Hewlett Packard
Enterprise

Cray ClusterStor data services Installation Guide 2.0 (S-1238 Revision B)

Part Number: S-1238
Published: December 2021
Edition: 1

Cray ClusterStor data services Installation Guide 2.0 (S-1238 Revision B)

Abstract
This guide contains instructions for installing and configuring ClusterStor data services

Part Number: S-1238
Published: December 2021
Edition: 1

Table of contents

Notices

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

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

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

Acknowledgments

Support and other resources

Accessing Hewlett Packard Enterprise Support

- For live assistance, go to the Contact Hewlett Packard Enterprise Worldwide website:

  https://www.hpe.com/info/assistance

- To access documentation and support services, go to the Hewlett Packard Enterprise Support Center website:

  https://www.hpe.com/support/hpesc

Information to collect

- Technical support registration number (if applicable)
- Product name, model or version, and serial number
- Operating system name and version
- Firmware version
- Error messages
- Product-specific reports and logs
- Add-on products or components
- Third-party products or components

Accessing updates

- Some software products provide a mechanism for accessing software updates through the product interface. Review your product documentation to identify the recommended software update method.
- To download product updates:

  Hewlett Packard Enterprise Support Center

  https://www.hpe.com/support/hpesc

  Hewlett Packard Enterprise Support Center: Software downloads

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

  My HPE Software Center

  https://www.hpe.com/software/hpesoftwarecenter
- To subscribe to eNewsletters and alerts:

  https://www.hpe.com/support/e-updates
- To view and update your entitlements, and to link your contracts and warranties with your profile, go to the Hewlett Packard Enterprise Support Center More Information on Access to Support Materials page:

  https://www.hpe.com/support/AccessToSupportMaterials

IMPORTANT:

Access to some updates might require product entitlement when accessed through the Hewlett Packard Enterprise Support Center. You must have an HPE Passport set up with relevant entitlements.

Remote support

Remote support is available with supported devices as part of your warranty or contractual support agreement. It provides intelligent event diagnosis, and automatic, secure submission of hardware event notifications to Hewlett Packard Enterprise, which initiates a fast and accurate resolution based on the service level of your product. Hewlett Packard Enterprise strongly recommends that you register your device for remote support.

If your product includes additional remote support details, use search to locate that information.

HPE Get Connected

https://www.hpe.com/services/getconnected

HPE Pointnext Tech Care

https://www.hpe.com/services/techcare

HPE Complete Care

https://www.hpe.com/services/completecare

Warranty information

To view the warranty information for your product, see the links provided below:

HPE ProLiant and IA-32 Servers and Options

https://www.hpe.com/support/ProLiantServers-Warranties

HPE Enterprise and Cloudline Servers

https://www.hpe.com/support/EnterpriseServers-Warranties

HPE Storage Products

https://www.hpe.com/support/Storage-Warranties

HPE Networking Products

https://www.hpe.com/support/Networking-Warranties

Regulatory information

To view the regulatory information for your product, view the Safety and Compliance Information for Server, Storage, Power, Networking, and Rack Products, available at the Hewlett Packard Enterprise Support Center:

https://www.hpe.com/support/Safety-Compliance-EnterpriseProducts

Additional regulatory information

Hewlett Packard Enterprise is committed to providing our customers with information about the chemical substances in our products as needed to comply with legal requirements such as REACH (Regulation EC No 1907/2006 of the European Parliament and the Council). A chemical information report for this product can be found at:

https://www.hpe.com/info/reach

For Hewlett Packard Enterprise product environmental and safety information and compliance data, including RoHS and REACH, see:

https://www.hpe.com/info/ecodata

For Hewlett Packard Enterprise environmental information, including company programs, product recycling, and energy efficiency, see:

https://www.hpe.com/info/environment

Documentation feedback

Hewlett Packard Enterprise is committed to providing documentation that meets your needs. To help us improve the documentation, use the Feedback button and icons (located at the bottom of an opened document) on the Hewlett Packard Enterprise Support Center portal (**https://www.hpe.com/support/hpesc**) to send any errors, suggestions, or comments. All document information is captured by the process.

