a fresh configuration, to update the VLANs or IP addresses on mgmtsw1, or for any reason.

Method 1 - Software method

If the spine switch is reachable from the admin node, you can prevent a networking loop when mgmtsw1 is factory reset. From the admin node, complete the following steps:

a. Enter the following command to disable the redundant port that connects mgmtsw0 to mgmtsw1

```
# switchconfig port -s mgmtsw0 --disable -p 2/48
```

b. Enter the following command to reset mgmtsw1 back to factory default settings:

```
# switchconfig reset_factory_defaults -s mgmtsw1 --force
```

c. Wait 3~10 minutes for mgmtsw1 to come back online. Enter the following command:

```
# ping mgmtsw1
```

d. Enter the following command to reconfigure **mgmtsw1**:

```
# switchconfig_configure_node --node mgmtsw1
```

Wait for this command to complete.

e. After mgmtsw1 is configured correctly, enter the following command to re-enable the redundant port that you disabled earlier in this procedure.

```
# switchconfig port -s mgmtsw0 --enable -p 2/48
```

f. (Conditional) Reapply lost configuration attributes.

Complete this step if, for example, mgmtsw1 had any nodes that require a switch configuration that was lost in this procedure.

For example, these nodes might be compute nodes that use 802.3ad (LACP) bonding.

To reapply any lost configuration settings, use commands such as the following:

```
# switchconfig_configure_node --node service100
```

Method 2 - Manual method

If you need to reset all management switches on your cluster or have lost full connectivity to the management fabric, you need physical access to the cluster hardware. This method, Method 2, is the same as Method 1, but the initial step is different. Rather than use the switchconfig command to disable ports, start the procedure by doing one of the following to replace Step a:

- Unplug all redundant switch cabling from one end of the wire for all cabling between management switches and for all cabling between management switches and chassis controllers.

  OR

- Attach a serial connection to a management switch, open up a serial console session, and use the vendor-specific methods to temporarily disable the redundant ports until switches can be successfully configured again.

The following figure shows an example topology with 3 management switches (1 spine switch stack and 2 leaf switch stacks) and which cables to disconnect.

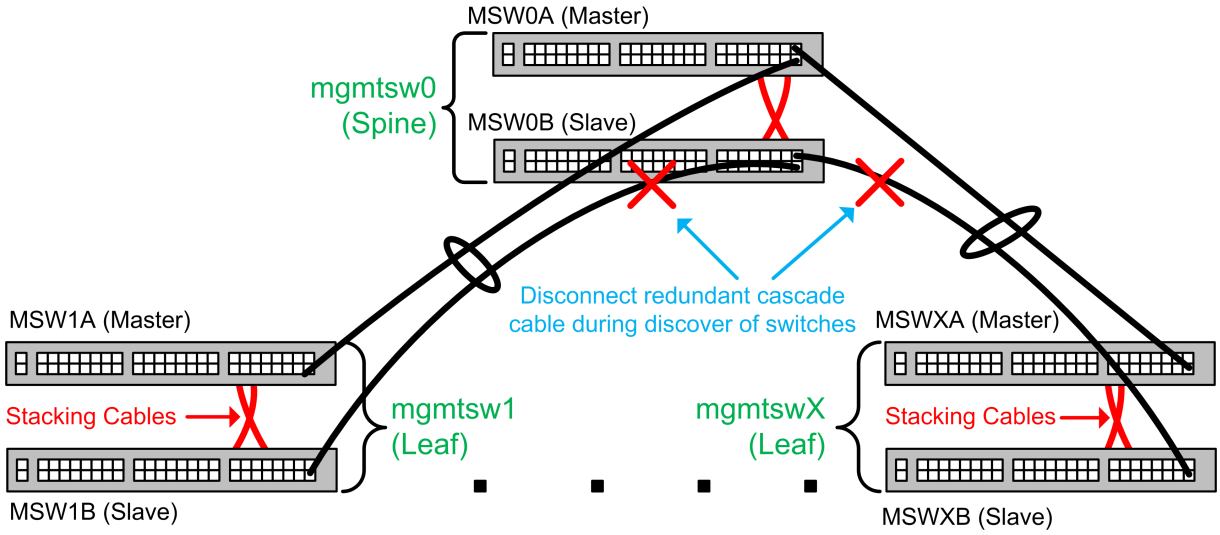

Figure 5: Cables to disconnect

Chassis controllers failed to configure

If you suspect that the chassis controllers failed to configure automatically, look in the following log file for information regarding the status of chassis controllers in the system:

`/opt/clmgr/log/cmcdetectd.log`

If you find errors in the preceding log files, power on the chassis controllers and complete the following steps to configure the chassis controllers:

1. **Reviewing the chassis controller configuration**
2. Complete one of the following procedures:
   - **Method 1 - Configuring the chassis controller switches manually**
   - **Method 2 - Configuring the chassis controller switches manually**

`cmcdetectd` daemon

The `cmcdetectd` daemon runs on the admin node. This daemon uses a specific `tcpdump` command to listen to DHCP packets generated by the chassis controller on the management network. When the `cmcdetectd` daemon receives a DHCP packet generated by the chassis controller, it takes the following actions:

- It inspects the information located in the packet.
- It determines the appropriate VLAN and bonding settings to apply to the attached management switch ports connected to the chassis controller in question.

Chassis controllers are cabled in the same manner as other redundantly cabled components in the cluster. For example, assume that rack 1, chassis controller 1, port 1 is connected to `mgmtsw0` port 1/0/11. In this case, rack 1, chassis controller 1, port 2 must be connected to `mgmtsw0` port 2/0/11, and so on.

Reviewing the chassis controller configuration

About this task

The following procedure explains how to review the current chassis controller configuration.

Procedure

1. From a local console, use the `ssh` command to open a remote session to the admin node.
2. Enter the following command to ensure that the `cmcdetectd` service is running:

   ```
   # systemctl status cmcdetectd
   ```

   If the `cmcdetectd` service is not running, enter the following command:

   ```
   # systemctl start cmcdetectd
   ```

3. Enter the following command to monitor the progress of `cmcdetectd`:

   ```
   # tail -f /opt/clmgr/log/cmcdetectd.log
   ```

4. (Optional) Use the `switchconfig unset` command to reset the management switch ports.

   If you know the management switch and the ports to which a chassis controller connects, you can reset the management switch ports back to default settings.

   This practice ensures that the `cmcdetectd` service receives the DHCP packets from the chassis controller.

   Example: Assume that rack 1, chassis controller 1 is plugged into `mgmtsw0` ports 1/0/11 and 2/0/11. Issue the following command to set both ports back to default settings:

   ```
   # switchconfig unset --switches mgmtsw0 --ports 1/0/11 --redundant yes
   ```

5. Flip the power breakers on for the chassis controllers, one rack at a time.

   Notice that `cmcdetectd` runs in a serial fashion. It handles one chassis controller configuration at a time to prevent configuration conflicts on the management switches.

6. Verify the configuration.

   After the `cmcdetectd` service configures a chassis controller, the system logs an entry to the following file:

   `/etc/cmc-switch-info.txt`

   Verify that this file contains the correct entries.

   Example 1. The chassis controller entry might look as follows:

   ```
   # cat /etc/cmc-switch-info.txt
   mac_address=00:fd:45:ff:3b:46, mgmtsw=mgmtsw0, vlans=None, default_vlan=2001, bonding=manual,
   ports=1/0/11, redundant=yes, cmc_type=nonice, cmc_hostname=r1c1
   # VLAN to management switch configuration
   vlan=2001, mgmtsw=mgmtsw0, configured=no, chassis_type=nonice, vlan_type=multinet
   ```

7. Back up the `/etc/cmc-switch-info.txt` file to another server at your site.

   If you ever have to perform a disaster recovery, this information can be useful.

Method 1 - Configuring the chassis controller switches manually

About this task

To reapply the chassis controller configuration on management switches quickly, complete this procedure. Use this procedure when the management switches are reset to factory settings or when a switch is replaced. Use this procedure if you have the `/etc/cmc-switch-info.txt` file.

NOTE: To configure L3 VLAN routing settings, you might need to change the `configured` VLAN property to `no` in order for `cmcdetectd` to re-apply a configuration to a management switch if the management switch has been reset. Example:

`vlan=####, … configured=no, …`

Procedure

1. Make sure that all chassis controllers are configured.
2. Run the following command to configure the chassis controller switches:

```
# cmcdetectd --switchconfig
=== Reading the CMC-Switch configuration file: /etc/cmc-switch-info.txt ===

=== Running switchconfig for VLANs ===

VLAN 101 on management switch mgmtsw0 is an ICE VLAN, skipping L3 configuration

=== Running switchconfig for all CMCs ===

configuring CMC 08-00-69-16-C0-12 on managemenent switch mgmtsw0...
command: switchconfig set --switches mgmtsw0 --ports 1/11 --default-vlan 101 --bonding manual --redundant yes --vlans 3

saving configuration on management switch(es) mgmtsw0...
command: switchconfig config --switches mgmtsw0 --save

=== Results ===

Component               Function            Result
-----------------------------------------------------------
CMC 08-00-69-16-C0-12    switchconfig set    successful
```

Method 2 - Configuring the chassis controller switches manually

About this task

Use this procedure if you do not have the `/etc/cmc-switch-info.txt` file.

This manual configuration method requires that you provide the following information:

- The rack VLANs in which the chassis controllers reside.
- The physical ports and management switches to which the chassis controllers are cabled.

If you cannot provide this information, do not use this method to configure the chassis controllers.

Procedure

1. From a local console, use the `ssh` command to open two remote sessions to the admin node.
2. In one remote session window, enter the following command to monitor the `switchconfig` log file:

   `# tail -f /opt/clmgr/log/switchconfig.log`

   Your goal is to make sure that the commands being sent are completing successfully.
3. In the other remote session window, use the `switchconfig` command to configure the management switch manually.

   As inputs to the `switchconfig` command, use the rack VLAN information and information about the physical port location of the chassis controller.

For example, assume that chassis controller `r1i0c` (that is, rack 1 chassis controller 0 using VLAN 101) is connected to management switch `mgmtsw0` on port `1:11` and `mgmtsw0` port `2:11`. Use one of the following commands:

```
# switchconfig set --switches mgmtsw0 --default-vlan 101 --vlan 3 \
--bonding manual --ports 1:11 --redundant yes
```

or

```
# switchconfig set -s mgmtsw0 -d 101 -v 3 -b manual -p 1:11 -r yes
```

For more information about the `switchconfig set` command, enter the following:

```
# switchconfig set --help
```

4. Flip the power breakers on for the chassis controllers, one rack at a time.

   Because the management switch is already configured, no additional tasks should be needed.

5. (Optional) Validate the chassis controller configuration.

   Enter the following `switchconfig` command to display the bonding and VLAN configuration of the chassis controllers that are connected to the management network:

```
# switchconfig sanity_check -s mgmtsw0

======== Beginning Sanity Check on mgmtsw0 ========

checking port-channel sharing configuration on mgmtsw0...(address-based L2/L3/L3_L4 = static port-channel,
address-based L2/L3/L3_L4 lacp = LACP port-channel)

port-channel group master is 1:11 with the following ports in a port-channel: 1:11, 2:11 in bonding mode: address-based L2
```

Node provisioning takes too long or fails to complete

Symptom

Node provisioning (or imaging) takes too long or fails to complete.

Cause

If you use UDPcast, and provisioning takes too long or fails to complete, consider using BitTorrent or `rsync`. Sometimes a different file transfer method helps a node discovery command to complete more quickly.

Action

1. Review the information in this step and select a different file transport method.

   When you configure the cluster components, consider the following:

   - The types of nodes you have
   - The file transfer methods
   - Whether you want to modify the node characteristics that currently exist in the cluster definition file

   The file transfer method directly affects the time it takes to install software on each node. This process includes the following event sequence:

- A `cm node add` or a `cm node discover add` command completes and returns you to the system prompt.
- The compute nodes install themselves with software from the admin node.
- The node comes up. At this point, if you issue a `cm power` command query to the node, the node responds with `ON`.

Regardless of the cluster type or node type, the default transfer method is BitTorrent. Other file transfer methods are `rsync` and UDPcast.

Table 5: Compute node characteristics and image information shows the node types and includes information about file transfer methods that are appropriate for each node.

Table 5: Compute node characteristics and image information

	Compute nodes with root disks	Compute nodes with NFS root file systems	Nodes with `tmpfs` root file systems
Transport path	The admin node installs the flat compute nodes. Uses UDPcast, BitTorrent, or `rsync`.	N/A	On clusters without leader nodes, the admin node installs the nodes using UDPcast, BitTorrent, or `rsync`.
Software to be installed	The image resides on the admin node.	N/A	N/A
Boot persistent?	Yes.	For compute nodes, boot persistence is possible depending on the configuration.	No. All is lost on reboot.
Node image memory use	No.	For compute nodes, node image memory use depends on the configuration.	Yes. The root file system consumes system memory.
RPM installation notes	N/A	For compute nodes, NFS solutions that use overlay let RPMs be installed.	You can install RPMs on the nodes. However, each node receives a new image when the node boots, so RPM images are not boot-persistent.
Image root file system notes	N/A	For compute nodes, the overlay solutions are writeable. The overmount solutions are writeable to some locations.	N/A

The other consideration when choosing a file transport method is the method itself. **Table 6: File transfer methods** shows the available file transport methods.

Table 6: File transfer methods

	`rsync`	BitTorrent	UDPcast
Status	Not default	Default	Not default
Performance	Slower performance when pushing images to more than two nodes simultaneously.	Midrange performance.	Fastest performance.
Method	Pushes the image to all nodes, in separate sessions, over the cluster network. This action can consume all bandwidth when more than two nodes are involved.	Transfers the node image as a `tar` file that is divided into pieces. The individual nodes receive the pieces and assemble the pieces into an image. After the node assembles the image, it boots. When you use this method, the miniroot is always transferred using `rsync`. The other image components are transferred using BitTorrent.	Transfers the node image in a multicast stream, which has one sender and many listeners.
Encryption?	Yes	Yes	Yes
Appropriateness	Suited for a small number of nodes (2-5). If you have many nodes, run a node discovery command multiple times, and target different groups of nodes each time.	Suited for a large number of nodes.	Most efficient for large numbers of nodes. Requires the switches to be configured for multicast traffic. Some switches might require additional configuration. Switches shipped with HPE clusters require no additional configuration.
Files transferred	Kernels, `initrd`, and the miniroot file system.	Kernels, and `initrd`. The system uses `rsync` to transfer the miniroot file system	The miniroot file system.

In addition to the tables, you can use the following figure to help you select a transport method:

Figure 6: Selecting a transport method for provisioning clusters that do not have leader nodes

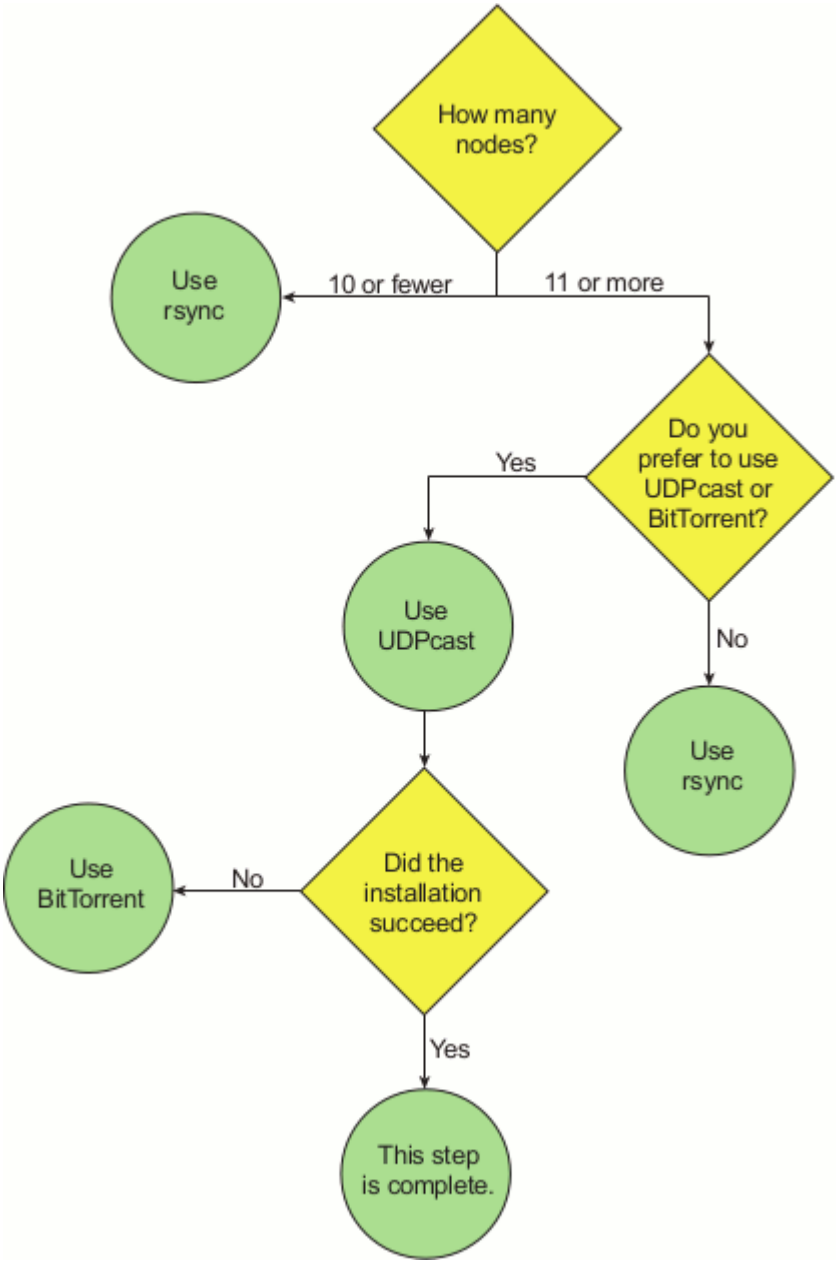

Figure 6: Selecting a transport method for provisioning clusters that do not have leader nodes

2. Update the cluster definition file to specify the alternative file transport method.

   Specify one of the following settings:

   - `transport=udpcast`
   - `transport=bt`
   - `transport=rsync`

3. Run the `cm node add` command again, and specify the newly updated cluster definition file.
4. Evaluate the provisioning time with the new cluster definition file.

