- 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.

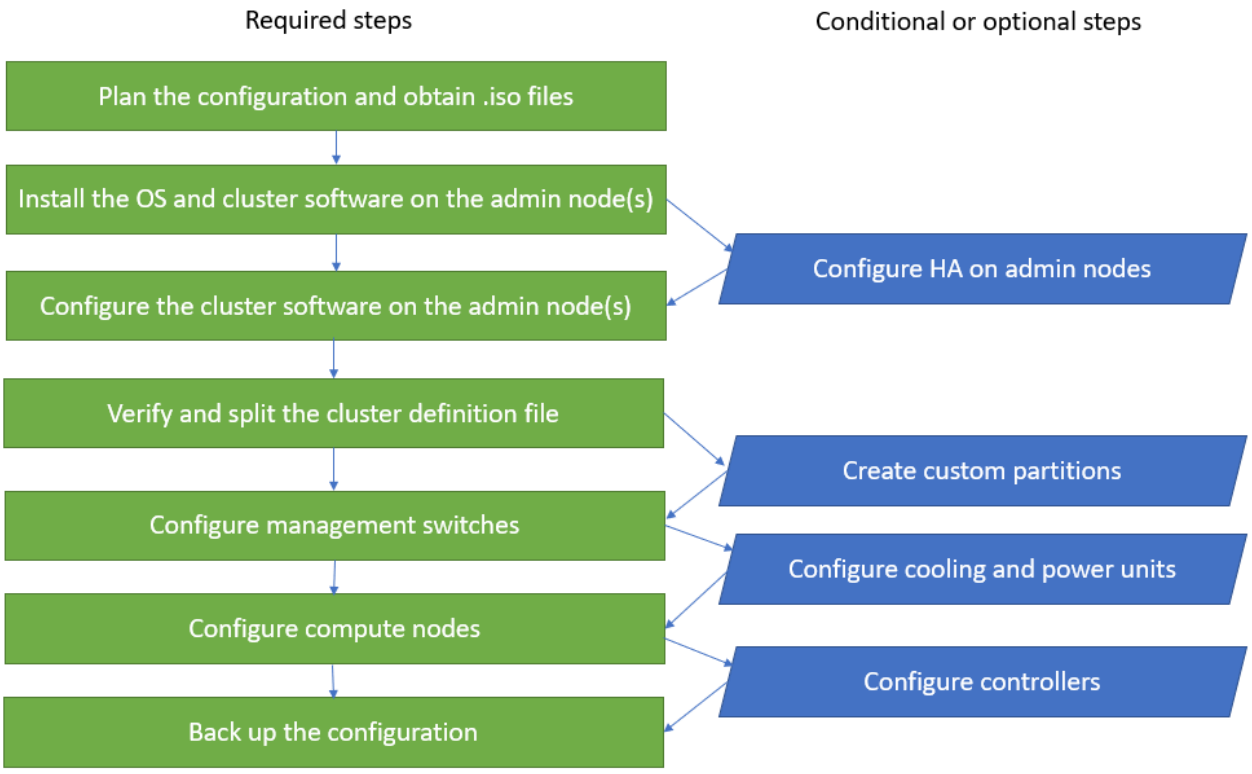

Figure 1: Installation process flow for clusters without leader nodes

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.

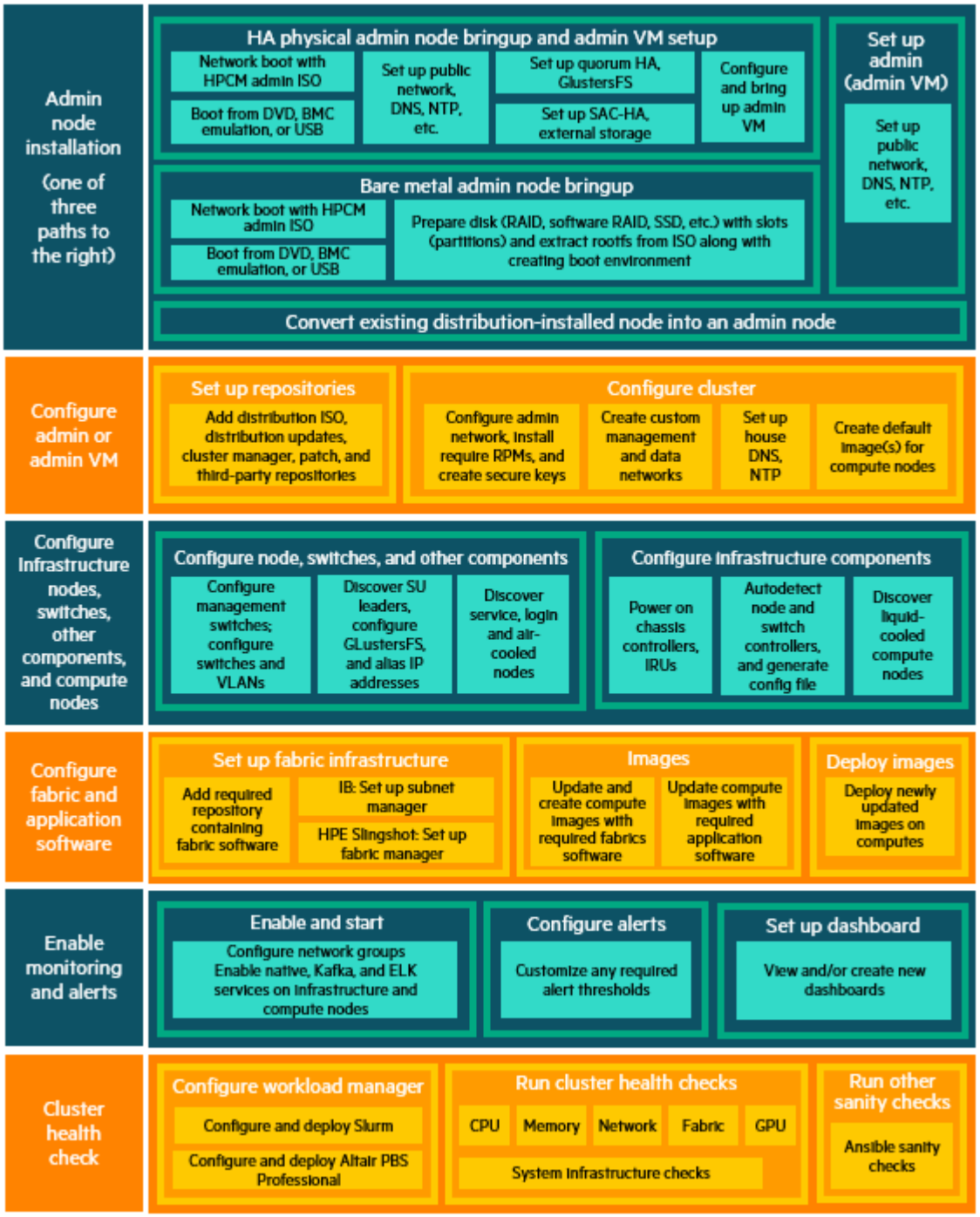

Figure 2: Installation process

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:

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.

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.