About the Cray ClusterStor data services Installation Guide

The Cray ClusterStor data services Installation Guide S-1238 includes procedures to install the ClusterStor data services software onto compatible servers. Perform these procedures before or after system delivery to the customer site.

Table 1: Record of revision

Publication Title	Date	Updates
Cray ClusterStor data services Installation Guide S-1238 (2.0) Revision B	December 2021	Many corrections throughout the guide. Simplified and clarified many procedures. Added new procedure for setting the BIOS time on nodes.
Cray ClusterStor data services Installation Guide S-1238 (2.0) Revision A	October 2021	Updated "Add or Remove Users in Keycloak" to use *DOMAIN_NAME*.
Cray ClusterStor data services Installation Guide S-1238 (2.0)	October 2021	New install process based on new base platform. New network cabling and VLAN setup. Updates to Emitter service configuration. CSMS CLI and remote `kubectl` setup. Kibana setup procedure no longer required.
Cray ClusterStor data services Installation Guide S-1238 (1.1)	May 2021	Added procedure for Aruba management switch configuration.
Cray ClusterStor data services Installation Guide S-1238 (1.0)	January 2021	Removed node requirements, SLES HA no longer needed
Cray ClusterStor data services Installation Guide S-1238 (0.3)	September 2020	The base OS file system is now XFS, DL325 is the basis for hardware requirements.
Cray ClusterStor data services Installation Guide S-1238 (0.2)	February 2020	Corrections and updates to match changes in the installer.
Cray ClusterStor data services Installation Guide S-1238 (0.1)	November 2019	Initial Release

Scope and audience

This publication is for:

- HPE service staff
- HPE manufacturing staff
- Customer personnel

This publication assumes that the reader is familiar with Cray ClusterStor storage systems and Lustre commands, terminology, and architecture. This guide contains procedures and reference information to support installing release 2.0 on compatible hardware. It does not include information about post-installation configuration, which is contained in the Cray ClusterStor data services Administration Guide S-1237.

Typographic conventions

`Monospace`	Indicates program code, command-line prompts, program output, file names and paths, and other software constructs.
`Monospaced Bold`	Indicates commands to enter on a command line or in response to an interactive prompt.
Oblique or *Italics*	Indicates user-supplied values in commands or syntax definitions.
Proportional Bold	Indicates a GUI Window, GUI element, cascading menu (Ctrl > Alt > Delete), or key strokes (press Enter).

`\` (backslash)	At the end of a command line, indicates a shell or command-line continuation character. The `bash` shell parses lines joined by a backslash as a single line.

Trademarks

Command Example Conventions

Host names and accounts in command prompts

The host name in a command prompt indicates on what host or type of host to run the command. The prompt also indicates the account that must run the command.

- The `root` or super-user account always has the `#` character at the end of the prompt.
- Any non-`root` account is indicated with a `$`. A user account that is not `root` or `admin` is referred to as "user".