   In some cases, any file transfer method can result in nodes that do not complete the data transfer. If the data transfer does not finish, reinstall the software on the failed node.

   Use the `cm node provision` command to install the software:

   `cm node provision [--transport method] -n failed_node_name`

   For *method*, specify one of the following data transport methods:

   - `bittorrent`
   - `rsync`
   - `udpcast`

   If you continue to have problems, you might have to change the transport method again. If UDPcast and BitTorrent both fail, specify `rsync`.

   Your site network configuration can affect the speed at which the `cm node provision` command can push software to nodes.

Suppressing nonfatal messages in the authentication agent

Symptom

The system issues the following erroneous message when you use `ssh` to log into the admin node as the root user:

```
Could not open a connection to your authentication agent.
```

You can safely ignore this message. Alternatively, use the command in this topic to start the agent in a way that suppresses this message.

Action

Start the `ssh` agent in a way that suppresses nonfatal messages about the authentication agent.

For example, in the `bash` shell, enter the following command:

```
# exec ssh-agent bash
```

To suppress the erroneous message, run this command after each boot during the installation process. After installation, the system no longer issues this message.

Verifying that the `clmgr-power` daemon is running

About this task

The following procedure explains how to make sure that the `clmgr-power` daemon is running properly.

Procedure

1. Log into the admin node as the root user, and enter the following command to make sure that the `clmgr-power` daemon is running:

   ```
   # systemctl status clmgr-power
   ```

   For example, the following example shows the daemon running as expected on a SLES system:

   ```
   # systemctl status clmgr-power
      clmgr-power.service - clmgr power
      Loaded: loaded (/usr/lib/systemd/system/clmgr-power.service; enabled;
   vendor preset: enabled)
      Active: active (exited) since Tue 20XX-06-26 08:28:07 CDT; 1 day 5h ago
    Main PID: 4183 (code=exited, status=0/SUCCESS)
      CGroup: /system.slice/clmgr-power.service
              └─4942 clmgr-power /usr/bin/twistd --originalname -o -r poll --
   logfile /opt/clmgr/log/clmgr-power.log --pidfile /var/ru...
   .
   .
   .
   ```

   If the daemon is not running, enter the following command to start the daemon:

   ```
   # systemctl start clmgr-power
   ```

2. Use a text editor to open file `/opt/clmgr/log/clmgr-power.log`, which is the log file for the `clmgr-power` daemon on the admin node.

3. Verify that the log entries indicate a running daemon.

   For example, the following log file entries show that the `clmgr-power` daemon is running as expected:

   ```
   20XX-05-03 14:16:07+0000 [-] Log opened.
   20XX-05-03 14:16:07+0000 [-] twistd 14.0.2 (/usr/bin/python 2.7.9) starting up.
   20XX-05-03 14:16:07+0000 [-] reactor class: twisted.internet.pollreactor.PollReactor.
   20XX-05-03 14:16:07+0000 [-] Changing process name to clmgr-power
   20XX-05-03 14:16:07+0000 [-] Log opened.
   20XX-05-03 14:16:07+0000 [-] twistd 14.0.2 ( 2.7.9) starting up.
   20XX-05-03 14:16:07+0000 [-] reactor class: twisted.internet.pollreactor.PollReactor.
   20XX-05-03 14:16:07+0000 [-] PBServerFactory starting on 8800
   .
   .
   .
   ```

   If the log file entries show traceback activity, the daemon might not be running correctly. If you see traceback entries, and you need help to interpret them, contact your technical support representative.

Using the `switchconfig` command

The `switchconfig` command displays switch settings and enables you to configure switches.

To retrieve help output online, which includes examples, enter the following:

```
# switchconfig --help
```

The preceding command displays all the possible subcommands. To retrieve more information about an individual subcommand, specify the following:

```
switchconfig subcommand --help
```

Nodes are taking too long to boot

Symptom

Nodes are taking too long to boot.

Cause

Generally, the 802.3ad (LACP) bonding mode provides more bandwidth and redundancy than the active-backup bonding mode. However, the bonding mode on a node must match the management Ethernet switch to which it is connected.

If the following are all true, use the procedure in this topic to verify the bonding mode and if necessary, to update the bonding mode:

- The `cm node add` command or the `cm node discover add` command has run.
- The node in question is configured into the cluster.
- You need to change the bonding mode for the node.

The following procedure works on any type of node, including an admin node.

Action

1. Log into the admin node as the root user.
2. Use the `cm node show` command in the following format to display the bonding mode:

   `cm node show --mgmt-bonding -n` *`hostname`*

   For *hostname*, specify the hostname of the node you want to verify.
3. Use the `cm node set` command to reset the bonding mode on a given node.

   Use the command in one of the following formats:

   `cm node set -n` *`hostname`* `--mgmt-bonding active-backup`

   Or

   `cm node set -n` *`hostname`* `--mgmt-bonding 802.3ad`

   Example:

   ```
   admin# cm node set -n n0 --mgmt-bonding 802.3ad
   ```
4. Reset the node.

   Use the `cm power reset` command in the following format:

   `cm power reset -t` *`target_type hostname`*

   For *target_type*, specify `node`.

   For *hostname*, specify a node hostname.

   Example:

   ```
   admin~# cm power reset -t node n0
   ```

Wait for the node to reboot fully.

5. Use the `switchconfig_configure_node` command in the following format to configure the management switch attached to the node:

   ```
   switchconfig_configure_node --node hostname
   ```

Nodes fail to boot

Symptom

Legacy BIOS software has the following network bootloaders:

- GRUB version 2
- iPXE

UEFI systems ((Arm (AArch64) and x86_64) have two network bootloaders:

- GRUB version 2
- iPXE-direct

GRUB version 2 was the default bootloader for a long time, but newer platforms use iPXE-direct by default.

If you are having trouble booting, for example, if GRUB version 2 or iPXE-direct starts but is unable to transfer data over the network, you can try the other bootloader.

NOTE: iPXE-direct does not support the disk bootloader feature, which is the feature that network boots GRUB version 2 then causes GRUB version 2 to chainload locally on disk to the currently selected slot.

A node can fail to boot for many reasons. This topic explains one remedy, which is to try booting the node with GRUB version 2, rather than the default of iPXE.

For additional information, see the following:

`dhcp_bootfile`

Action

1. Log into the admin node as the root user.
2. Use the following command to verify whether the node is enabled to load iPXE:

   ```
   cm node show --dhcp-bootfile -n node
   ```

   For *node*, specify the hostname of the node that did not boot.
3. Use the following command to specify that iPXE load first and that iPXE load GRUB version 2:

   ```
   cm node set --dhcp-bootfile method -n node
   ```

   Select one of the following for *method*:

method	Appropriateness
`ipxe-direct`	`ipxe-direct` uses iPXE itself to load the kernel and `inittramfs` directly, without using GRUB version 2. Available on all systems (UEFI, x86_64, UEFI Arm (AArch 64), and legacy x86_64) except HPE Apollo 80 systems. Default.
`ipxe`	`ipxe` is available as a workaround for legacy x86_64 (non-UEFI) systems. When specified, the node network boots iPXE, and iPXE loads GRUB version 2 over the network. This method works around problems with some types of BIOS.
`grub2`	`grub2` is available on UEFI and legacy systems, Arm (AArch64), and x86_64. Sometimes a system or BIOS version does not work well with GRUB version 2 network booting. HPE Apollo 80 nodes require grub2.

4. Boot the node.

   If the node still does not come up, the following are some additional actions you can take:

   - Attach a console or review the console log.
   - If node is in a shell, check its date.
   - Inspect the node entry in the cluster definition file for missing or incorrect parameters.
   - Verify that BIOS settings are correct.

Cannot find the management switch that a node is plugged into

Symptom

You cannot find the management switch that a node is plugged into.

Action

1. From the admin node, use the `arp` command to find the MAC address of the node.

   The command format is as follows:

   `arp` *`hostname`*

   For *hostname*, enter the hostname of the node.

2. Use the `switchconfig find` command.

   The `switchconfig find` command returns the switch upon which a given node MAC address exists. The command searches multiple switches and displays information about physical ports and switches.

   NOTE: Some long commands in this topic use the \ character to continue the command to a second line.

Example 1. The following command searches all management switches for MAC address `00:25:90:96:4e:ac`:

```
admin:~ # switchconfig find --switches all --macs 00:25:90:96:4e:ac
mac-address               switch        find_method     port
-------------------------------------------------------------------
00:25:90:96:4e:ac         mgmtsw3        lldp           1:1
```

The preceding output shows the following:

- The MAC address is found on `mgmtsw3`, port `1:1`.
- The command used the link layer discovery protocol (LLDP) to determine its findings.

Log files

The main log file directories are the `/opt/clmgr/log` directory and the `/var/log` directory.

The following are some other log files that might interest you:

- On the admin node, the `/var/log/messages` resides on the admin node.
- `/opt/clmgr/log/cmcdetectd.log`

  On the admin node, `cmcdetectd` logs its actions as it configures the switches for chassis controllers in the system. Watch for progress or errors here.
- `/var/log/dhcpd`

  This file contains DHCP messages.

  This file resides on the admin node.
- `/opt/clmgr/log/switchconfig.log`

  On the admin node, there is a `switchconfig` command-line tool. This tool is largely used as nodes are configured into the cluster. Its actions are logged in this log file.

Ensuring that the hardware clock has the correct time

Some software distributions do not synchronize the system time to the hardware clock as expected. As a result, the hardware clock is not synchronized with the system time, which is the correct condition. At shutdown, the system time is copied to the hardware clock, but sometimes this synchronization does not happen.

To set the compute node hardware clocks properly, check the following:

- Use the `chronyc sources` command to show synchronization.

  In the output, note the following:
  - The carat (^) adjacent to the hostname or IP address shows that the node is an NTP server.
  - The asterisk (*) shows the NTP server to which the system is synchronized.
  - The plus sign (+) shows an NTP server that is a combined source.
  - The minus sign (–) shows an NTP server that is not a combined source.
- Use the `chronyc tracking` command to show the state of a node.

- To set the hardware clock to the system clock, enter the following command:

  ```
  # hwclock --systohc
  ```

- To set the hardware clock to the system clock on the compute nodes, enter the following command:

  ```
  # clush -g compute hwclock --systohc
  ```

- To confirm the current hardware clock time, enter the `hwclock` command without options, as follows:

  ```
  # hwclock
  Thu 26 Jan 20XX 10:57:27 PM CST  -0.750431 seconds
  ```

- To confirm the current hardware clock on compute nodes, enter the following command:

  ```
  # clush -b -g compute date
  node0: Tue Apr 18 15:00:45 PDT 20XX
  ```

Switch wiring rules

Some clusters have a redundant management network (stacked pairs of switches). Other clusters have cascaded switches, in which switch stacks are cascaded from the top-level switch. When configuring cascaded switches, it is impossible to know the connected switch ports of all trunks in advance, so you start with only one cable and add the second one later on.

When trunks are configured, it is often hard to find the MAC address of both legs of the trunk. The difficulty arises because the trunked connection just uses one MAC for the connection. Therefore, you can rely on rules that infer the second port connection based on the first port connection.

The following are some simple wiring rules:

- In a redundant management network (RMN) configuration, use the same port number in both switches for a particular piece of equipment. That is, make sure to assign the same port number in each stack to the following components:
  - Admin nodes
  - Compute nodes with services installed upon them
  - Chassis controllers

  For example, if you connect chassis controller `r1i0c` chassis controller 0 port to switch A, port 2, then `r1i0c` chassis controller 1 port must go to switch B port 2.
- When adding cascaded switch stacks, all switch stacks must cascade from the primary switch stack. In other words, there is always only, at most, one switch hop.
- When configuring cascaded switch pairs in an RMN setup, observe the following:
  - If you are connecting switch stack 1, switch A, port 48 to switch stack 2, then connect the second trunked connection to stack 2, switch B, port 48.
  - Until the cascaded switch stack is configured into the cluster database, leave one trunk leg unplugged temporarily to prevent looping.
  - The node discovery commands tell you when it is safe to plug in the second leg of the trunk. This notification avoids circuit loops.

Bringing up the second NIC in an admin node when it is down

About this task

The logical interface, `bond0`, can contain one or more physical NICs. It is possible for these physical NICs to be administratively down or unplugged. The following procedure explains how to determine link status of the physical NICs under bond0.

The following procedure explains how to detect this situation.

Procedure

1. Check the Ethernet port of the add-in card and confirm that it is lit.
2. Confirm that the add-in card connection to the management switches is using port `0`.

   Make sure that port `1` is not connected.

   This step verifies the wiring.
3. Examine the following file to see whether the second, redundant Ethernet interface link is down:

   `/proc/net/bonding/bond0`
4. Use the `ethtool` command to determine if the content of the `Link detected:` field is `no`.

   For example:

   `# ethtool management_interface1`
5. Enter the following command to bring up the interface:

   `# ip link set management_interface1 up`
6. To verify that the link is detected, run the following command:

   `# ethtool management_interface2`
7. In the preceding command output, search for `yes` in the `Link detected` field.

Miniroot operations

The following topics can help you troubleshoot a suspected miniroot kernel problem:

- **Miniroot functioning**
- **Entering rescue mode**
- **Logging into the miniroot to troubleshoot an installation**

Miniroot functioning

The cluster manager miniroot is a small Linux environment based on the same RPM repositories that generated the root image itself.

The cluster manager software uses the miniroot to install the software and to boot the computes nodes over the cluster network.

The miniroot is a small, bootable file system. It includes kernel modules such as disk drivers, Ethernet drivers, and other software. These software drivers are associated with a specific kernel number. As new driver updates become available,

the operating systems distribute additional kernels. The system requires at least one kernel to be associated with a specific node image. You can associate more than one kernel with a specific node image.

When the cluster manager boots a node, the cluster manager uses the images that reside in the admin node image repository. Because the nodes boot over the network, it is important that the images in the admin node image repository include the correct kernels. That is, it is important that the following are identical:

- The kernel in the on-node image. This image is the installed image that resides on the node while the node is running.
- The kernel in the node image repository on the admin node. There can be multiple node images for a single node type in the image repository.
- The kernels in the kernel repository on the admin node. There can be multiple kernels in the kernel repository. The `cm node provision` command includes a kernel from the repository when it builds a node image. These kernels reside in the following directory on the admin node:

  `/opt/clmgr/tftpboot/images`

Use the cluster manager `cm node provision` command to update images. When you use this command, the cluster manager ensures synchronization between the on-node image and the image in the admin node image repository. Do not change an on-node image manually without using the cluster manager commands. If you omit the command and subsequently boot the node, one of the following occurs:

- The boot fails

  Or

- The cluster manager detects a mismatch between the following:
  - The kernel loaded over the network
  - The kernel and associated modules in the image itself

  The mismatch can lead to a node that boots but has no network, for example. Therefore, it is important that all the images in the image repository on the admin node contain the on-node images with all the kernels in use.

If you update any images manually, use the following command:

`cm image update -i` *`image`* `-k`

The preceding command has the following effects:

- The command synchronizes the kernels and the `initrd` daemon in the images.
- The command writes a copy of the kernel to the `/opt/clmgr/tftpboot/images` directory for future use when performing network boots.

Entering rescue mode

To go into miniroot rescue mode, enter commands such as the following:

```
# cm node set -n n1 --kernel-extra-params "rescue=1"
# cm node refresh netboot -n n1
```

The preceding command includes the `rescue=1` kernel command-line argument. This argument ensures that the kernel command line includes `rescue=1`.

To remove `rescue=1`, use commands such as the following:

```
# cm node unset --kernel-extra-params -n n1
# cm node refresh netboot -n n1
```

Logging into the miniroot to troubleshoot an installation

The miniroot brings up an `ssh` server for its operations. If an installation fails, first look to the serial console using the `conserver` command and any console log files.

To examine the situation from a separate session, specify port 40. The miniroot environment listens for `ssh` connections on port 40.

For example, assume that the node that failed to install is `n0`. The following command logs you into the miniroot on node `n0` from the admin node:

```
admin# ssh -p 40 root@n0
miniroot#
```

At this point, you can run typical Linux commands to debug the problem. HPE supports only a subset of the standard Linux commands on the miniroot.

Troubleshooting an HA admin node configuration

The following list shows the commands that you can use to troubleshoot an HA admin node configuration problem:

- To verify the network configuration, examine the `/etc/hosts` file.

  For example:

  ```
  # cat /etc/hosts
  100.10.10.10    acme-admin1
  100.10.10.11    acme-admin2
  100.100.0.1     acme-admin1-ptp
  100.100.0.2     acme-admin2-ptp
  111.22.222.222  acme-admin1-head
  111.22.222.223  acme-admin2-head
  111.40.40.100   acme-admin1-bmc
  111.40.40.101   acme-admin2-bmc
  ```

- To verify the firewall, use the following commands:

  - On RHEL platforms, enter the following command:

    ```
    # cat /etc/firewalld/zones/public.xml | grep service
    <service name="dhcpv6-client"/>
    <service name="ssh"/>
    <service name="high-availability">
    ```

  - On SLES platforms, enter the following command:

    ```
    # cat /etc/sysconfig/SuSEfirewall2 | grep FW_CONFIGURATIONS_EX
    FW_CONFIGURATIONS_EXT="cluster sshd vnc-server"
    ```

Troubleshooting UDPcast transport failures from the admin node

About this task

You might encounter one of the following situations when you use the UDPcast (multicast) transport method during installation:

- The client side waits forever for a `udp-receiver` process to complete.
- The `udp-receiver` processes repeatedly attempts to provision a node.