`install#`	Run the command as `root` on the system that hosts, serves, and executes the files that install ClusterStor data services on the three management nodes.
`ext-adm#`	Run the command as `root` on an external administration system for the ClusterStor data services cluster.
`csms$`	Run the command on a host that has (or will soon have) the CSMS CLI installed and initialized.
`VM-W#`	Run the command on a ClusterStor data services Kubernetes worker node virtual machine as `root`.
`CDS#`	Run the command on a ClusterStor data services node as `root`.
`CDS$`	Run the command on a ClusterStor data services node as a user.
`CDS1#`	Run the command on the primary ClusterStor data services node as `root`.
`CDS1$`	Run the command on the primary ClusterStor data services node as a user.
`MGMT0#`	Run the command on the primary ClusterStor management node as `root`.
`MGMT0$`	Run the command on the primary ClusterStor management node as `admin`.
`MGMT1#`	Run the command on the secondary ClusterStor management node as `root`.
`MGMT1$`	Run the command on the secondary ClusterStor management node as `admin`.
`OSS#`	Run the command on an OSS node as `root`.
`OSS$`	Run the command on an OSS node as a user.
`MGS#`	Run the command on an MGS node as `root`.
`MGS$`	Run the command on an MGS node as a user.
`MDS#`	Run the command on an MDS node as `root`.
`MDS$`	Run the command on an MDS node as a user.
`sw0#`	Run the command on the primary management switch with administrative privileges.
`sw1#`	Run the command on the secondary management switch with administrative privileges.
`client$`	Run the command on a Lustre client node as any user.
`client#`	Run the command on a Lustre client node as `root`.
`user@`*`hostname`*`$`	Run the command on the specified system as any user.

Lustre file system names

The name of the Lustre file system seen in command examples is `cls12345`, with a mount point of /lus on the Lustre clients. The following example demonstrates this convention:

```
client$ lfs df -h
UUID                       bytes        Used   Available  Use% Mounted on
cls12345-MDT0000_UUID        2.0T       59.1G        1.9T   4% /lus[MDT:0]
cls12345-MDT0001_UUID        2.0T       97.6M        1.9T   1% /lus[MDT:1]
cls12345-OST0000_UUID      112.0T        5.1T      105.8T   5% /lus[OST:0]
cls12345-OST0001_UUID      112.0T        5.1T      105.8T   5% /lus[OST:1]
cls12345-OST0002_UUID       15.3T      425.8G       14.8T   3% /lus[OST:2]
cls12345-OST0003_UUID       15.3T      423.4G       14.8T   3% /lus[OST:3]

filesystem_summary:       254.6T       11.0T      241.1T   5% /lus
```

Command Prompt inside a Kubernetes pod

If executing a shell inside a container of a Kubernetes pod where the pod name is *podName*, the prompt changes to indicate that it is inside the pod. Not all shells are available within every pod, the following is an example using a commonly available shell.

```
kubectl exec -it podName /bin/sh
pod#
```

Directory Path in Command Prompt

Example prompts do not include the directory path, because long paths can reduce the clarity of examples. Usually, the commands can run in any directory. When a command must run within a specific directory, examples use the `cd` command to change into the necessary directory.

For example, here are actual prompts as they appear on the system:

```
client:~ # cd /etc
client:/etc# cd /var/tmp
client:/var/tmp# ls file
```

And here are the same prompts as they appear in this publication:

```
client# cd /etc
client# cd /var/tmp
client# ls file
```

About ClusterStor data services Networks

External Access to ClusterStor data services

ClusterStor data services provides services for data movement as well as:

- Kibana
- Elasticsearch
- Grafana
- Monitoring services
- Various management services

Access to services is load balanced across the service nodes. ClusterStor data services uses BGP routing on the management switches and a software-based load balancer running on the service nodes. The table in this topic describes the externally exposed platform services available over the Access Network (AN). Each of the services described requires an IP address on the AN subnet to be reachable. The load balancer assigns these IP addresses, using a subset of the AN subnet designated for dynamic allocation.

Services under Istio Ingress Gateway and Keycloak Gatekeeper Ingress share an ingress, so those services all use the IP allocated to that ingress.

The endpoint of each service has a DNS name (grafana.mycds.example.com, for example) served by the authoritative subdomain DNS service of the platform. This ClusterStor data services DNS service allows hosts on the site network to resolve the service names. Therefore, external hosts can access each of these services by name rather than finding the allocated IP. The ClusterStor data services installation software prepends the DNS name to the `MERCURY_SYSTEM_EXTERNAL_DOMAIN` specified in the system_install_env.list file used during installation.

Table 2: ClusterStor data services Service Access

Service	DNS Name	External Port
Ingress Gateway	N/A (multiple services)	80/443, 8081, 8888
REST service APIs (through ingress gateway)	`api`	443
Authentication (through ingress gateway)	`auth`	
External DNS	N/A	53
Web UI Ingress	N/A (multiple services)	443
Sysmgmt-health (through web UI ingress)	`grafana`	443
SMA Grafana	`sma-grafana`	443
SMA Kibana	`sma-kibana`	443
Rsyslog-Aggregator	`rsyslog`	514/8514

The API Gateway URL for accessing the APIs on a specific system is https://api. *SYSTEM_NAME*. *DOMAIN_NAME*/apis/.

Service Security and Resiliency

The ClusterStor data services management services consists of RESTful micro-service APIs. All of the supported API services provide an HTTP interface and are accessible through a single gateway URL. The API gateway for the system is available at a predictable URL based on the domain name of the system. It acts as a single HTTPS endpoint for terminating Transport Layer Security (TLS) using the configured certificate authority. All services and the API gateway are not dependent on any single node. This resilient arrangement ensures that services remain available during possible underlying hardware and network failures.

Admins and users must retrieve a token for authentication before attempting to access APIs through the gateway and present a valid token with each API call. The gateway makes the authentication and authorization decisions, which prevent unauthorized API calls from reaching the underlying micro-services.

Traffic Between ClusterStor data services and the Site Network

The traffic between ClusterStor data services and the customer site network consists of:

- Production API access to ClusterStor data services from the CSMS CLI, Lustre-generated requests, and other sources.
- Access to ancillary services. Such services include authentication, management, configuration, software updates, hardware control, monitoring, and logging.

- Authentication traffic from ClusterStor data services to the customer site LDAP, Active Directory, or other authentication service.

ClusterStor data services Switches

Generally, ClusterStor data services will share management switches with the ClusterStor E1000. Therefore, the switch configuration instructions required will depend on the switch model installed for the E1000. If ClusterStor data services will have its own management switches, they should be Aruba 8360.

ClusterStor data services VLANs

ClusterStor data services release 2.0 requires these VLANs:

- **VLAN2 for the Node Management Network (NMN):** This VLAN enables the ClusterStor data services nodes to communicate with each other. This communication flows over a private network coordinated by Kubernetes. The IP addresses of the nodes must remain under Kubernetes control. This private network and the internal DNS service ensure this control. This VLAN secures the internal traffic against external sniffing.
- **VLAN4 for the Hardware Management Network (HMN):** This VLAN provides access to the iLO ports, giving administrators node power control over the ClusterStor data services nodes. This VLAN also connects the same interfaces on those nodes as the NMN and AN. The base platform isolates this traffic. As a result, the base management platform retains exclusive control over ClusterStor data services node power and enables discovery. This VLAN also prevents exposure of the PXE boot service beyond the PXE host.
- **VLAN7 for the Access Network (AN):** This VLAN provides high-availability access to ClusterStor data services through a fixed endpoint. The AN uses a load balancer (MetalLB) and BGP to advertise routes to the ClusterStor data services cluster for a given service IP. This high-availability model is a change from the pairwise failover model of ClusterStor Neo. In the Neo, all services had two fixed failover endpoints, while in ClusterStor data services the service endpoints are dynamic.

The base management platform for ClusterStor data services includes an Access Network (AN). The AN provides:

- External access to services provided by ClusterStor data services.
- External access to the underlying management platform.
- A path for communication from the E1000 MMU and workload managers to ClusterStor data services.

The AN is a VLAN within the management network infrastructure of the platform. A virtual router is implemented in the platform management switches. This virtual router routes between endpoints on the AN and the External Administrator Network (EAN). The EAN connects the ClusterStor data services system to the customer site network.