If either of the preceding conditions exist, you have an error situation.

The `systemimager-server-flamethrowerd` service manages the `udp-sender` instances. The following procedure explains how to remedy this situation by stopping and restarting UDPcast `flamethrower` services.

The following procedure explains another UDPcast troubleshooting strategy:

Troubleshooting UDPcast transport failures from the switch

Procedure

1. As the root user, log in to the node that serves UDPcast.

   Log into the admin node if your goal is to restart UDPcast services for compute nodes.
2. Use one of the following commands to stop the `systemimager-server-flamethrowerd` services:

   From an admin node or a high availability (HA) admin node, enter the following command:

   ```
   # systemctl stop systemimager-server-flamethrowerd
   ```
3. Enter the following command to check for `udp-sender` processes that did not stop:

   ```
   # ps -ef | grep udp-sender
   ```
4. (Conditional) Enter one or more `kill -9 process_ID` commands to stop `udp-sender` processes that are still running.
5. Start the `systemimager-server-flamethrowerd` service.

   Use one of the following commands:

   From an admin node or an HA admin node, enter the following command:

   ```
   # systemctl start systemimager-server-flamethrowerd
   ```

Troubleshooting UDPcast transport failures from the switch

About this task

UDPcast relies on IGMP technology. The IGMP technology determines the physical ports that subscribe to specific multicast addresses at layer 2 (data link) in the OSI model. In some scenarios, IGMP can be problematic for UDPcast.

The following `switchconfig` commands show the parameters that retrieve IGMP status information:

- To view global IGMP status on a management switch:

  ```
  switchconfig igmp --switches mgmtswX --info
  ```
- To the IGMP status for a specific VLAN on a management switch:

  ```
  switchconfig igmp --switches mgmtswX --info --vlan VLAN_#
  ```

You can disable IGMP on the management switches. Disabling and enabling IGMP have the following effects:

- When IGMP is enabled, a layer-2 multicast tree is created on the Ethernet switches. This tree determines the ports to which the UDPcast traffic is forwarded. In some cases, in the UDPcast code, the `IGMP Join` packets from the `udp-receiver` clients do not reach the Ethernet switches. In these cases, the multicast tree is not formed.
- When IGMP is disabled globally, the Ethernet switches convert all multicast packets to broadcast packets. In this case, the packets are nearly guaranteed to reach every host in a VLAN. Thus, the reduced performance increases reliability.

The following `switchconfig` commands show the parameters that disable IGMP on the management switches:

- To disable IGMP globally on a management switch:

  `switchconfig igmp --switches mgmtsw`*`X`* `--disable`

- To disable IGMP on a specific VLAN on a management switch:

  `switchconfig igmp --switches mgmtsw`*`X`* `--disable --vlan` *`VLAN_#`*

To re-enable IGMP on global or per-VLAN basis, replace `--disable` with `--enable`. In addition, if necessary, use the `--version` parameter to specify the IGMP version. You can specify `--version 2` or `--version 3` . The default version is version 2.

The following command re-enables IGMP with IGMP version 3 on `mgmtsw0`:

```
# switchconfig igmp --switches mgmtsw0 --enable --version 3
```

Connecting to the virtual admin node in a cluster with a high availability (HA) admin node

Procedure

1. Log into one of the physical admin nodes as the root user.
2. Use the `crm_mon` command to determine which physical node hosts the virtual admin node at this time.
3. Log into the node that hosts the virtual admin node
4. In a terminal window, connect to the virtual admin node.

   For a system admin controller high availability (SAC HA) virtual admin node, enter the following command:

   ```
   # virsh console sac
   ```

   For a quorum HA virtual admin node, enter the following command:

   ```
   # virsh console adminvm
   ```

Nodes configured but with mismatched BIOS settings

Symptom

This problem presents itself when nodes that are identical, or are presumed to be identical, exhibit different behaviors. For example, some nodes might PXE boot and others might not.

You can use the remedy in this topic to analyze BIOS differences in the following additional circumstances:

- To compare the BIOS differences between two or more nodes. For example, you can see the boot order for different nodes easily.
- To retrieve information about node differences related to Hyperthreading or other settings.

In addition to using the commands in this topic for troubleshooting, you can use these commands to adjust BIOS settings for performance.

Cause

New nodes were configured into the cluster, but the BIOS settings on the new nodes do not match the BIOS settings on the other nodes.

Action

1. Log into the admin node as the root user.
2. Use the `cm node show` command to display the cluster nodes.

   For example:

   ```
   # cm node show
   n1
   n2
   n3
   n4
   ```

3. Use the `cm node bios show` to display the BIOS setting differences.

   The format is as follows:

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

   For *nodes*, specify the node hostnames.

   You can use a mouse to click on the highlighted lines in the output to display the differences for each node.

   For example:

   ```
   # cm node bios show -n n2,n[3-4] --cmdiff
   ```

4. Adjust the BIOS settings as needed.

   The cluster manager provides the following commands:

   - `cm node bios show`, which shows BIOS settings for nodes.
   - `cm node bios set`, which lets you set BIOS features.
   - `cm node bios reset`, which lets you reset the BIOS to factory settings.

   Use the BIOS documentation and HPCM to adjust the BIOS settings.

Cluster manager cannot find a suitable disk

Symptom

When installing the admin node using the installation media, or using the installation environment on any node to install to disk drives, the installer might fail if it cannot determine the disk upon which to install the operating system. In this case, miniroot exits and displays instructions that describe how to proceed. The possibilities are as follows:

- If you reinstall the cluster manager, the cluster manager overwrites existing disks with new cluster manager data and labels. In this case, the cluster manager recognizes the existing disks as belonging to a previous cluster manager installation. The installation proceeds as expected without issuing any messages.
- If you reinstall the cluster manager onto disks that were not part of a previous cluster manager installation, the cluster manager does not recognize the disks. In this case, the cluster manager miniroot exits and issues instructions in a message.
- If you start an installation, or a reinstallation, and there is more than one blank disk device, the cluster manager does not arbitrarily choose a disk. In this case, the cluster manager miniroot exits and issues instructions in a message.

If your console terminal did not buffer the instructional messages, you can find the messages in the following file:

`/tmp/si.log`

Before proceeding to the solutions that follow, complete the following prerequisites:

1. Copy or write down all disk devices in the following directory:

   `/dev/disk/by-path/`

2. Determine the disk you want to use for the installation.

Solution 1

Cause

The cluster manager did not find any suitable disks. The problem could be that there are too many empty disks or that there are no empty disks. If you have a RAID controller, the RAID controller might not be configured properly.

Action

1. Observe the messages that the installer displays.

   The installer lists possible disks for the installation and prompts you to enter one of the disks that it found.

   For example:

   ```
   .
   .
   .
   Listing of block devices from lsblk
   NAME     MAJ:MIN RM  SIZE RO TYPE MOUNTPOINTS
   /dev/sda   8:0    0  1.8T  0 disk
   /dev/sdb   8:16   0  1.8T  0 disk
   /dev/sdc   8:32   0  1.8T  0 disk
   /dev/sr0  11:0    1  7.1G  0 rom
   ! - for subshell (exit returns), or disk /dev device to force and re-try:
   ```

2. (Optional) Press `!` and press Enter to enter a temporary shell.

   By entering this temporary shell, you can examine the system.

   The temporary shell includes tools, such as `sgdisk`, for managing this situation. For example, you can use `sgdisk` in the following way to completely wipe one or more devices:

   `sgdisk --zap-all /dev/disk/by-path/`*`target_device_name`*

   For *target_device_name*, specify a device name.

   To exit the temporary shell, type `exit`. When you exit the temporary shell, the prompt reappears.

3. Enter one of the disks shown to start the installation.

   For example, enter `/dev/sda` and press Enter.

4. (Conditional) Install other disks.

   Complete this step if you have multiple, similar nodes with the same disk installation problem.

   Use the following command:

   ```
   cm node provision -n nodes --force-disk /dev/disk/by-path/target_device_name
   ```

   The variables are as follows:

Variable	Value
nodes	One or more node hostnames. Use a comma to separate hostnames. You can specify multiple node hostnames if all the nodes have identical hardware.
target_device_name	The by-path identifier. Do not identify nodes by using `/dev/sda` formats when you use the `cm` commands.

Solution 2

Cause

Too many blank disks were found, and you want an MD RAID array.

Action

Use the cm node provision command to set up an MD RAID array.

For example:

```
cm node provision -n nodes --force-disk=
/dev/disk/by-path/1st_target_device_name, \
/dev/disk/by-path/2nd_target_device_name, \
/dev/disk/by-path/3rd_target_device_name, \
/dev/disk/by-path/4th_target_device_name \
--md-metadata md --md-raidlevel 10
```

Socket failure when connecting to the configuration manager

Symptom

A socket failure can occur when you attempt the following:

- You run the following command:

  ```
  cm node update config --sync -n admin
  ```

- You change the IP address of the admin node.

In the preceding cases, the cluster manager might issue the following messages:

```
ERROR: Socket failure connecting to configuration manager ('172.xx.xx.xx', 1030): Connection refused
ERROR: Retrying in 0.500 seconds
ERROR: Socket failure connecting to configuration manager ('172.xx.xx.xx', 1030): Connection refused
ERROR: Failed to contact configuration manager
```

Cause

These messages reflect a failure to connect with the configuration manager.

Action

1. Log into the admin node as the root user.
2. Enter the following command to restart the configuration manager service:

   ```
   # systemctl restart config_manager.service
   ```

Replacing and servicing nodes

You can install and configure a spare node to replace failed system disks.

The failed node can be any kind of node, including an admin node. The cold spare can be a shelf spare or a factory-installed cold spare that shipped with your system. The replacement process applies equally to the case where the spare is actually the failed node itself with a motherboard replacement.

As part of maintaining the cluster, make sure that you always have the following two types of spare nodes:

- One spare for the admin node.
- One spare for a compute node.

NOTE: If you are using multiple root slots, the installation procedures affect only the current slot.

For information about other hardware operations and about replacing other types of cluster components, see the following:

HPE Performance Cluster Manager Administration Guide

Replacing a node

About this task

The procedure in this topic explains how to replace an entire node, including the system disks in the node. This procedure does not preserve the original system disks.

For example, you can use this procedure to replace service nodes on any cluster.

You can use the procedure in this topic to replace nodes that are not automatically managed by any of the following:

- `cmcdetectd`
- `cmcinventory`

Procedure

1. Verify that you have an appropriate spare node.

   A cold spare node is equivalent to one of the nodes on your running cluster. The spare sits on a shelf or is a factory preinstalled node. The cold spare is intended to be used in an emergency.

   Make sure that HPE supplied your spares. HPE does not support spares not supplied by HPE.

   The following are some reasons to have the two types of spares:

   - Admin node BIOS settings are different from BIOS settings for other nodes.

     For example, the boot order is different for each node type.

Attempts to configure the node into a cluster will fail.

- Depending on your site policy, the node controllers of an admin node might or might not be configured to use DHCP by default.
- The management cards of compute nodes must be configured to use DHCP by default.

  Otherwise, attempts to configure the node into the cluster will fail.

2. Connect a keyboard, video screen, and mouse to the node.
3. If possible, power down the failed node.
4. Examine and label the power cables.

   Before you disconnect any cables, make sure they are labeled. Make sure that you are familiar enough with the cabling to re-cable the new node at the end of this procedure.
5. Disconnect all power cables.
6. Unplug the Ethernet cables used for system management.

   To avoid confusing them, note the plug number and label the cables. It is important that they stay in the same jacks in the new node. This connection is vital to proper system management and communication.

   NOTE: The Ethernet cables must be connected to the same plugs on the cold spare unit.
7. Remove any peripheral components, such as a keyboard, video screen, or mouse, from the node.
8. Remove the failed node from the rack.
9. Install the shelf spare node into the rack.
10. Connect the Ethernet cables in the same way they were connected to the replaced node.
11. Connect AC power.
12. Connect to the node through the node controller or attach a keyboard, video screen, and mouse to the node.
13. From the admin node, update the following in the cluster manager database:

    - The MAC address of the spare
    - The MAC address of the node controller in the spare

    When you update the preceding address information in the database, you ensure that the cold spare can boot and function properly. If necessary, use the BIOS to retrieve the new MAC addresses. For more information about how to retrieve the MAC address of the spare, see the node controller documentation for the spare. For example, see the iLO server guide for the spare.

    From the admin node, query and set the MAC addresses in the database. The following table shows the command parameters that you can use:

    Example:

    The following example displays the MAC address of compute node `n0`:

    ```
    # cm node show -M -n n0
    NODE   NETWORK.NAME   IPADDRESS    SUBNETMASK    MACADDRESS
    n0     None           172.23.0.3   255.255.0.0   00:25:90:fd:3c:28
    ```

    NOTE: The preceding output has been truncated from the right for inclusion in this documentation.

Example:

The following example sets the MAC address of `n0`:

```
# cm node set --mac-address 00:25:90:04:4e:01 -n n0
```

Example:

The following example shows the MAC address of the node controller on `n1`:

```
# cm node show -B -n n1
NODE                  CARDIPADDRESS         CARDMACADDRESS       CARDTYPE
PROTOCOL
n1                    172.24.0.11           00:25:90:cd:7d:83    IPMI
dcmi,ipmi
```

Example:

The following example sets the MAC address of the node controller on `n0`:

```
# cm node set --bmc-mac-address 00:25:90:03:51:1d -n n0
```

14. Power up the replaced node.

Replacing failed system disks in a node that uses a disk drive for its root file system

About this task

The procedure in this topic explains how to replace system disks. The procedure assumes that the rest of the node is operating appropriately. You can reinstall the system disks into a replacement node.

Procedure

1. Verify that you have an appropriate spare system disk.

   If necessary, obtain new system disks from HPE.

   A spare system disk is equivalent to one of the system disks on your running cluster. The spare sits on a shelf. The cold spare is intended to be used in an emergency.

   Make sure that HPE supplied your spares. HPE does not support spares not supplied by HPE.

2. Connect a keyboard, video screen, and mouse to the node.
3. Power down the node that contains the failed system disks.
4. Remove the failed system disks from the node.
5. Install the new system disks into the node.
6. Use the node controller to connect to the failed node or attach a keyboard, video screen, and mouse to the failed node.

   If your system disks were part of a RAID, use the RAID controller interface to configure the disks into a RAID. The RAID controller interface is often part of the BIOS. See the RAID documentation for the node.

7. Power up the node.
8. (Conditional) Configure the RAID controller to use the new system disks.
9. From the admin node, install the cluster manager software on the new system disks:

   `cm node provision -n` *`hostname`* `-i` *`image`*

The variables are as follows:

Variable	Specification
hostname	The hostname of the node with the new system disks.
image	The name of the image that had been installed on the failed system disks.

Replacing a node and reinstalling the original system disks

About this task

Use the procedure in this topic if a node is no longer functioning, but the system disks within the node are still useful. This procedure explains the following:

- Removing good disks from a failed node
- Preserving the removed disks
- Installing the preserved disks from the failed node into a new node

Procedure

1. Verify that you have an appropriate spare node.

   A cold spare node is equivalent to one of the nodes on your running cluster. The spare sits on a shelf or is a factory preinstalled node. The cold spare is intended to be used in an emergency.

   Make sure that HPE supplied your spares. HPE does not support spares not supplied by HPE.

   As part of maintaining the cluster, make sure that you always have the following types of spare nodes:

   - One spare for the admin node.
   - One spare for a compute node.

   The following are some reasons to have the two types of spares:

   - Admin node BIOS settings are different from BIOS settings for other nodes.

     For example, the boot order is different for each node type.

     Attempts to configure the node into a cluster will fail.
   - Depending on your site policy, the node controllers of an admin node might or might not be configured to use DHCP by default.
   - The management cards of compute nodes must be configured to use DHCP by default.

     Otherwise, attempts to configure the node into the cluster will fail.

2. Connect a keyboard, video screen, and mouse to the node.
3. If possible, power down the failed node.
4. Examine and label the power cables.

Before you disconnect any cables, make sure they are labeled. Make sure that you are familiar enough with the cabling to re-cable the new node at the end of this procedure.

5. Disconnect all power cables.
6. Unplug the Ethernet cables used for system management.

   To avoid confusing them, note the plug number and label the cables. It is important that they stay in the same jacks in the new node. This connection is vital to proper system management and communication.

   NOTE: The Ethernet cables must be connected to the same plugs on the cold spare unit.

7. Remove any peripheral components, such as a keyboard, video screen, or mouse, from the node.
8. Remove the failed node from the rack.
9. Remove the system disks from the failed node.

   That is, open the failed node and remove the system disks.
10. Remove the system disks from the new node.

    That is, pull the current system disks, using their carriers, and set the disks aside.
11. Insert the preserved disks from the failed node into the new node (the shelf spare).
12. Insert the new node, with the preserved disks, back into the rack.
13. Connect AC power to the new node.
14. Connect a keyboard, video screen, and mouse to the new node.
15. From the admin node, update the cluster manager database.

    Update the following in the cluster database:

    - The MAC address of the spare
    - The MAC address of the node controller in the spare

    When you update the preceding address information in the database, you ensure that the cold spare can boot and function properly. If necessary, use the BIOS to retrieve the new MAC addresses. For more information about how to retrieve the MAC address of the spare, see the node controller documentation for the spare. For example, see the iLO server guide for the spare.