Incoming access to ClusterStor data services requires a service ingress IP address on the External Administration Network (EAN). The EAN refers to an access point from the customer site network, but not a particular switch. The EAN connects to ClusterStor data services through the Administrative Network (AN) VLAN on the ClusterStor data services switch. A proxy service routes traffic to the correct ClusterStor data services host. This routing minimizes the number of EAN connections required to one per switch and allows for dynamic service placement and node deployment.

Network connection details

ClusterStor data services use the network connections listed in the following table:

Table 3: ClusterStor data services network connections

Network	Provides Connectivity Between	Purpose	Details
Node Management Network (NMN)	ClusterStor data services nodes	Carries all ClusterStor data services inter-node traffic.	• Uses a private VLAN on the ClusterStor Management Switch. ClusterStor, however does not manage ClusterStor data services nodes. • This network connection exists as an untagged VLAN on the two bonded 10GbE network interfaces.

Network	Provides Connectivity Between	Purpose	Details
Access Network (AN)	ClusterStor data services Emitter running on the ClusterStor MDS and the ClusterStor data services nodes	Provides load-balanced access to ClusterStor data services from the External Administrator Network. Carries the data movement requests relayed from Lustre to ClusterStor data services. Transports all ClusterStor data services requests (both Lustre and API). ClusterStor data services presents its Data Movement API on the AN.	• Uses a private VLAN on the ClusterStor Management Switch. ClusterStor, however does not manage ClusterStor data services nodes. • Lustre generates requests on the MMU and routed through the SMU to the EAN. From the EAN, the requests flow to the ClusterStor data services nodes over the AN. • This network connection exists as a tagged VLAN on the two bonded 10GbE network interfaces.
External Administrator Network (EAN)	Customer site administrative network to the ClusterStor management switches and either: • The ClusterStor SMU • The ClusterStor SSMU	Enables ClusterStor data services to receive external requests for data movement.	• This network propagates to all the ClusterStor data services nodes through the AN VLAN. • Requires one 10Gb Ethernet connection per ClusterStor data services switch (not per node).
High Speed Network (HSN)	ClusterStor nodes and ClusterStor data services nodes	Carries Lustre file system traffic needed for bulk data movement and Scalable Search.	• ClusterStor data services uses the Lustre HSN. • The network adapter on the node must match the Lustre interconnect type and speed. One option is: ○ P23664-B21 HPE HDR IB/EN 200Gb 1-port in Slot 1 or 2. • All HSN interfaces must connect to the HSN using appropriate cables.
Hardware Management Network (HMN)	Hardware Administrators and ClusterStor data services nodes	Provides provisioning and management of ClusterStor data services nodes. Used for discovery, installation, and power control.	• This network connection exists as a tagged VLAN on the two bonded 10GbE network interfaces. • The iLO interfaces, switches, and Power Distribution Units (PDUs) also connect to the HMN, where it is not tagged.

More information

Installation Prerequisites

Installation Overview

The process of installing the ClusterStor data services software stack on the physical ClusterStor data services nodes (servers) consists of several phases. Perform these phases in this order:

Procedure

1. Verify that the site meets the networking requirements in Installation Prerequisites. Then determine the configuration file settings listed in that topic.
2. Perform Configure the Aruba Management Switch .
3. Perform Prepare for Installation.
4. Perform Edit the system_install_env.list file.
5. Perform Set BIOS Clock on ClusterStor data services Nodes .
6. Perform Run `b_splat_bmc` .
7. Perform Run the Final Installation Containers .
8. Perform Add or Remove Users in Keycloak.
9. Perform Configure an External Administration System .
10. Perform Install CSMS CLI.
11. Perform Configure the Emitter User in Keycloak .
12. Perform Configure the Emitter Service .

Installation Prerequisites

ClusterStor data services has network connectivity requirements. Customer sites must satisfy these requirements before starting the install process. This topic explains these requirements in detail.

Customer network requirements

Customer sites must provide:

- A /26 subnet that can route to and from the site network for the Access Network (AN)
- Two IP addresses on the site network for the External Administration Network (EAN).
- A subdomain to use for administrating ClusterStor data services. This subdomain must take the form of *SUBDOMAIN.DOMAIN.DOMAIN_EXTENSION* (cds.cray.com, for example). The `MERCURY_SYSTEM_EXTERNAL_DOMAIN` variable in the system_install_env.list file stores this subdomain.

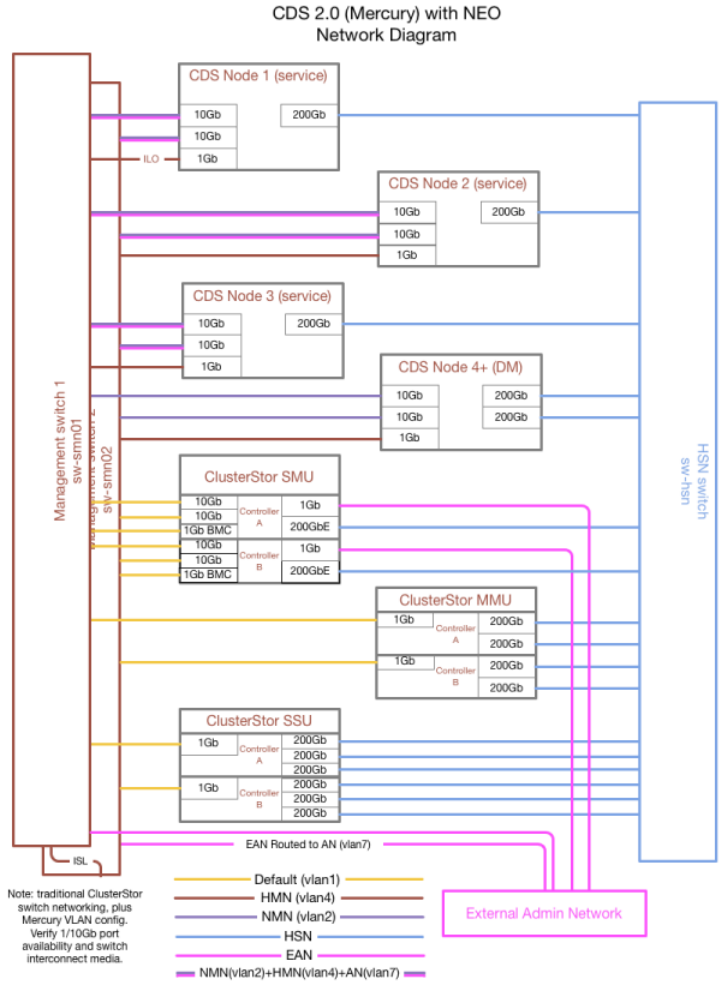
Network Connections

The following diagram details the ClusterStor data services node networking connections as well as example connections between those nodes and the ClusterStor system. ClusterStor network configuration is outside the scope of this document. Refer instead to the ClusterStor installation and administration documentation.

Customer sites must use this diagram to physically connect the following before installing the ClusterStor data services software:

- ClusterStor data services nodes
- ClusterStor nodes
- ClusterStor management switches
- Customer switch (for the EAN)
- ClusterStor High Speed Network (HSN) switch

Figure 1: ClusterStor data services networking diagram



Networking Changes from Release 1.0 to 2.0

Customer sites that already have ClusterStor data services release 1.0 deployed must make the following cabling and networking changes before installing release 2.0:

- **First ClusterStor data services node:** Disconnect the 10Gb interface from the EAN and reattach it to the second management switch.
- **Second ClusterStor data services node:**
  - Disconnect the 10Gb interface from the EAN and reattach it to the second management switch.
  - Move move the connection from the iLO to the second management switch.
- **Third ClusterStor data services node:** Connect a second 10Gb interface to the second management switch.
- **Management switches:**
  - Move EAN connections from the ClusterStor data services nodes to the management switches.
  - Remove VLAN22.