    From the admin node, query and set the MAC addresses in the database. The following table shows the command parameters that you can use:

    Example 1. The following example displays the MAC address of compute node `n0`:

    ```
    # cm node show -M -n n0
    NODE   NETWORK.NAME   IPADDRESS    SUBNETMASK    MACADDRESS
    n0     None           172.23.0.3   255.255.0.0   00:25:90:fd:3c:28
    ```

    NOTE: The preceding output has been truncated from the right for inclusion in this documentation.

    Example 2. The following example sets the MAC address of `n0`:

    ```
    # cm node set --mac-address 00:25:90:04:4e:01 -n n0
    ```

Example 3. The following example shows the MAC address of the node controller on `n1`:

```
# cm node show -B -n n1
NODE                  CARDIPADDRESS         CARDMACADDRESS       CARDTYPE
PROTOCOL
n1                    172.24.0.11           00:25:90:cd:7d:83    IPMI
dcmi,ipmi
```

Example 4. The following example sets the MAC address of the node controller on `n0`:

```
# cm node set --bmc-mac-address 00:25:90:03:51:1d -n n0
```

16. Power up the replaced node.
17. (Conditional) Interrupt the boot-up sequence in BIOS and enter the RAID configuration tool.

    Complete this step if the disk or disks being replaced represent a RAID configuration.

    The RAID controller facilitates the importing of drives and volumes into the new node. After the RAID is configured, the node might reboot or you might have to reset the node. Typically, the node boots normally.

    For information, see the RAID documentation for the node.

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

  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 Account 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 Tech Care Service

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.

For more information about CSR, contact your local service provider.

Warranty information

To view the warranty information for your product, see the **warranty check tool**.

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 (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. This process captures all document information.

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.

Installing the operating system and the cluster manager separately

Procedure

1. **Preparing to install the operating system and the cluster manager separately**
2. **Installing and configuring the operating system**
3. **Installing the cluster manager**

Preparing to install the operating system and the cluster manager separately

Procedure

1. Verify that this installation method can work for you by making sure that all of the following are true:

   - You want to install the operating system yourself so you can customize it.
   - You need only one slot. This procedure results in only one slot.
   - You do not need a high availability admin node.
   - You want to install the admin node manually.

2. Use your hardware documentation to connect the cluster hardware to your site network, and assign roles to each server.

   The admin node needs access to the following:

   - Compute nodes
   - Node controllers (the iLOs or baseboard management controllers (BMCs))
   - GUI clients

   Although it is not strictly required, each component type typically resides on a separate network. Using independent networks ensures good network performance and isolates problems if network failures occur.

   Configure the NICs on the admin node as follows:

   - Connect one NIC to a network established for compute node administration. The IP address of this NIC is needed during configuration of the admin node.
   - Connect a second NIC to the network connecting the admin node to the GUI clients.
   - A third NIC is typically used to provide access to the network connecting all the compute node controllers.

3. Use one of the following methods to prepare the installation files:

   - **Method 1 - Download the cluster manager release software**

If you download the cluster manager `.iso` file to a network location at your site, you can install from that network location or you can write the installation files to a USB device. To complete the installation over a network connection, modify the instructions accordingly.

- **Method 2 - Obtain a media kit from HPE**

  If you obtain a media kit from HPE, the media kit includes installation DVDs . If you obtain a media kit, use the instructions in the `README` file on the installation DVDs to create a `cm-admin-install .iso` file.

  You can install the software from that network location or you can write the installation files to a USB device. To complete the installation over a network connection, modify the instructions accordingly.

The installation instructions assume that you have the software on a USB device. To write the installation software to a USB device, use the instructions in the following topic:

(Conditional) Preparing a USB device

For more information about preinstallation steps, see the following:

Preparing to install the operating system and the cluster manager simultaneously on the admin node

Installing and configuring the operating system

Procedure

1. Obtain the operating system installation software.

   For information about the operating system installation software, see the following:

   HPE Performance Cluster Manager operating system releases supported

2. Install an operating system on the admin node with the following characteristics:

   - Create a static IP address on the admin node.
   - Configure the admin node to use the network time protocol (NTP) server at your site. Configure the time zone for your site.
   - Set the admin node to use your site domain name server (DNS).
   - For the internal traffic between the admin node and other nodes, allow all incoming and outgoing traffic. Configure the admin node NIC as a trusted interface or internal zone.
   - If you install the RHEL operating system on the admin node, do not configure SELinux. The cluster manager disables SELinux.
   - Configure the root file system with enough space to hold all the system images the cluster needs.
   - Design the operating system as a conventional operating system with typical installation packages.
   - Ensure that the Java 17 packages are installed.

Installing the cluster manager

About this task

The following procedure explains how to run the installation script that installs the cluster manager on the admin node.

Procedure

1. Obtain the cluster manager installation software from the following website:

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

2. Insert the cluster manager admin installation USB device into the USB drive in the admin node or mount the cluster manager admin installation `.iso` image.

   If you downloaded the cluster manager software, be aware that the `.iso` files on the HPE website use the HPE part number format (for example, `Q9V62-11049.iso`). You can rename the files before or after download. If you download the software as an `.iso` file, use the `mount` command to mount the files and give the download a more descriptive name.

   In the following example, the `mount` command specifies a new, more descriptive name for the `.iso` files:

   ```
   # ls -lh
   .
   .
   -rw-r---r-- 1 linuxdev linuxdev 6.5G Apr 1 02:47 cm-admin-install-1.10-rhel8.X-x86_64.iso
   .
   .
   .
   # mount -o ro,loop cm-admin-install-1.10-rhel8.X-x86_64.iso /mnt
   ```

   For more information about HPE part numbers, see the cluster manager release notes.

3. Enter the following command to change your working directory to the mount point:

   ```
   # cd /mnt
   ```

4. Enter the following command to start the installation script:

   ```
   # ./standalone-install.sh
   ```

   The installation script starts and begins the installation. The script is included on the cluster manager admin installation USB device and on the corresponding `.iso` files. Respond to the prompts for the following information:

   a. The full path to the operating system distribution `.iso` file.

      This is the path that you used to install the admin node. The installer uses this information to set up repositories and start the installation.

   b. After the installer lists all the packages, the installer prompts you to enter **y** or **n** to proceed.

   c. After all packages are added, the installer prompts you to log out and log back in again.

   The script prompts you for information from time to time. Respond to the prompts with information about your cluster environment.

5. Reboot the cluster.

   For example:

   ```
   # reboot
   ```

6. Proceed to one of the following:

   - **Using the cluster definition file to specify the cluster configuration**

     If you have a copy of the cluster definition file, use this procedure.

   - **Using the menu-driven cluster configuration tool to specify the cluster configuration**

     If you do not have a copy of the cluster definition file, use this procedure. This procedure guides you through the menu-driven configuration tool.

Upgrading the operating system and reinstalling the cluster manager

About this task

There are situations in which you might want to reinstall all or most of the software on a factory-configured cluster. A common case is the following:

- You are satisfied with the cluster configuration. That is, you want to reinstall the cluster system as it was configured at the factory with a minimum of changes.

  And

- You want to upgrade the cluster operating system to the next major release.

The procedure in this topic assumes that the cluster is intact and that you can back up the configuration files you need.

Procedure

1. Back up the cluster.

   For information about how to back up the cluster, see the following:

   HPE Performance Cluster Manager Administration Guide

2. Install the cluster software on the admin node.

   a. Complete the procedures in the following topics:

      - **Installing the operating system and the cluster manager simultaneously on the admin node**
      - **(Optional) Configuring a system admin controller high availability (SAC HA) admin node**. Complete this procedure if the admin node is an HA admin node.
      - **Configuring the cluster software on the admin node**

   b. (Conditional) Add operating system updates or cluster manager updates.

3. (Conditional) Preserve the existing switch configuration.

   Complete this step if you want to retain the current network, VLAN, and IP configuration in the cluster. That is, complete this step if you did not reset the management switches when you backed up the configuration.

   Enter the following command to omit the switch configuration:

   ```
   # cadmin --enable-discover-skip-switchconfig
   ```

   The command in this step ensures that the cluster manager does not overwrite or configure new settings on the management switches that are added back to the cluster.

4. Run the `cm node add` command to configure the cluster.

5. (Conditional) Plug in the redundant cables.

   Complete this step if you disconnected the redundant cables earlier in this procedure.

NOTE: If the network becomes unstable when adding the redundant cabling, you can attempt to reconfigure the switches in the foreground and watch the progress. To watch the progress, enter the following command:

`switchconfig_configure_node --node mgmtsw`*X*

For *X*, specify the number of the management switch.

For example, to reconfigure `mgmtsw0` and `mgmtsw1`, issue the following command:

```
# switchconfig_configure_node --node mgmtsw0,mgmtsw1
```

6. Direct the system to enable top-level switch configuration when you run a node discovery command in the future:

   ```
   # cadmin --disable-discover-skip-switchconfig
   ```

7. (Conditional) Recreate or import custom images, repositories, or files that you backed up.

   Complete this step as needed.

   For example, if you have custom repositories for NVIDIA or Mellanox OFED, copy back the repositories that you copied off.

   Import images, add or recreate repositories, and create custom images as necessary.

8. Enter the following command to reboot the cluster:

   ```
   # cm power reboot -t system
   ```

Subnetwork information

Cluster hardware components can be connected to multiple networks. Generally, a network is assigned to a single subnet.

A **subnet** is a logical subdivision of an IP network. A subnet keeps broadcast traffic from the various hosts within the subnet contained in its own subnet. This action helps clusters to scale properly. Additionally, if layer-3 IP routing is not configured, the components that reside in a subnet can communicate only with other components within the subnet.

The cluster management software uses a variety of networking concepts to accomplish the architecture design goals for various cluster types. These concepts include the following:

- Virtual Local Area Network (VLAN / 802.1Q) tagging
- Supernetting
- Layer 3 IP routing
- Subinterfaces

Network and subnet information within a cluster

Table 7: Network and subnet information shows the following for the networks that the cluster management software uses:

- The names of the components on the networks
- The default allocation of the system-wide IP address ranges on the networks

Generally, a node and its node controller reside in the same VLAN. However, these components do not reside in the same IP subnet range. This separation prevents cross-communication between the host node and its management interface.

Table 7: Network and subnet information shows the components in a given VLAN as either an **untagged port** or a **tagged port**.

The following are additional notes:

- At a minimum, all switchports are put into a VLAN as an untagged port. This is also known as a **native VLAN** or a **default VLAN** in some networking nomenclature.
- All traffic coming from a component that is otherwise untagged is put into an untagged VLAN.
- A switchport is not required to allow tagged VLANs.

  Some switchports allow a tagged VLAN. These switchports forward the traffic coming out of the VLAN when the traffic coming out of a component with a VLAN tag matches the switchport configuration.
- A switchport can allow zero, one, two, or many tagged VLANs at the same time.

Table 7: Network and subnet information

VLAN #	Subnet name	IP range / subnet mask	Nodes in Subnet
1	`head`	`172.23.0.0/16`	Admin `bond0` (untagged) Generic compute (untagged) Management switch VLAN 1 (untagged)
1	`head-bmc`	`172.24.0.0/16`	Admin `bond0:bmc` (untagged) Generic compute node controller (untagged) Cooling devices such as ARCS (untagged) Cooling devices such as CDUs (untagged)
N/A	`ib0`	`10.148.0.0/16`	Any component with InfiniBand interfaces
N/A	`ib1`	`10.149.0.0/16`	Any component with InfiniBand interfaces

Naming conventions

The cluster management software has the following default naming formats for various components found within a cluster:

- If a component is configured into the cluster by using a node discovery command, you can specify a custom hostname of your choice.

  If you do not specify a hostname, the cluster management software uses the default naming convention.
- If a component is automatically added to the database, the naming convention is predetermined and cannot be specified at this time.

In a cluster definition file, for any given component, you can append the following parameter to assign a custom hostname to any component:

`hostname1=`*`hostname`*

For example, an entry for a compute node might look like this:

```
internal_name=service1,mgmt_net_name=head,hostname1=r01n01,...
```

In the preceding example, the ellipsis (`...`) at the end represents the fact that you could specify many other configuration attributes on this line.

Table 8: Naming conventions includes information about various components, default naming conventions for each component, and the type of cluster in which these components can be found. The variables *T*, *X*, *Y*, and *Z* are always positive integer numbers. The examples represent the hostnames that can be seen once a component is added to the cluster management software.

Table 8: Naming conventions

Component	Internal name format	Examples	Found in

Admin node	*admin*	`myadmin`
		`sleet`
		`snow`
			All clusters

Ethernet management switch	`mgmtsw`*X*	`mgmtsw0` (spine) `mgmtsw1` (leaf)	All clusters
Ethernet data switch	`dataswX`	`datasw0` (spine) `datasw1` (leaf)	Clusters with an Ethernet high-speed fabric
InfiniBand data switch	`ibsw`*X*	`ibsw0` `ibsw1`	Clusters with an InfiniBand high-speed fabric
Compute node	`service`*X*	`service1` `service100`	Clusters with generic compute resources
Node controller interfaces	`service`*X*`-bmc`	`service1-bmc`	Clusters that contain components that have a node controller. Node controllers can be of type iLO or BMC.
Power distribution unit (PDU)	`pdu`*X*	`pdu0` `pdu1`	Clusters with PDU hardware
HPE Adaptive Rack Cooling System (ARCS) device	`cooldev`*X*	`cooldev0`, `cooldev1`	Clusters other than HPE Apollo 2000 clusters.

Default partition layout information

The default partition layout uses the GUID partition table (GPT) and the GRUB version 2 boot system. Alternatively, to create a custom partitioning scheme for the cluster, see the following:

(Optional) Configuring custom partitions on the admin node

Partition layout for a one-slot cluster

Table 9: Partition layout for a single-boot cluster shows the partition layout for a one-slot cluster. This layout yields one boot partition. If you configure a single-slot system and later decide to add another partition, the addition process destroys all the data on your system.

Table 9: Partition layout for a single-boot cluster

Partition	File system type	File system label	Notes
1	Ext4	`sgidata`	Contains slot information. On the admin node, contains GRUB version 2 data for choosing root slots at boot time.
2	swap	`sgiswap`	Swap partition.
3-10	N/A	N/A	N/A
11	Ext4	`sgiboot`	Slot 1 `/boot` partition.
12-20	N/A	N/A	N/A
21	VFAT	`sgiefi`	Notice that the `/boot/efi` partition is used only on systems with UEFI BIOS.
22-30	N/A	N/A	N/A
31	XFS	`sgiroot`	Slot 1 `/` partition.

Partition layout for a two-slot cluster

Table 10: Partition layout for a dual-boot cluster shows the partition layout for a two-slot cluster. This layout yields two boot partitions.

Table 10: Partition layout for a dual-boot cluster

Partition	File system type	File system label	Notes
1	Ext4	`sgidata`	Contains slot information. On the admin node, contains GRUB version 2 data for choosing root slots at boot time.
2	swap	`sgiswap`	Swap partition.
3-10	N/A	N/A	N/A
11	Ext4	`sgiboot`	Slot 1 `/boot` partition.
12	Ext4	`sgiboot2`	Slot 2 `/boot` partition.
13-20	N/A	N/A	N/A
21	VFAT	`sgiefi`	Slot 1 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
22	VFAT	`sgiefi2`	Slot 2 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
23-30	N/A	N/A	N/A
31	XFS	`sgiroot`	Slot 1 `/` partition.
32	XFS	`sgiroot2`	Slot 2 `/` partition.

Partition layout for a five-slot cluster

Table 11: Partition layout for a quintuple-boot cluster shows the partition layout for a five-slot cluster. This layout yields five boot partitions.

Table 11: Partition layout for a quintuple-boot cluster

Partition	File system type	File system label	Notes
1	Ext4	`sgidata`	Contains slot information. On the admin node, contains GRUB version 2 for choosing root slots at boot time.
2	swap	`sgiswap`	Swap partition.
3-10	N/A	N/A	N/A
11	Ext4	`sgiboot`	Slot 1 `/boot` partition.
12	Ext4	`sgiboot2`	Slot 2 `/boot` partition.
13	Ext4	`sgiboot3`	Slot 3 `/boot` partition.
14	Ext4	`sgiboot4`	Slot 4 `/boot` partition.
15	Ext4	`sgiboot5`	Slot 5 `/boot` partition.
16-20	N/A	N/A	N/A
21	VFAT	`sgiefi`	Slot 1 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
22	VFAT	`sgiefi2`	Slot 2 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
23	VFAT	`sgiefi3`	Slot 3 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.

Table Continued

Partition	File system type	File system label	Notes
24	VFAT	`sgiefi4`	Slot 4 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
25	VFAT	`sgiefi5`	Slot 5 `/boot/efi` partition. EFI BIOS clusters only. On x86_64 BIOS clusters, this partition is unused.
26-30	N/A	N/A	N/A
31	XFS	`sgiroot`	Slot 1 / partition.
32	XFS	`sgiroot2`	Slot 2 / partition.
33	XFS	`sgiroot3`	Slot 3 / partition.
34	XFS	`sgiroot4`	Slot 4 / partition.
35	XFS	`sgiroot5`	Slot 5 / partition.