HSN Internet Protocol Addresses

ClusterStor data services requires 33 contiguous IP addresses on the High Speed Network (HSN). Each of the three ClusterStor data services nodes requires 11 IP addresses. These IP addresses are for the `MERCURY_SRN_ADDRESSES_1`, `MERCURY_SRN_ADDRESSES_2`, and `MERCURY_SRN_ADDRESSES_3` variables in the system_install_env.list installation configuration file.

This allocation is separate from the ClusterStor Internet Protocol address allocation.

Site Integration Information

Customer sites must compile several site-specific configuration settings. Record these settings in the system_install_env.list file during the installation process. The ClusterStor data services installer software then uses these settings to configure the Access Network and DNS to integrate properly with the site network.

The following table lists and describes the subset of settings in system_install_env.list relevant to site network integration. The system_install_env.list file requires separate settings for each server. Therefore the *X* at the end of the variable name in the following table will be either `1`, `2`, or `3`, depending on the ClusterStor data services node configured.

The following table uses the example base AN subnet address of 10.113.125.64/26 to explain the process for calculating the required Internet Protocol site addresses. This example subnet provides IP addresses from 10.113.125.65 - 10.113.125.126.

Variable in system_install_env.list	Description
`MERCURY_HOSTNAME_X`	Sets the hostname for the ClusterStor data services node
`MERCURY_CAN_CIDR_X`	The Internet Protocol assigned to the AN network interface in the base operating system. The last octet of `MERCURY_CAN_CIDR_1` must start +4 from the base Internet Protocol of the AN subnet allocated for the system. For example, if the allocated subnet is 10.113.125.64/26, then this variable will be `10.113.125.68` for the first node. The +4 offset accounts for the management switch configuration using the initial IP addresses in the subnet. The `MERCURY_CAN_CIDR_X` addresses can be calculated from the base AN network Internet Protocol address: • `MERCURY_CAN_CIDR_2` is offset from the base AN Internet Protocol address by 15. This results in an example address of 10.113.125.79. • `MERCURY_CAN_CIDR_3` s offset from the base AN Internet Protocol address by 26. This results in an example address of 10.113.125.90.

Variable in system_install_env.list	Description
`MERCURY_CAN_ADDRESSES_X`	A range of 10 IP addresses used to configure the AN interfaces on the virtual machines of the node. These IP addresses must immediately follow the Internet Protocol assigned in `MERCURY_CAN_CIDR_X` for the same node. The Internet Protocol range for the second node immediately follows the range for the first node. The `MERCURY_CAN_ADDRESSES_X` are a range of 10 IP addresses immediately following the `MERCURY_CAN_CIDR_X` Internet Protocol address of the node. The `MERCURY_CAN_CIDR_X` of the next node picks up at the next contiguous Internet Protocol address. These addresses can be calculated from the `MERCURY_CAN_CIDR_X`: • `MERCURY_CAN_ADDRESSES_1` range is the set of IP addresses starting with +1 and ending at +10 from the `MERCURY_CAN_CIDR_1` address. • `MERCURY_CAN_ADDRESSES_2` range is the set of IP addresses starting with +1 and ending at +10 from the `MERCURY_CAN_CIDR_2` address. • `MERCURY_CAN_ADDRESSES_3` range is the set of IP addresses starting with +1 and ending at +10 from the `MERCURY_CAN_CIDR_3` address. For example, if this range for the first node is `"10.113.125.69-10.113.125.78"` the range for the second node will be `10.113.125.80-10.113.125.89`. The range for the third node would be `10.113.125.91-10.113.125.100`
`MERCURY_DEFAULT_GATEWAY`	This variable sets the default route for the ClusterStor data services nodes. The value must be the Internet Protocol address assigned to VLAN 7 on the first management switch. This address is the first usable address of the AN subnet. For a AN subnet of 10.113.125.64/26, this value would be `10.113.125.65`.
`MERCURY_SYSTEM_EXTERNAL_DOMAIN`	This variable sets the subdomain, combining the system name and site domain. External hosts use this subdomain to access services provided by this system.
`MERCURY_SYSTEM_DNS_ADDRESS`	This Internet Protocol address must begin +62 from the base IP of the AN subnet allocated for the system. For example, if the allocated subnet is 10.113.125.64/26, then this value would be `10.113.124.126`. This address is the last usable Internet Protocol address in the AN subnet. This variable sets the Internet Protocol address for the external subdomain of the system (`MERCURY_SYSTEM_EXTERNAL_DOMAIN`). Add this Internet Protocol address to the DNS of the site.