Specifying configuration attributes

The cluster manager include the following types of configuration attributes:

- Node attributes
- Cluster attributes

The cluster definition file includes all the node attributes and cluster attributes assigned in the cluster. In the cluster definition file, node attributes are found in the following sections:

- `[discover]`
- `[nic_templates]`
- `[templates]`

In the cluster definition file, cluster attributes are found in the following sections:

- `[attributes]`
- `[dns]`
- `[images]`
- `[networks]`

To obtain a copy of the cluster definition file, enter the following command:

```
# cm system show configfile --all
```

The cluster manager commands and utilities use the configuration attributes to define individual nodes and to define the general cluster in the following way:

- The `configure-cluster` command, starts a menu-driven utility that you can use at any time to specify cluster attributes. The command is as follows:

  ```
  # configure-cluster
  ```

  Alternatively, you can provide a cluster definition file, populated with attributes, as input to the `configure-cluster` command. The format is as follows:

  `configure-cluster --configfile` *`cluster_definition file`*

- The `cm node add` command and the `cm node discover` command assign node attributes to nodes when they configure nodes into the cluster.

NOTE: In many cases, you can set or clear attributes on a command line. On a command line, the attribute name often uses underscore characters (_). In the cluster definition file, the attribute name often uses hyphens (–). For more information, see the manpages for the individual commands.

Provisioning attributes

image

Specifies the image for a node.

Values = The name of the image.

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

kernel

Specifies the kernel for a node.

Values = The version of the kernel.

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

nfs_writable_type

Specifies the type of writable area for NFS root file systems. Only valid when `rootfs=nfs` is in effect. For more information, see the `cinstallman`(1) manpage.

Values = `nfs-overmount`, `nfs-overlay`, `tmpfs-overmount`, or `tmpfs-overlay`.

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command

- `cm node add` command
- `cm node set` command
- `cm node show` command

rootfs

Sets the root file system type for a node. For more information, see the `cinstallman`(1) manpage.

Values = `disk`, `tmpfs`, `nfs`, `custom`

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

tpm_boot

Enables the node to boot, or not, as a trusted platform module (TPM).

Values = `yes` or `no`

Default = `no`

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

transport

Sets the image transport method.

Values = `rsync`, `bt`, `udpcast`

Default = `bt`

Range = N/A

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

Management network attributes

The management network attributes provide information about the redundant management network and the switch management network.

`redundant_mgmt_network`

Specifies the default setting for the redundant management network. If no value is supplied at configuration time, the installer populates all nodes with the default value.

Values = `yes` (default) or `no`.

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

`switch_mgmt_network`

Specifies the default setting for the switch management network. If no value is supplied at configuration time, the installer populates all nodes with the default value.

Values = `yes` (default) or `no`.

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

Console server attributes

The admin node is a management node.

On a management node, there are files in the `/var/log/consoles` directory for each subordinate node. The files contain log information from the node controllers on the subordinate nodes.

On the admin node, the `/var/log/consoles` directory contains log information for each node under admin node control.

The console server options let you control the quantity and frequency of log information that is collected. The cluster manager software logs node controller output to the `/var/log/consoles` directory. In the `/var/log/consoles` directory, there is a file for each node in the cluster. If you tune the console server options, you can limit the amount of traffic between the console and the cluster. Set these options if you want to minimize network contention.

conserver_logging

Specifies console server logging. If set to `yes`, the console server logs messages to the console through IPMItool. This feature uses some network bandwidth.

Values = `yes` (default) or `no`.

Accepted by:

- Cluster definition file
- `cadmin` command
- `cattr` command
- `cm node add` command
- `cm node set` command

conserver_ondemand

Specifies console server logging frequency. When set to `no`, logging is enabled all the time. When set to `yes`, logging is enabled only when someone is connected.

Values = `yes` or `no` (default).

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

conserver_timestamp

Prints a date and time in front of every line of console output in the console log managed by the console server.

Values = `yes` or `no` (default).

Enter the following commands to enable `conserver_timestamp`:

1. `# cattr set CONSERVER_TIMESTAMP yes`
2. `# cm node update config --sync conserver -n admin`

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command

console_device

Specifies the console device.

Values = the device hostname

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

Networking attributes

mgmt_bmc_net_if

Assigns an IP address to the management interface in the `mgmt-bmc` network with a label of `bmc`. Typically used so a node, such as a fabric management node (FMN), that needs to reach devices in the `mgmt-bmc` network can obtain an IP address to get to the `mgmt-bmc` network if the `mgmt-bmc` gateway is set.

Values = `yes` or `no`.

Default = `no`.

Range = NA

Accepted by:

- Cluster definition file
- `cm node nic set` command
- `cm node nic show` command
- `cm node add` command

mgmt_bmc_net_if_interface

Specifies the NIC to use if the `mgmt-bmc` network is separate from the management network. Typically, this attribute is not used. By default, the interface is on the bonded NIC in the management network.

Values = the name of the NIC

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cm node nic set` command
- `cm node nic show` command
- `cm node add` command

mgmt_bmc_net_if_ip

Specifies the management card (iLO or BMC) IP address.

Values = The IP address of the interface

Default = None. If not specified, the next available IP address in the management BMC network is assigned.

Range = NA

Accepted by:

- Cluster definition file
- `cm node nic set` command
- `cm node nic show` command
- `cm node add` command

mgmt_net_interfaces

Configures the system to associate and set up the specified interface or interfaces in Linux.

By default, `eth0` and `eth1` are used on compute nodes.

For more information about predictable interface naming, see the documentation for your operating system. Generally, different types of compute hardware and NICs have different naming styles. Common formats include the following:

- `eno#`
- `ens#f#`
- `enp#f#`

If you specify more than one address, include the addresses in a comma-separated string, and enclose the string in quotation marks (`" "`). If you use quotation marks on a command line, remember that quotation marks must be escaped with backslash (`\`) characters. If using predictable network names, the specified names are used.

Values = At least one, and up to 64 interface hostnames. If you specify more than one interface, use a comma (`,`) to separate hostnames.

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

mgmt_net_macs

Specifies MAC addresses for the management network. If you specify more than one, include the addresses in a comma-separated string, and enclose the string in quotation marks (`" "`). If you use quotation marks on a command line, remember that quotation marks must be escaped with backslash (`\`) characters. Specify to avoid network sniffing discovery.

Values = the interface MAC address or MAC addresses

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

mgmt_net_name

Specifies the name of the management network.

Used primarily on large clusters without leader nodes that have multiple routed networks for management.

Values = `head` or a network name

Default = `head`

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command

- `cm node set` command
- `cm node show` command

mtu

Specifies jumbo frames on any network. Specify this attribute in the `[networks]` section of the cluster definition file. All nodes that you include in this network inherit the `mtu` setting for the network.

Values = An integer value.

Default = 1500

Range = 68 through 65,536

Accepted by:

- Cluster definition file
- `cm network add` command
- `cm network set` command

network_group

You can use the `network_group` attribute in the `[discover]` section or the `[templates]` section of the cluster definition file. When specified, the installer creates a network group and assigns the nodes to the network group during installation. The installer configures the nodes into the network group you specify.

To use native monitoring, configure the compute nodes into network groups. It is common to configure all the compute nodes under a common switch into a network group. Limit the number of nodes in a network group to 519.

Values = any network group name. Specify one network group name. A node can appear in only one network group.

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cm group network` command
- `cm node set` command
- `cm node show` command

For information about network groups, see the following:

HPE Performance Cluster Manager Administration Guide

Monitoring attributes

monitoring_kafka_elk_alerta_enabled

Specifies whether monitoring, through Kafka, OpenSearch, and Alerta are enabled in the cluster. When set to `yes`, monitoring is enabled.

Values = `yes` or `no` (default).

Accepted by:

- Cluster definition file
- `cm monitoring kafka` command
- `cm monitoring elk` command
- `cm monitoring alerta` command

monitoring_native_enabled

Specifies whether the cluster manager native monitoring is enabled in the cluster. When set to `yes`, monitoring is enabled.

Values = `yes` or `no` (default).

Accepted by:

Cluster definition file

monitoring_timescale_access

Specifies whether the node is running TimescaleDB/Postgres and is configured as an access node.

Values = `yes` or `no` (default).

Accepted by:

`cm monitoring timescaledb` command

monitoring_timescale_data

Specifies that the node is configured to host a TimescaleDB data node.

Values = `yes` or `no` (default).

Accepted by:

`cm monitoring timescaledb` command

monitoring_timescale_enabled

Specifies whether monitoring, through TimescaleDB is enabled in the cluster. When set to `yes`, monitoring is enabled.

Values = `yes` or `no` (default).

Accepted by:

`cm monitoring timescaledb` command

Switch attributes

discover_skip_switchconfig

Signals the installer to omit the switch configuration steps. When set to `yes`, the installer does not configure the switches. Set this option to `yes` when you want to perform a quick configuration change, but you do not need to update the switch configuration. This value is not saved in the cluster definition file, but it can be specified there.

Values = `yes` or `no`

Default = `no`

Accepted by:

- Cluster definition file
- `cm node add` command

mgmtsw_isls

Configures all ISL LAGs when the switch is configured into the cluster.

Values:

- For a single LAG, separate ports by a comma.

  For example: `mgmtsw_isls="X/X/p1,X/X/p2"`
- For multiple LAGs, use a semi-colon `(;)` to separate the LAGS.

  For example: `mgmtsw_isls="X/X/p1,X/X/p2;X/X/p3,X/X/p4"`

Default = NA

Range = A comma-separated list with valid switch port syntax such as `1/1/1` or `1/1`

Accepted by:

- Cluster definition file
- `cm node add` command

mgmtsw_partner

Specifies the partner of a VXS partner management switch. Specify this attribute if you have also specified the `type=dual-spine` or `type=dual-leaf` attributes.

Values = the hostname of a VSX partner management switch

Default = NA

Accepted by:

- Cluster definition file
- `cm node add` command

net

For management switches, specifies the name of the served management network when configuring a management leaf switch that is dedicated to the supplied management networks.

Values = `head`, `head-bmc`, `head/head-bmc`, `ib0`, or `ib1`

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command

type

Specifies the type of management switches. If not specified for a management switch, the cluster manager uses link layer discovery protocol (LLDP) to determine which switch is connected directly to the admin node.

There can be only one spine switch or one dual-spine pair of switches per cluster.

Values = `dual-spine`, `dual-leaf`, `spine`, or `leaf`

Default = NA

Accepted by:

- Cluster definition file
- `cm node add` command

Miscellaneous attributes

alias_groups

Defines node aliases, which are additional ways to refer to a node, at cluster configuration time. This attribute adds the `alias_groups` keyword within the node definitions of the `[discover]` section of the cluster definition file.

The format is as follows:

`alias_groups="`*`group1`*`:`*`alias1`*`,`*`group2`*`:`*`alias2`*`,..."`

The variables are as follows:

Variable	Specification
groupX	The name for the group of nodes.
aliasX	The name for the individual nodes.

For example, in the `[discover]` section, a node with `hostname1=node1` could define an aliases called `work1` in the alias group `work-nodes` by adding the following to the configuration definition file:

```
alias_groups="work-nodes:work1"
```

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

architecture

Specifies the processor architecture type on a node.

Values = `x86_64` or `aarch64`

Default = `x86_64`

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

baud_rate

Specifies the baud rate of the serial console device.

Values = a positive integer value

Default = `115200`

NOTE: This attribute is required.

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

bmc_password

Specifies the node controller password.

Values = a cluster-specific node controller password.

NOTE: This attribute is case sensitive and is required.

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

bmc_username

Specifies the node controller username.

Values = a cluster-specific node controller username.

NOTE: This attribute is case sensitive and is required.

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

card_type

Specifies the type of node controller in the node.

Values = Valid values include `IPMI` (default), `bmx`, `iLO`, and `ILOCM`. Specifically, the cluster manager supports the card types defined in the following file:

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

In the `cmuserver.conf` file, see the `CMU_VALID_HARDWARE_TYPES` field.

Default = `IPMI`.

NOTE: This attribute is case sensitive and is required.

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

cluster_domain

This configuration attribute specifies the cluster domain name. Hewlett Packard Enterprise recommends that users change this value.

Values = must be a standard domain name.

Accepted by:

- Cluster definition file
- Cluster configuration tool
- `cadmin` command
- `cattr`

custom_groups

You can use the `custom_groups` attribute in the `[discover]` section or the `[templates]` section of the cluster definition file. When specified, the installer creates custom node groups during installation and configures nodes into the custom groups you identify.

Values = any custom group name, as follows:

- To specify one custom group, use the following syntax:

  `custom_groups=`*`name`*

- To specify two or more custom groups, use the following syntax:

  `custom_groups="`*`name`*`,`*`name`*`"`

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cm group custom set` command
- `cm node` commands such as `cm node set` and `cm node show`

custom_partitions

The custom partitioning feature is not enabled by default. The cluster manager supports this feature on nodes with root disks.

The `custom_partitions` variable specifies the name of the custom partition file. The file resides in the following directory:

`/opt/clmgr/image/scripts/pre-install`

Values = any file name that ends in `.cfg`

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

dhcp_bootfile

This attribute specifies whether to load iPXE (the default) or GRUB version 2 first when a node is first configured.

In some cases, a node can fail to boot over the network with the default settings. For example, a node might hang when it tries to load the kernel and `initrd` during the boot from its system disk. In this case, modify the DHCP boot file setting.

The settings are as follows:

- The default boot loader is `dhcp_bootfile=ipxe-direct`. In this case, the server boot agent uses a special iPXE binary on UEFI and legacy BIOS systems to directly load the kernel and `initrd`. It avoids GRUB version 2.

  Arm (AArch64) nodes and HPE Apollo 20 compute nodes require `ipxe-direct`.
- On legacy (non-UEFI) x86_64 systems, you can specify `dhcp_bootfile=ipxe` to work around some boot problems by network-booting iPXE and letting iPXE load GRUB version 2 over the network.

  For more information, see the following:

  Nodes fail to boot
- When you specify `dhcp_bootfile=grub2`, GRUB version 2 loads first.

  HPE Apollo 80 nodes require `dhcp_bootfile=grub2`

Use the `cm node set` command to specify the new boot order.

Values = `ipxe-direct` (default), `ipxe`, and `grub2`.

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

The following other topics might also interest you:

- **dhcpd_default_lease_time**
- **dhcpd_max_lease_time**
- **Nodes fail to boot**

dhcpd_default_lease_time

This attribute specifies the default DHCP lease time, which the cluster manager sets at 180 seconds.

The cluster manager includes other configuration attributes that let you control DHCP, but Hewlett Packard Enterprise recommends that you use the default values whenever possible.

You can also use the `cadmin` command to set this default lease time.

Values = `180` seconds (default) or another integer value.

Accepted by:

- Cluster definition file
- `cadmin` command

The following topics might also interest you:

- **dhcp_bootfile**
- **dhcpd_max_lease_time**
- **Nodes fail to boot**

dhcpd_max_lease_time

This attribute specifies the maximum DHCP lease time, which the cluster manager sets at 300 seconds.

The cluster manager includes other configuration attributes that let you control DHCP, but Hewlett Packard Enterprise recommends that you use the default values whenever possible.

You can also use the `cadmin` command to set this maximum lease time.

Values = `300` seconds (default) or another integer value.

Accepted by:

- Cluster definition file
- `cadmin` command

The following topics might also interest you:

- **dhcp_bootfile**
- **dhcpd_default_lease_time**
- **Nodes fail to boot**

disk_bootloader

After installation, specifies whether the node can boot from the on-disk bootloader. When enabled, it is no longer possible to control kernel boot parameters centrally.

Values = `yes` or `no`. If the node uses an NFS root file system or a `tmpfs` root file system, specify `disk_bootloader=no`.

Default = `no`

Range = NA

Accepted by:

- Cluster definition file
- `cm node add` command
- `cm node set` command
- `cm node show` command

domain_search_path

Specifies the domain search path for the cluster.

Values = one or more domains. If you specify multiple domains, use a comma (`,`) to separate each domain.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node show` command

geolocation

Specifies information about the physical location of the node at your site in a human-readable form.

Values = An alphanumeric string of up to 128 characters. The string can include spaces and special characters. If you include spaces, enclose the string in quotation mark (") characters.

For more information, see the following:

(Conditional) Configuring power distribution units (PDUs) into the cluster

Default = NA

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cattr` command
- `cm node add` command

- `cm node set` command
- `cm node show` command

hostname1

Specifies a site-specific, custom hostname for a node. Users can specify this name when they want to log into the node. The hostname appears in most cluster manager output.

Values = a hostname

Default =

Range = NA

Accepted by:

- Cluster definition file
- `cadmin` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

internal_name

Defines the function of a component in the cluster definition file. Formerly `temponame` (now deprecated). This name can match the hostname. This name never changes for the life of the node.

Values = A name such as `service0` or `mgmtsw0`.

Default = NA

Range = NA

Accepted by:

Cluster definition file

kernel_distro_params

Specifies kernel parameters for the operating system distribution that runs on the cluster. Typically, this setting includes parameters suggested by the distribution.

You can set this parameter on either a node basis or an image basis. The node setting overrides the image setting.

Values = "*arg=value arg=value ...*"

Accepted by:

- `cadmin` command
- `cattr` command
- `cm node add` command

- `cm node set` command
- `cm node show` command

For example:

```
# cm node set --kernel-distro-params "cgroup_disable=memory max_cstate=1" --nodes service1
For ICE compute nodes, the image must be pushed to the leaders to take effect.
Refreshing netboot files for impacted nodes...

# cm node show --kernel-distro-params --nodes service1
cgroup_disable=memory,max_cstate=1

# cm node unset --kernel-distro-params --nodes service1
For ICE compute nodes, the image must be pushed to the leaders to take effect.
Refreshing netboot files for flat compute, leader nodes...

# cm node show --kernel-distro-params --nodes service1
None
```

For information about specifying parameters on a command line, see the help output (`-h`) for a specific command.

kernel_extra_params

Specifies additional kernel parameters for the operating system distribution that runs on the cluster. This attribute sets parameters in addition to the standard parameters suggested by the distribution.

You can set this parameter on either a node basis or an image basis. The node setting overrides the image setting.

Values = "*arg=value arg=value ...*"

Accepted by:

- `cadmin` command
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

For example:

```
# cm node set --kernel-extra-params "cgroup_disable=memory rescue-1 fips=1" --nodes service1
For ICE compute nodes, the image must be pushed to the leaders to take effect.
Refreshing netboot files for impacted nodes...

# cm node show --kernel-extra-params --nodes service1
cgroup_disable=memory,rescue-1,fips=1

# cm node unset --kernel-extra-params --nodes service1
For ICE compute nodes, the image must be pushed to the leaders to take effect.
Refreshing netboot files for flat compute, leader nodes...

# cm node show --kernel-extra-params --nodes service1
None
```

For information about specifying parameters on a command line, see the help output (`-h`) for a specific command.

For information about the federal information processing standards (FIPS) parameter used in the preceding examples, see the following:

HPE Performance Cluster Manager Administration Guide

mgmt_net_def_gw

Specifies whether there is a user-defined default gateway for a node or a set of nodes. The cluster manager assumes one of the following for the default gateway:

- By default, when you specify `mgmt_net_def_gw=yes` in the cluster definition file, the cluster manager uses the gateway value of the management network upon which the node is located.

  In the following example, the cluster manager uses the gateway value of the management network as the gateway for `service0`:

  ```
  internal_name=service0, mgmt_bmc_net_macs="aa:aa:aa:bb:cc:dd",
  mgmt_net_macs="aa:bb:cc:dd:ee:ff", template_name=compute, mgmt_net_def_gw=yes
  ```

- If you also specify both a `mgmt_net_def_gw=yes` and a `mgmt_net_def_gw_ip` configuration attribute for a node or nodes, the cluster manager uses the specified value of the `mgmt_net_def_gw_ip` configuration attribute as the gateway value.

  In the following example, the cluster manager uses `173.23.200.200` as the gateway value for `service1`:

  ```
  internal_name=service1, mgmt_bmc_net_macs="aa:aa:aa:bb:cc:ee", mgmt_net_macs="aa:bb:cc:dd:ee:11",
  template_name=compute, mgmt_net_def_gw=yes, mgmt_net_def_gw_ip=172.23.200.200
  ```

For more information, see the following:

mgmt_net_def_gw_ip

Values = `yes` or `no`

Default = `no`

Range = NA

Accepted by:

Cluster definition file

mgmt_net_def_gw_ip

Defines the default gateway IP address for a node or a set of nodes.

By default, the `mgmt_net_def_gw_ip` address is the value in the management network gateway. Use the `cm network show` command to display this value. For example:

```
# cm network show -d -w head | grep gateway
  gateway:              172.23.200.200
```

In the cluster definition file, if you specify the `mgmt_net_def_gw_ip` configuration attribute, also specify the `mgmt_net_def_gw` configuration attribute.

For more information, see the following:

mgmt_net_def_gw

Values = a single IP address in dotted-decimal format. For example: `172.23.255.254`.

Default = NA

Range = NA

Accepted by:

Cluster definition file

name

In the `[templates]` section of the cluster definition file, the `name=` field defines the name for a particular node template.

Values = a custom name for the node template

Default = NA

Range = NA

Accepted by:

Cluster definition file

node_notes

Specifies node-specific information in a human-readable form. You can include any information about the node in the note.

Values = An alphanumeric string of up to 8192 characters.

Default = NA

Range = NA

Accepted by:

- `cadmin` command
- `cattr` command
- `cm node set` command
- `cm node show` command

predictable_net_names

Specifies whether the cluster uses predictable network names to describe the network interface cards (NICs). When set to `yes`, predictable network names are enabled.

Values = `yes` (default) or `no`.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

template_name

Identifies the custom template for the cluster manager to use when configuring this node. The custom template is defined in the cluster definition file.

Values = the name of a template in the cluster definition file

Default = NA

Range = NA

Accepted by:

Cluster definition file

type

Specifies the type of management switches. If not specified for a management switch, the cluster manager uses link layer discovery protocol (LLDP) to determine which switch is connected directly to the admin node.

There can be only one spine switch or one dual-spine pair of switches per cluster.

Values = `dual-spine`, `dual-leaf`, `spine`, or `leaf`

Default = NA

Accepted by:

- Cluster definition file
- `cm node add` command

Predictable network interface card (NIC) names

By default, the cluster manager assigns **predictable names** to the Ethernet NICs within a node. This practice ensures that each NIC name is boot persistent. Predictable names are different for different types of nodes with different types of motherboards.

Predictable names are the same across like hardware. For example, if your cluster has only one type of compute node, then the predictable names are the same for all compute nodes in the cluster.

The cluster manager also supports legacy names as NIC names. For example, `eth0`, `eth1` are legacy names. Legacy NIC names can change when you boot the cluster. For example, assume that the cluster includes multiple adapters and NICs in a given node. For this cluster, the Linux mechanisms that maintain persistent names in the wanted order can fail to rename NICs properly.

NOTE: Do not mix predictable NIC names and legacy NIC names in the same cluster. The cluster manager does not use predictable names for InfiniBand devices.

The following table shows comparable predictable NIC names and legacy NIC names for an example cluster.

Table 12: Example cluster - using predictable NIC names and legacy NIC names

Node type and role	Network role	Example predictable name	Example legacy name
CH-C1104-GP2 admin node	House network	`ens20f0`	`eth0`
	Management #1	`ens20f1`	`eth1`
	Management #2	`ens20f2`	`eth2`
C1104-TY13 compute	Management network	`enp1s0f0`	`eth0`
	House network	`enp1s0f1`	`eth1`

The cluster manager includes the following commands for adding, deleting, modifying, and displaying NIC information:

- `cm node nic add`
- `cm node nic delete`
- `cm node nic set`
- `cm node nic show`

For information about these commands, enter the command and add the `-h` parameter. For example:

```
# cm node nic add -h
```

Configuring a new switch

About this task

New switches require some preliminary configuration before you configure them into a cluster. After you complete the preliminary configuration, you can run a node discovery command from the admin node.

The procedures in this topic apply to both stacked and nonstacked switches. Complete the procedures in this topic under the following circumstances:

- You want to add a switch to the cluster.
- You want to replace an existing switch for which you have no backup data. In this situation, proceed as if you want to add a switch.

The HPE Performance Cluster Manager Release Notes list the switches that are supported.

NOTE: To replace an existing switch for which you have backup data, use the procedure in the following:

HPE Performance Cluster Manager Administration Guide

Procedure

1. **(Conditional) Configuring an Extreme Networks switch**
2. **(Conditional) Configuring an HPE FlexFabric switch or an HPE FlexNetwork switch**
3. **Running the `cm node add` command for a new switch**

(Conditional) Configuring an Extreme Networks switch

Procedure

1. Access the switch through a console cable.
2. Log in with the default credentials.

   These credentials are one of the following:

   - Username = `admin`, password = `admin`

     Or

   - Username = `admin`, password = *<blank>*

     For *<blank>*, simply press Enter.

3. Enter the following commands:

```
enable dhcp vlan default
enable flooding all_cast ports all
enable jumbo-frame ports all
```

```
enable lldp ports all
enable loopback-mode vlan default
```

4. Enter the following command to retrieve the switch MAC address:

   ```
   show switch  | grep MAC
   ```

   For example:

   ```
   Slot-1 mgmtsw8.3 # show switch  | grep MAC
   System MAC:       02:04:96:8B:CC:A8
   ```

   Record the MAC address that this command returns. The cluster definition file requires you to specify the switch MAC address in a slightly different format. To specify the MAC address in this example in the cluster definition file, reformat the address as follows:

   ```
   mgmt_net_macs="02:04:96:8b:cc:a8"
   ```

(Conditional) Configuring an HPE FlexFabric switch or an HPE FlexNetwork switch

Procedure

1. Access the switch through a console cable.
2. Log in with the default credentials.

   The username is `admin`, and the password is `admin`.

3. Enter the following commands:

   ```
   system-view
   interface Vlan-interface 1
   ip address dhcp-alloc
   quit
   local-user admin
   password simple admin
   service-type telnet
   authorization-attribute user-role network-admin
   quit
   telnet server enable
   line vty 0 63
   authentication-mode scheme
   user-role network-admin
   quit
   undo stp global enable
   save safely force
   ```

4. Enter the following command to retrieve the switch MAC address:

   ```
   display int vlan 1 | include hardware
   ```

   For example:

   ```
   display int vlan 1 | include hardware
   IP packet frame type: Ethernet II, hardware address: d894-03fe-07b1
   ```

Record the MAC address that this command returns. The cluster definition file requires you to specify the switch MAC address in a slightly different format. To specify the MAC address in this example in the cluster definition file, reformat the address as follows:

```
mgmt_net_macs="d8:94:03:fe:07:b1"
```

Running the `cm node add` command for a new switch

Procedure

1. Log into the admin node as the `root` user.
2. Edit the cluster definition file to include the new switch.

   Example 1. The following is a cluster definition file example for a spine switch:

   ```
   # Spine Switch
   internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="02:04:96:8b:cc:a8",
   redundant_mgmt_network=yes, net=head/head-bmc, type=spine, ice=yes
   ```

   Example 2. The following is a cluster definition file example for a leaf switch:

   ```
   # Leaf Switch
   internal_name=mgmtsw1, mgmt_net_name=head,
   mgmt_net_macs="d8:94:03:fe:07:b1", redundant_mgmt_network=yes, net=head/head-bmc, type=leaf, ice=yes
   ```

   NOTE: If the cluster does not have HPE SGI 8600 ICE hardware, set the `ice=` configuration attribute to `no`. Example: `ice=no`.

   It is possible that you cannot locate the cluster definition file. In this case, see the following topic for information about how to create a new one in a location of your choosing:

   Preparing to install the operating system and the cluster manager simultaneously on the admin node

   For more information and for cluster definition file examples, see the following:

   Cluster definition file examples with node templates, network interface card (NIC) templates, and predictable names

3. Run the `cm node add` command in the following format:

   To configure one management switch, use the following format:

   ```
   cm node add -c path_to_configfile
   ```

   For *path_to_configfile*, specify the full path to the cluster definition file.

   For example:

   ```
   # cm node add -c mgmtsw.config
   Config file: mgmtsw.config
   Add - All nodes in the mgmtsw.config will be added to the database.
   hog1: fastdiscover: Config file parse step: , 0.06s
   hog1: fastdiscover: new nodes step: , 0.09s
   hog1: fastdiscover: Script time: , 0.16s

   Refreshing the netboot environment for nodes in the config file...

   Updating admin node configs...
   Configuration manager initiating node configuration.
   0 of 1 nodes completed in 34.0 seconds, averaging 0.0s per node
   1 of 1 nodes completed in 34.8 seconds, averaging 34.1s per node
   Node configuration complete.

   Updating SU leader configs...
   Configuration manager initiating node configuration.
   ```

```
0 of 3 nodes completed in 7.1 seconds, averaging 0.0s per node
3 of 3 nodes completed in 7.6 seconds, averaging 6.8s per node
Node configuration complete.

Performing switch configuration...

Please view '/opt/clmgr/log/switchconfig.log' to verify no switch configuration error occurred during this process.
```

4. (Optional) Monitor the `cm node add` command progress.

   After the `cm node add` process is complete, it takes time for the switches to obtain an IP address via DHCP and become fully configured.

   To monitor the switch configuration progress, enter the following command:

   ```
   # tail -f /opt/clmgr/log/switchconfig.log
   ```

   After the `cm node add` command completes successfully, optionally continue to the next step.

5. (Optional) Enter the following command to verify that the firmware version matches this installation of the cluster manager:

   ```
   switchconfig sanity_check -s mgmtswX | grep firmware
   ```

   For *X*, specify the management switch number as it appears in the cluster definition file.

   Example 1: The following command is for an Extreme Networks switch:

   ```
   admin:~ # switchconfig sanity_check -s mgmtsw8 | grep firmware
   checking switch firmware on mgmtsw8 ...
   Switch installed in Slot-1 has firmware 16.2.5.4 installed (good)
   Switch installed in Slot-2 has firmware 16.2.5.4 installed (good)
   ```

   Example 2: The following command is for an HPE switch:

   ```
   admin:~ # switchconfig sanity_check -s mgmtsw6 | grep firmware
   checking switch firmware on mgmtsw6 ...
   mgmtsw6 slot 1 (5510 24G 4SFP+ HI 1-slot Switch) has firmware '7.1.070
   Release 1309P07-US' installed (recommended: '7.1.070 Release 1309P07' or
   '7.1.070 Release 1309P07-US')
   mgmtsw6 slot 2 (5510 24G 4SFP+ HI 1-slot Switch) has firmware '7.1.070
   Release 1309P07-US' installed (recommended: '7.1.070 Release 1309P07' or
   '7.1.070 Release 1309P07-US')
   ```

6. (Conditional) Upgrade or downgrade the firmware to match the recommended release.

   Complete this step if the switch firmware does not match the firmware version recommended by the cluster manager.

   The cluster manager has support for some brands of switches to simplify the upgrade or downgrade process. Enter the following command and view the help output for more information:

   ```
   # switchconfig update_firmware --help
   ```

   For example, assume that you need to update `mgmtsw0`, which is an HPE FlexFabric 5940 48SFP + 6QSFP + switch. You want to update the switch to firmware version 7.1.070 Release 2612P08-US. Enter the following command:

   ```
   # switchconfig update_firmware --switches mgmtsw0 --update --firmware file 5940-CMW710-R2612P08-US.ipe
   ```

   The `switchconfig` command in this example assumes that file `5940-CMW710-R2612P08-US.ipe` resides in the default TFTP directory, which is `/opt/clmgr/tftpboot/`. To guide you through the firmware upgrade or downgrade process, the command prompts you to answer a series of questions. Optionally, add the `--force` option to suppress the prompts and just upgrade the switch firmware.

   The cluster manager supports the `switchconfig update_firmware` command on the following switches:

- HPE FlexNetwork switches
- HPE FlexFabric switches
- HPE Aruba switches

Configuring a serial console

About this task

If the system console you typically use is a graphical console, you can configure a serial console to make remote maintenance easier.

Procedure

1. Use a text editor to open file `/etc/default/grub`.
2. On the `GRUB_CMDLINE_LINUX_DEFAULT` line, edit the line to include `console=`*`settings`* and remove `splash=silent quiet`.

   For *settings*, specify values that match your system settings. For example:

   ```
   GRUB_CMDLINE_LINUX_DEFAULT="console=ttyS0,115200n8
   intel_idle.max_cstate=1 processor.max_cstate=1 net.ifnames=1
   biosdevname=0 numa_balancing=disable"
   ```

3. On the `GRUB_TERMINAL` line, edit the line to change `gfxterm` to `console`.

   For example:

   ```
   GRUB_TERMINAL="console"
   ```

4. Enter the following command to apply the changes made in `/etc/default/grub` to the GRUB configuration file:

   ```
   # /usr/sbin/grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

   Later, if you want to remove the serial setup and switch back to the graphical console by default, complete the following steps:

   a. From the `GRUB_CMDLINE_LINUX_DEFAULT=` line, remove all the `console=` parameters.

   b. Change the `GRUB_TERMINAL` line back to `GRUB_TERMINAL="gfxterm"`

Using Aruba switches

The general procedure for connecting Aruba switches to a cluster includes the following steps:

- Remove the switch and cables from the packing box, power on, and connect a console cable to the Aruba switch.
- Configure basic settings. For example, configure an administrator account, ssh capabilities, and a static IP address.
- Assign an IP address to either the `mgmt` interface or the `vlan1` interface, and update the firmware to the correct revision.
    - The approved firmware for HPE Performance Cluster Manager 1.5 is 10.06.0010 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.5.1, which is HPE Performance Cluster Manager 1.5 with patches, is 10.07.0004 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.6 is 10.07.0010 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.7 is 10.08.1030 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.8 is 10.09.1040 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.9 is 10.10.1040 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
    - The approved firmware for HPE Performance Cluster Manager 1.10 is 10.11.1040 across all Aruba switches (6300, 6400, 8320, 8325, 8360).
- If applicable, configure either VSF (single control plane stacking on 6300 switches) or VSX (dual control plane stacking on 8320, 8325, 8360 switches).
- If VSX dual-spine switches are connected to the admin node, you configure the VSX LAG on both VSX spine switches going to the admin node.
- Obtain the MAC address on VLAN 1, and configure the switch on the cluster manager admin node.

Configuring basic settings on Aruba switches

About this task

The procedure in this topic explains how to configure the following:

- `ssh` capabilities through a console cable. This is required. This capability allows the cluster manager and users to connect to each switch remotely.
- Ports. Aruba 8320/8325/8360 switches require you to enable all ports. By default, all ports are disabled.
- An L3 IP address. You need to assign an L3 IP address to each switch on interface VLAN 1 in order for the cluster manager to reach the switch in-band.
- To facilitate firmware upgrades, you can assign an IP address to either interface `vlan1` or interface `mgmt`. This is optional.

Procedure

1. Connect to the console port of the switch, and open a CLI session.

   - For the username, specify `admin`.
   - For the password, leave it blank. You can press **Enter**.

   For information about console access, see the following:

   https://www.arubanetworks.com/techdocs/AOS-CX/10.10/HTML/fundamentals_6300-6400/Content/Chp_IniCfg/con-con-por.htm

2. Enter the following commands to configure basic connectivity:

   ```
   configure
   lldp
   user admin password plaintext admin
   ssh server vrf mgmt
   ssh server vrf default
   ```

   On Aruba 6300 switches, also enter the following command to enable the `ssh` server globally:

   ```
   ssh server
   ```

3. (Conditional) Enable ports.

   Complete this step on Aruba 8320, 8325, and 8360 switches.

   By default, all Aruba ports are shut down by default.

   Enter the following commands:

   ```
   configure
   interface 1/1/1-1/1/56
   no shutdown
   no routing
   ```

   NOTE: Your interface range might differ depending on the port count and switch model. Adjust the range accordingly. Enter the following command to display all the switch ports:

   `show interface brief`

4. Assign an IP address to the switch.

   Choose an IP address that matches the cluster manager discovery line.

   Follow the steps in one of the following scenarios:

   Scenario 1. In this scenario, the regular switchports, that is the *data plane*, connect to the admin node. This scenario assumes that you want to connect the switch to the `head` network on the admin node with the default subnet of `172.23.0.0/16`.

   The commands are as follows on switch `sw-spine01`:

   ```
   configure       # Spine Switch 8325 Pair in VSX (sw-spine01/sw-spine02)
   hostname sw-spine01
   interface vlan1
   ip address 172.23.255.252/16
   no shutdown
   no ip dhcp      # on 6300 switches only (8320/8325/8360 switches do not support DHCP on VLAN 1)
   ```

The commands are as follows on switch `sw-spine02`:

```
configure
hostname sw-spine02
interface vlan1
ip address 172.23.255.253/16
no shutdown
no ip dhcp      # on 6300 switches only (8320/8325/8360 switches do not support DHCP on VLAN 1)
```

Scenario 2. In this scenario, only the `mgmt` RJ45 port connects to the admin node. This scenario assumes that you want to use the `mgmt` RJ45 port, and you created your own `mgmt` subnet of `10.10.0.0/16` off of `eno4`on an admin node with an IP address of `10.10.0.1/16`. The commands are as follows:

The commands are as follows on switch `sw-spine01`:

```
configure        # Spine Switch 8325 Pair in VSX (sw-spine01/sw-spine02)
hostname sw-spine01
interface mgmt
ip static 10.10.0.100/16
```

The commands are as follows on switch `sw-spine02`:

```
configure
hostname sw-spine02
interface mgmt
ip static 10.10.0.101/16
no shutdown
```

Aruba firmware levels

The cluster manager interoperates with the following Aruba firmware versions:

HPE Performance Cluster Manager release version	Aruba firmware version
1.5	10.06.0010
1.5 plus patches	10.07.0004
1.6	10.07.0010
1.7	10.08.1030
1.8	10.09.1040
1.9	10.10.1040
1.10	10.11.1021

Obtain firmware from the following Aruba support portal:

https://asp.arubanetworks.com/downloads

You can log into the Aruba support portal with your HPE login. On the portal, search for the string that matches the firmware version. For example, search for `10.11.1021`.

HPE Performance Cluster Manager 1.5 firmware file names

The HPE Performance Cluster Manager 1.5 release supports firmware version 10.06.0010. The Aruba 8360 switch supports only firmware version 10.06.*XXXX*. The file names are as follows:

- Aruba 6300/6400 firmware file name: `ArubaOS-CX_6400-6300_10_06_0010.swi`
- Aruba 8320 firmware file name: `ArubaOS-CX_8320_10_06_0010.swi`
- Aruba 8325 firmware file name: `ArubaOS-CX_8325_10_06_0010.swi`
- Aruba 8360 firmware file name: `ArubaOS-CX_8360_10_06_0010.swi`

HPE Performance Cluster Manager 1.6 firmware file names

The HPE Performance Cluster Manager 1.6 release supports firmware version 10.07.0010. The file names are as follows:

- Aruba 6300/6400 firmware file name: `ArubaOS-CX_6400-6300_10_07_0010.swi`
- Aruba 8320 firmware file name: `ArubaOS-CX_8320_10_07_0010.swi`
- Aruba 8325 firmware file name: `ArubaOS-CX_8325_10_07_0010.swi`
- Aruba 8360 firmware file name: `ArubaOS-CX_8360_10_07_0010.swi`

HPE Performance Cluster Manager 1.7 firmware file names

The HPE Performance Cluster Manager 1.7 release supports firmware version 10.08.1030. The file names are as follows:

- Aruba 6300/6400 firmware file name: `ArubaOS-CX_6400-6300_10_08_1030.swi`
- Aruba 8320 firmware file name: `ArubaOS-CX_8320_10_08_1030.swi`
- Aruba 8325 firmware file name: `ArubaOS-CX_8325_10_08_1030.swi`
- Aruba 8360 firmware file name: `ArubaOS-CX_8360_10_08_1030.swi`

HPE Performance Cluster Manager 1.8 firmware file names

The HPE Performance Cluster Manager 1.8 release supports firmware version 10.09.1040. The file names are as follows:

- Aruba 6300/6400 firmware file name: `ArubaOS-CX_6400-6300_10_09_1040.swi`
- Aruba 8320 firmware file name: `ArubaOS-CX_8320_10_09_1040.swi`
- Aruba 8325 firmware file name: `ArubaOS-CX_8325_10_09_1040.swi`
- Aruba 8360 firmware file name: `ArubaOS-CX_8360_10_09_1040.swi`

HPE Performance Cluster Manager 1.9 firmware file names

The HPE Performance Cluster Manager 1.9 release supports firmware version 10.10.1040. The file names are as follows:

- Aruba 6300/6400 firmware file name: `ArubaOS-CX_6400-6300_10_10_1040.swi`
- Aruba 8320 firmware file name: `ArubaOS-CX_8320_10_10_1040.swi`
- Aruba 8325 firmware file name: `ArubaOS-CX_8325_10_10_1040.swi`
- Aruba 8360 firmware file name: `ArubaOS-CX_8360_10_10_1040.swi`

HPE Performance Cluster Manager 1.10 firmware file names

The HPE Performance Cluster Manager 1.10 release supports firmware version 10.11.1021. The file names are as follows:

- Aruba 6300/6400 Firmware File Name: `ArubaOS-CX_6400-6300_10_11_1021.swi`
- Aruba 8320 Firmware File Name: `ArubaOS-CX_8320_10_11_1021.swi`
- Aruba 8325 Firmware File Name: `ArubaOS-CX_8325_10_11_1021.swi`
- Aruba 8360 Firmware File Name: `ArubaOS-CX_8360_10_11_1021.swi`

Upgrading Aruba switch firmware manually

Procedure

1. Open a console to the switch, and enter the following command:

   ```
   show image | include Active
   Management Module 1/1 (Active)
   Active Image       : primary
   ```

   The preceding output shows that the switch has the `primary` image booted. Subsequent steps in this procedure download the new firmware to `secondary`.

2. Log into the admin node as the root user, and enter the following command to ensure the correct location on the admin node:

   ```
   # ls -l ArubaOS-CX_*
   -rwxr-xr-x 1 root root 642950360 Nov 20 16:05 ArubaOS-CX_6400-6300_10_06_0010.swi
   -rwxr-xr-x 1 root root 385303399 Nov 20 16:05 ArubaOS-CX_8320_10_06_0010.swi
   -rwxr-xr-x 1 root root 425134317 Nov 20 16:05 ArubaOS-CX_8325_10_06_0010.swi      # This is the Aruba 8325 Firmware
   ```

3. Upload the firmware to the Aruba switch.

   Use the `tftp` command with one of the following parameters:

   - If you configured the switch with `interface vlan1`, also specify `vrf default`.
   - If you configured the switch with `interface mgmt`, also specify `vrf mgmt`.

   Example 1. The following example specifies `interface vlan1` on admin node `bond0 = 172.23.0.1`:

   ```
   8325# copy tftp://172.23.0.1/ArubaOS-CX_8325_10_10_1040.swi secondary vrf default
   The secondary image will be deleted.

   Continue (y/n)? y
     % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                    Dload  Upload   Total   Spent    Left  Speed
     0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
     0  405M    0 3060k    0     0   738k      0  0:09:22  0:00:04  0:09:18  740k
   ```

   Example 2. The following example specifies `interface mgmt` on admin node `eno4 = 10.10.0.1`:

   ```
   8325# copy tftp://10.10.0.1/ArubaOS-CX_8325_10_10_1040.swi secondary vrf mgmt
   The secondary image will be deleted.

   Continue (y/n)? y
     % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                    Dload  Upload   Total   Spent    Left  Speed
     0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
   100  405M  100  405M    0     0  11.9M      0  0:00:33  0:00:33 --:--:-- 11.9M
   100  405M  100  405M    0     0  11.9M      0  0:00:33  0:00:33 --:--:-- 11.9M
   ```

```
Verifying and writing system firmware...
8325#
```

4. After the download completes, set the default boot to `secondary`:

```
8325# boot set-default secondary
Default boot image set to secondary.
```

5. Enter the following command to boot to the secondary partition, and enter **y** when prompted:

```
8325# boot system

Do you want to save the current configuration (y/n)? y
The running configuration was saved to the startup configuration.

Checking for updates needed to programmable devices...
```

It can take 3 to 7 minutes for the switch to reboot.

6. After the reboot completes, verify the firmware version:

```
8325# show version
--------------------------------------------------------------------------------
ArubaOS-CX
(c) Copyright 2017-2023 Hewlett Packard Enterprise Development LP
--------------------------------------------------------------------------------
Version      : GL.10.10.1040
Build Date   : 2023-02-14 04:44:29 UTC
Build ID     : ArubaOS-CX:GL.10.10.1040:8e70e02edd94:202302140343
Build SHA    : 8e70e02edd94009b1b2f593ff7a9836fa28ed244
Active Image : primary

Service OS Version : GL.01.08.0003
BIOS Version       : GL-01-0013
```

Upgrading Aruba switch firmware using the cluster manager `switchconfig` command

Procedure

1. Log into the admin node as the root user.
2. Download the firmware to a location on the cluster admin node.

   For example, download it to `/opt/clmgr/tftpboot`.

3. Verify the software download:

```
# ls /opt/clmgr/tftpboot/ArubaOS-CX_8325*
/opt/clmgr/tftpboot/ArubaOS-CX_8325_10_10_1040.swi
```

4. Use the `switchconfig` command to update the firmware.

   For example:

```
switchconfig update_firmware --switches sw-spine01 --update --firmware-file ArubaOS-CX_8325_10_10_1040.swi
```

   Follow the prompts to save, reboot, or complete other tasks. To avoid prompts, also specify the `--force` parameter, which performs the upgrade and reboots. The entire process can take from 10-15 minutes.

Configuring Aruba VSF

About this task

You can configure VSF only on Aruba 6200/6300 switches. VSF requires an interswitch link. VSF is also known as *single control pane stacking*.

The example in this topic assumes two Aruba 6300 switches with the following cabling:

- Aruba6300-1 1/1/25 <---> Aruba6300-2 1/1/26
- Aruba6300-1 1/1/26 <---> Aruba6300-2 1/1/25

Procedure

1. Console into **each** Aruba switch and enter the following commands:

   ```
   configure
   vsf member 1
   link 1 1/1/25
   link 2 1/1/26
   exit
   ```

2. On the bottom Aruba switch, enter the following command to set it to member 2:

   ```
   vsf renumber-to 2
   ```

3. Enter `yes` to save and reboot.

4. Wait for the reboot to complete.

5. Verify the switch roles.

   On the top (conductor) switch, enter the commands shown in the following examples:

   ```
   mgmtsw3# show vsf member 1 | include Status
           Status                          : Conductor
   mgmtsw3# show vsf member 2 | include Status
           Status                          : Member
   ```

   The output shows that `member 1` is the conductor switch, and `member 2` is a member switch.

6. Enter the following commands to enable a member switch to assume the role of a conductor switch if the current conductor switch fails:

   ```
   configure
   vsf secondary-member 2
   This will save the configuration and reboot the specified switch.
   Do you want to continue (y/n)? y
   ```

   Remember that Aruba switches are single-control-plane switches. For such switches, when there are two or more physical switches, the cluster manages them both as if they were a single, logical, management switch. For example `mgmtsw3`. When you configure these switches into the cluster, configure them as `type=leaf` in the cluster definition file.

7. (Conditional) Plug in the VSF ISL cables.

   Complete this step if the cables are currently not plugged in.

Configuring Aruba VSX

About this task

You can configure VSX only on Aruba 6400/8320/8325/8360 switches. A VSX requires an inter-switch link (ISL). The maximum is four (4) links.

Procedure

1. Verify that all inter-switch links (ISLs) are in a link aggregation group (LAG).

   When VSX is enabled, use the methods in this appendix section to create LAGs. LAGs are also called channel groups, port channels, or bonds. The following are the two types of LAGs:

   - LACP (802.3ad). Used to connect to end nodes, such as some service nodes or the admin node.
   - Static. Used for interswitch links, chassis management modules (CMMs), and chassis management controllers (CMCs).

2. Update the firmware on all switches.
3. Verify that the same fimware revision is installed on all the switches.
4. Verify that the VSX LAG ID matches the lowest port number in the LAG.

   For example, this means that interfaces (1/1/47 and 1/1/48) use LAG 47.

5. Verify that each VSX pair on the cluster uses a different system MAC address.

   Example dual-spine syntax. The first dual-spine VSX pair uses MAC `02:02:00:00:01:00`.

   Example dual leaf syntax. The MAC addresses are as follows:

   - The first dual-leaf VSX pair uses MAC `02:00:00:00:01:00`.
   - The second dual-leaf VSX pair uses MAC `02:00:00:00:02:00`.
   - The third dual-leaf VSX pair uses MAC `02:00:00:00:03:00`.

Configuring Aruba VSX dual spine switches

Prerequisites
Configuring Aruba VSX

About this task

The procedure in this topic assumes two Aruba 8325 switches connected in the following way:

- `sw-spine01` 1/1/47 <--> `sw-spine02` 1/1/47
- `sw-spine01` 1/1/48 <--> `sw-spine02` 1/1/48

Procedure

1. Log into `sw-spine01`, and enter the following commands:

```
configure
interface lag 47
no shutdown
no routing
vlan trunk allowed all
exit
int 1/1/47-1/1/48
lag 47
no shutdown
exit
vsx
system-mac 02:02:00:00:01:00
inter-switch-link lag 47
role primary
```

2. Log into `sw-spine02`, and enter the following commands:

```
configure
interface lag 47
no shutdown
no routing
vlan trunk allowed all
exit
int 1/1/47-1/1/48
lag 47
no shutdown
exit
vsx
system-mac 02:02:00:00:01:00
inter-switch-link lag 47
role secondary
```

Configuring Aruba VSX dual leaf switches

Prerequisites

Configuring Aruba VSX

About this task

The procedure in this topic assumes two Aruba dual leaf switches connected in the following way:

- `d100sw1` 1/1/49 <--> `d100sw2` 1/1/49
- `d100sw1` 1/1/50 <--> `d100sw2` 1/1/50

Procedure

1. Log into `d100sw1`, and enter the following commands:

```
configure
interface lag 49
```

```
no shutdown
no routing
vlan trunk allowed all
exit
int 1/1/49-1/1/50
lag 49
no shutdown
exit
vsx
system-mac 02:00:00:00:01:00
inter-switch-link lag 49
role primary
```

2. Log into `d100sw2`, and enter the following commands:

```
configure
interface lag 49
no shutdown
no routing
vlan trunk allowed all
exit
int 1/1/49-1/1/50
lag 49
no shutdown
exit
vsx
system-mac 02:00:00:00:01:00
inter-switch-link lag 49
role secondary
```

Configuring Aruba VSX keep alive

Prerequisites

Configuring Aruba VSX

Procedure

1. Determine a /30 subnet for the point-to-point link.

   If there are many pairs of VSX switches with keep-alive links, choose an unused /24 subnet and split it into many /30 subnets. For example:

```
Base Subnet = 10.1.0.0/24
VSX Keepalive #1 = 10.1.0.0/30 = VSX switch #1 (10.1.0.1/30) | VSX switch #2 (10.1.0.2/30)
VSX Keepalive #2 = 10.1.4.0/30 = VSX switch #1 (10.1.0.5/30) | VSX switch #2 (10.1.0.6/30)
VSX Keepalive #3 = 10.1.8.0/30 = VSX switch #1 (10.1.0.9/30) | VSX switch #2 (10.1.0.10/30)
VSX Keepalive #4 = 10.1.12.0/30 = VSX switch #1 (10.1.0.13/30) | VSX switch #2 (10.1.0.14/30)
...
```

   The examples in the rest of this procedure assume VSX switch #1 <-> VSX switch #2 connected for the keep-alive as port 1/1/41<->1/1/41.

2. Enter commands to configure the VSX keep alive on VSX switch 1.

   For example:

```
configure
interface 1/1/41
```

```
routing
no shutdown
ip address 10.1.0.1/30
vsx
keepalive peer 10.1.0.2 source 10.1.0.1
end
```

3. Enter commands to configure the VSX keep alive on VSX switch 2.

   For example:

```
configure
interface 1/1/41
routing
no shutdown
ip address 10.1.0.2/30
vsx
keepalive peer 10.1.0.1 source 10.1.0.2
end
```

Using `switchconfig` to configure an Aruba VSX spine pair and a VSX leaf pair in a link aggregate group (LAG)

About this task

In this example, assume that there are four links that connect four physical switches. That is, there are two pairs of VSX partner switches.

NOTE: If more than one link connects the switches before you configure the link aggregation group (LAG), the result is a network loop, which can cause the network to go down. Take one of the following precautions:

- Physically unplug the redundant cables.

  Or

- While configuring the LAGs, issue a `shutdown` command at one end of the links.

Assume the following configuration:

- `sw-spine01` 1/1/51 <-> `sw-leaf01` 1/1/51
- `sw-spine01` 1/1/52 <-> `sw-leaf02` 1/1/51
- `sw-spine02` 1/1/51 <-> `sw-leaf01` 1/1/52
- `sw-spine02` 1/1/52 <-> `sw-leaf02` 1/1/52

Procedure

1. Log into the admin node as the root user.
2. Enter the following commands to configure a 4-port VSX LAG on each pair:

```
# switchconfig set -s sw-spine01,sw-spine02 --default-vlan 1 --bonding manual --ports 1/1/51,1/1/52 --mlag
# switchconfig set -s sw-leaf01,sw-leaf02 --default-vlan 1 --bonding manual --ports 1/1/51,1/1/52 --mlag
```

Using switch commands to configure an Aruba VSX spine pair and a VSX leaf pair in a link aggregate group (LAG)

About this task

Complete the procedure in this topic if you are unable to complete the automated procedure in the following topic:

Using `switchconfig` to configure an Aruba VSX spine pair and a VSX leaf pair in a link aggregate group (LAG)

For information about assumptions and the example switch scenario, see the preceding link.

Procedure

1. Log into each of the switches.
2. Enter the following commands on each switch:

```
interface lag 51 multi-chassis static
    no shutdown
    no routing
    vlan trunk native 1
    vlan trunk allowed 1
    exit
interface 1/1/51
    no shutdown
    mtu 9198
    lag 51
    exit
interface 1/1/52
    no shutdown
    mtu 9198
    lag 51
    exit
```

Using `switchconfig` commands to configure an Aruba VSX spine pair and a VSX leaf pair in a link aggregate (LAG) group

About this task

In this example, assume that there are four links interconnecting four physical switches. There are two VSX switches (dual-control plane) to one VSF logical switch, which consists of two physical switches with a single control plane.

NOTE: If more than one link connects the switches together before the LAGs are configured, the result is a network loop, which can cause the network to go down. Take one of the following precautions:

- Physically unplug the redundant cables.

  Or

- While configuring the LAGs, from the switch command line at one end of the links, issue a `shutdown` command.

Assume the following configuration:

- `sw-spine01 1/1/51 <-> sw-leaf01 1/1/51`
- `sw-spine01 1/1/52 <-> sw-leaf01 1/1/51`

- `sw-spine02 1/1/51 <-> sw-leaf01 2/1/52`
- `sw-spine02 1/1/52 <-> sw-leaf01 2/1/52`

Procedure

1. Log into the admin node as the root user.
2. Enter the following command to disable 3 / 4 of the links, which prevents a networking loop and leaves `sw-spine01 1/1/51` up:

   ```
   # switchconfig port -s sw-spine02 --disable --ports 1/1/51,1/1/52
   # switchconfig port -s sw-spine01 --disable --ports 1/1/52
   ```

3. Enter the following commands to configure a 4-port LAG on the two VSX switches and the singe VSF switch:

   ```
   # switchconfig set -s sw-spine01,sw-spine02 --default-vlan 1 --bonding manual --ports 1/1/51,1/1/52 --mlag
   # switchconfig set -s sw-leaf01 --default-vlan 1 --bonding manual --ports 1/1/51,1/1/52,2/1/51,2/1/52
   ```

NOTE: If the `switchconfig` command is unavailable, complete the following sequence:

Enter the following commands on the two VSX switches:

```
interface lag 51 multi-chassis static
    no shutdown
    no routing
    vlan trunk native 1
    vlan trunk allowed 1
    exit
interface 1/1/51
    no shutdown
    mtu 9198
    lag 51
    exit
interface 1/1/52
    no shutdown
    mtu 9198
    lag 51
    exit
```

Enter the following commands on the single VSF switch:

```
interface lag 51
    no shutdown
    no routing
    vlan trunk native 1
    vlan trunk allowed 1
    exit
interface 1/1/51
    no shutdown
    mtu 9198
    lag 51
    exit
interface 1/1/52
    no shutdown
    mtu 9198
    lag 51
    exit
interface 2/1/51
    no shutdown
    mtu 9198
    lag 51
    exit
interface 2/1/52
    no shutdown
    mtu 9198
    lag 51
    exit
```

Configuring an Aruba 8325 VSX spine pair and the admin node as a link aggregate group (LAG)

About this task

The example in this procedure assumes the following configuration:

- `sw-spine01` 1/1/1 <-----> admin node `ens1f0`
- `sw-spine02` 1/1/1 <-----> admin node `ens1f1`
- The following information pertains to admin node `bond0`:
  - The bonding mode is IEEE 802.3ad dynamic link aggregation.
  - `ens1f0` and `ens1f1` are bonding subsidiary interfaces.

Procedure

1. Log into the admin node as the root user.
2. Make sure that VSX is enabled on the two Aruba 8325 switches.
3. Log into each switch, and use switch commands to configure the LAG.

   For example, log into switch `sw-spine01` and switch `sw-spine02`, and enter the following commands on each switch:

```
interface lag 1 multi-chassis
    no shutdown
    no routing
    vlan access 1
    lacp mode active
    lacp fallback
interface 1/1/1
    no shutdown
    mtu 9198
    lag 1
```

Adding Aruba switches to a cluster

About this task

You can use the `mgmtsw_isls=` configuration attribute to configure all ISL LAGs when the switches are configured into the cluster. When a LAG is identified with this configuration attribute, the LAG is configured into the cluster in one of the following ways:

- When you run the `cm node add` command and specify to configure the switch configuration file into the cluster. For example:

  ```
  # cm node add -c mgmtsw.config
  ```

- When you run a `switchconfig_configure_node` command in the following format:

  `switchconfig_configure_node --node` *`switch hostnames`*

  Typically, it is not necessary to run the `switchconfig_configure_node` command because the `cm node add` command calls the `switchconfig_configure_node` command. Use this command only if one of the following are true:

- When the `discover_skip_switchconfig=yes` parameter is specified in the cluster definition file for that particular node or switch.
- The `cadmin --enable-discover-skip-switchconfig` attribute is enabled globally.
- When you specify `--skip-switch-config` on the `cm node add` command line.

Procedure

1. Log into the admin node as the root user.
2. Create a cluster definition file, in the form of text file, that contains the configuration attributes for the new switch.

   Example 1. Assume that the switches are cabled as follows:

   - `sw-spine01 1/1/55 <----> sw-leaf01 1/1/51`
   - `sw-spine02 1/1/55 <----> sw-leaf01 1/1/52`

   The cluster definition file lines to define these switches are as follows:

   ```
   internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:33", redundant_mgmt_network=yes,
   net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine01, mgmtsw_partner=sw-spine02,
   mgmt_net_ip=172.23.255.252, mgmtsw_isls="1/1/55"

   internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:34", redundant_mgmt_network=yes,
   net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine02, mgmtsw_partner=sw-spine01,
   mgmt_net_ip=172.23.255.253, mgmtsw_isls="1/1/55"

   internal_name=mgmtsw2, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
   net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf01, mgmt_net_ip=172.23.255.247,
   mgmtsw_isls="1/1/51,1/1/52"
   ```

   Example 2. Assume that the switches are cabled as follows:

   - `sw-spine01 1/1/51 <---> sw-leaf02 1/1/51`
   - `sw-spine01 1/1/52 <---> sw-leaf02 2/1/51`
   - `sw-spine02 1/1/51 <---> sw-leaf02 1/1/52`
   - `sw-spine02 1/1/52 <---> sw-leaf02 2/1/52`
   - `sw-spine01 1/1/53 <---> sw-leaf03 1/1/51`
   - `sw-spine01 1/1/54 <---> sw-leaf03 2/1/51`
   - `sw-spine02 1/1/53 <---> sw-leaf03 1/1/52`
   - `sw-spine02 1/1/54 <---> sw-leaf03 2/1/52`

   The cluster definition file lines to define these switches are as follows:

   ```
   # NOTE the careful use of "," vs ";" in the mgmtsw_isls= parameter and compare it
   # to the cabling
   # the VSX switches here will create 2 distinct LAGs - "LAG 51 that includes 1/1/51 + 1/1/52"
   # and "LAG 53 that includes 1/1/53 + 1/1/54"
   internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:33", redundant_mgmt_network=yes,
   net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine01,
   mgmtsw_partner=sw-spine02, mgmt_net_ip=172.23.255.252, mgmtsw_isls="1/1/51,1/1/52;1/1/53,1/1/54"

   internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:34", redundant_mgmt_network=yes,
   net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine02,
   mgmtsw_partner=sw-spine01, mgmt_net_ip=172.23.255.253, mgmtsw_isls="1/1/51,1/1/52;1/1/53,1/1/54"
   ```

```
# Notice in the VSF switch entry - there is only commas, so it will create a single LAG -
# "LAG 51 will include 1/1/51 + 1/1/52 + 2/1/51 + 2/1/52"
internal_name=mgmtsw3, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf02, mgmt_net_ip=172.23.255.248,
mgmtsw_isls="1/1/51,1/1/52,2/1/51,2/1/52"

internal_name=mgmtsw4, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:11:22", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf03, mgmt_net_ip=172.23.255.249,
mgmtsw_isls="1/1/51,1/1/52,2/1/51,2/1/52"
```

Example 3. Assume that the switches are cabled as follows:

- `sw-spine01 1/1/49 <---> d100sw1 1/1/51`
- `sw-spine01 1/1/50 <---> d100sw2 1/1/51`
- `sw-spine02 1/1/49 <---> d100sw1 1/1/52`
- `sw-spine02 1/1/50 <---> d100sw2 1/1/52`

The cluster definition file lines to define these switches are as follows:

```
internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:33", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine01, mgmtsw_partner=sw-spine02,
mgmt_net_ip=172.23.255.252, mgmtsw_isls="1/1/49,1/1/50"

internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:34", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine02, mgmtsw_partner=sw-spine01,
mgmt_net_ip=172.23.255.253, mgmtsw_isls="1/1/49,1/1/50"

internal_name=mgmtsw5, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, hostname1=d100sw1, mgmtsw_partner=d100sw2,
mgmt_net_ip=172.23.255.250, mgmtsw_isls="1/1/51,1/1/52"

internal_name=mgmtsw6, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:11:22", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, hostname1=d100sw2, mgmtsw_partner=d100sw1,
mgmt_net_ip=172.23.255.251, mgmtsw_isls="1/1/51,1/1/52"
```

3. Combine the cluster definition lines from the preceding step into one file called `mgmtsw.config`.

The following lines appear in the combined cluster definition file:

```
[discover]
internal_name=mgmtsw0, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:33", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine01, mgmtsw_partner=sw-spine02,
mgmt_net_ip=172.23.255.252, mgmtsw_isls="1/1/49,1/1/50;1/1/51,1/1/52;1/1/53,1/1/54;1/1/55"

internal_name=mgmtsw1, mgmt_net_name=head, mgmt_net_macs="b8:d4:e7:11:22:34", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-spine, hostname1=sw-spine02, mgmtsw_partner=sw-spine01,
mgmt_net_ip=172.23.255.253, mgmtsw_isls="1/1/49,1/1/50;1/1/51,1/1/52;1/1/53,1/1/54;1/1/55"

internal_name=mgmtsw2, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf01, mgmt_net_ip=172.23.255.247,
mgmtsw_isls="1/1/51,1/1/52"

internal_name=mgmtsw3, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf02, mgmt_net_ip=172.23.255.248,
mgmtsw_isls="1/1/51,1/1/52,2/1/51,2/1/52"

internal_name=mgmtsw4, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:11:22", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=leaf, hostname1=sw-leaf03, mgmt_net_ip=172.23.255.249,
mgmtsw_isls="1/1/51,1/1/52,2/1/51,2/1/52"

internal_name=mgmtsw5, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:bb:cc", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, hostname1=d100sw1, mgmtsw_partner=d100sw2,
mgmt_net_ip=172.23.255.250, mgmtsw_isls="1/1/51,1/1/52"

internal_name=mgmtsw6, mgmt_net_name=head, mgmt_net_macs="88:3a:30:aa:11:22", redundant_mgmt_network=yes,
net=head/head-bmc, ice=no, type=dual-leaf, hostname1=d100sw2, mgmtsw_partner=d100sw1,
mgmt_net_ip=172.23.255.251, mgmtsw_isls="1/1/51,1/1/52"
```

4. Enter the following command to configure the switches into the cluster:

   ```
   # cm node add -c mgmtsw.config
   ```

5. Enter the following command to watch the configuration progress:

   ```
   # tail -f /opt/clmgr/log/switchconfig.log
   ```

Imaging nodes with UDP Multicast (UDPcast)

The cluster manager supports the following image transport methods for node imaging:

- BitTorrent (default)
- `rsync`
- UDPcast

Hewlett Packard Enterprise recommends that you use either the default method of BitTorrent or `rsync`. The cluster manager supports UDPcast in the cluster definition file and on the `cm node provision` command for legacy circumstances.

Managing UDP multicast (UDPcast) provisioning

UDPcast allows hundreds of nodes to join a multicast stream of the files being transported. With all the nodes sharing a single stream, the network is protected from being saturated by disjoint installations.

UDPcast overview

By default, BitTorrent is the tool used for multicast image provisioning operations.

As an alternative, you can specify `--transport udpcast` to specify UDPcast on the `cm node provision` command. UDPcast 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.

NOTE: Hewlett Packard Enterprise recommends that you do not change UDPcast settings unless advised to do so by a technical support representative.

You can tune the following configuration attribute settings:

- `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.

- `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.

UDPcast configuration attributes

admin_udpcast_mcast_rdv_addr

When UDPcast is used, this attribute specifies the UDPcast rendezvous (RDV) address at which admin node senders and non-admin receiver nodes find each other.

The receiver nodes are compute nodes.

Default = `239.255.255.1`.

Values = any valid IPv4 multicast address in the 224.0.0.0/4 range .

Accepted by:

- Cluster configuration tool
- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command

To set a node-specific value for this attribute, use the `cm node set` command in the following format:

```
cm node set --udpcast-mcast-rdv-addr value -n node
```

For *value*, specify the node-specific value.

For *node*, specify the node hostname.

If you specify a nondefault IP address, also use the `cm node provision` command to push an image and initiate changes on the receiver nodes.

- `cm node show` command

  To display the cluster-wide setting for this attribute, enter the following command:

  ```
  # cm node show --udpcast-mcast-rdv-addr -n admin
  ```

  To display a node-specific setting for this attribute, enter the `cm node show` command in the following format:

  ```
  cm node show --udpcast-mcast-rdv-addr -n node
  ```

  For *node*, specify the node hostname.

edns_udp_size

Specifies the `edns-udp-size` option in `/etc/named.conf`. This value is the default packet size, in bytes, that remote servers can receive.

Default = `512`.

Values = any positive integer number.

Accepted by:

`cattr` command

udpcast_max_bitrate

Specifies the maximum numbers of bits that are conveyed or processed per second. This attribute is expressed as a number followed by a unit of measure, such as `m`.

Default = `900m`.

Values = any positive integer number followed by a unit of measure. The default unit of measure is `m` (megabytes). For the list of units of measure, see the `udp-sender`(1) manpage.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

udpcast_max_wait

Specifies the greatest amount of time that can elapse between when the first client node connects and any other client nodes connect. Clients that connect after this time has elapsed receive their software in a subsequent broadcast.

Default = `10`.

Values = any positive integer number.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

udpcast_min_receivers

Specifies the minimum number of receiver nodes for UDPcast.

Default = `1`.

Values = any positive integer number.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

udpcast_min_wait

Specifies the minimum amount of time that the system waits, while allowing clients to connect, before the software broadcast begins. This specification is the time between when the first client node connects and any other client nodes connect. The UDPcast distributes the software to all clients that connect during this interval.

Default = `10`.

Values = any positive integer number.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

udpcast_rexmit_hello_interval

Specifies the frequency with which the UDP sender transmits `hello` packets.

For the admin node, the default is `5000` (5 seconds).

Values = any positive integer number. When set to `0`, this attribute is disabled.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command

The `--rexmit-hello-interval` setting is especially important when the rendezvous (RDV) address is not 224.0.0.1. 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 that the Ethernet switch detects. The Ethernet switch then updates its tables with this information, thus allowing the multicast packets to properly route through the switch. The problem is that sometimes the UDP receiver sends its connection packet before the switch has had a chance to update the switch routing. If the request packet is not detected by the UDP sender on the admin node, the UDP receiver could wait forever for a UDPcast stream. For example, the sender might not detect the packet because the packet was sent before the switch was set up to pass the packet.

The `udpcast_rexmit_hello_interval` value configures the UDP sender to send a `HELLO` packet at regular intervals and configures UDP receivers to respond to the packet. This way, even if the UDP receiver request is missed, the UDP receiver sends a fresh request after seeing a `HELLO` packet from the UDP sender.

By default, the cluster manager sets the `udpcast_rexmit_hello_interval` value to 5000 (5 seconds) for UDP senders running on the admin node. Enter the following command to display this value:

```
# cm node show --udpcast-rexmit-hello-interval --global
```

The admin node uses the global value when it serves the compute nodes using the UDPcast transport mechanism.

For more information, see the information about the `--rexmit-hello-interval` on the `udp-sender` manpage.

udpcast_ttl

Sets the UDPcast time to live (TTL), which specifies the number of VLAN boundaries a request can cross.

For the admin node, the default is `2`. The admin nodes serve the compute nodes.

When `udpcast_ttl=1`, the request cannot cross a VLAN boundary. When `udpcast_ttl=2`, the request can cross one VLAN boundary.

If your site has routed management networks, a data transmission might have to cross from one VLAN to another.

If your site has no routed management networks, or if your site policy requires, you can set `udpcast_ttl=1` for the admin node.

Values = any positive integer number.

Accepted by:

- Cluster definition file
- `cattr` command
- `cm node add` command
- `cm node set` command
- `cm node show` command