Variable in system_install_env.list	Description
`MERCURY_CAN_METALLBADDRESSES_DYNAMIC`	This variable sets a range of 11 IP addresses in the AN subnet. MetalLB uses this range as a dynamic pool for allocating IP addresses to services provided by the system. This range must immediately follow the Internet Protocol address range used for `MERCURY_CAN_ADDRESSES_3`. For example, if `MERCURY_CAN_ADDRESSES_3` is `10.113.125.91-10.113.125.100`, then this value will be `10.113.125.101-10.113.125.111`.
`MERCURY_CAN_METALLBADDRESSES`	This variable sets the remainder of the AN subnet IP addresses. MetalLB uses these addresses as a pool of static IP addresses assigned to certain services. One of these services is the authoritative DNS service for the system. For example, this range would be `10.113.125.112-10.113.125.126` if the other variables set according to the examples in this table.

Site DNS Configuration for ClusterStor data services

Administer ClusterStor data services through the following methods from an external administration system:

- Web GUIs
- `csms` CLI commands
- `kubectl` commands

The `csms` CLI and web GUIs use Uniform Resource Locators (URLs) that contain the names of API endpoints. Theses URLs are based on the ClusterStor data services subdomain (`MERCURY_SYSTEM_EXTERNAL_DOMAIN`). Customer sites must add entries for this subdomain to the site authoritative DNS server for the domain.
Sites can delegate the entire subdomain to the ClusterStor data services external DNS service. To configure the site DNS this way, create the following records in the DNS server for the primary domain of the site:

- Name Server (NS) records pointing to the authoritative name servers for the subdomain.
- Address (A) records for the subdomain name server: This subdomain is the external DNS server running in the ClusterStor data services system.

The installation software automatically configures the ClusterStor data services DNS server with the following:

- A Start of Authority SOA record for the subdomain
- Two or more NS records
- Address (A) records for the services available in the subdomain

More information

About ClusterStor data services Networks

Configure the Aruba Management Switch

Skip the Aruba switch configuration if the ClusterStor E1000 system and the ClusterStor data services nodes:

- share management switches, and
- were built, configured, and shipped together

Prerequisites

- Verify that the site network meets the requirements in Installation Prerequisites.

Procedure

1. Physically connect the ClusterStor management switches to one another and to the ClusterStor data services nodes according to Installation Prerequisites. Record the switch ports used for the inter-switch link (ISL) and ClusterStor data services management nodes connections.

   The model-specific ClusterStor management switch configuration procedures require these port numbers.

2. Inspect the current configuration of the ClusterStor data services management switches. Skip the rest of this procedure if the switches are already configured.

   Refer to the Aruba documentation for the specific switch model. Refer to Configure Aruba 6300M Switches or Configure Aruba 8360 Switches for the configuration required by ClusterStor data services.

3. Perform one of following procedures depending on the model of the Aruba management switches:

   - Perform Configure Aruba 6300M Switches if the ClusterStor management switches are Aruba model 6300M.
   - Perform Configure Aruba 8360 Switches if the ClusterStor management switches are Aruba model 8360.

Configure Aruba 8360 Switches

This procedure modifies an existing ClusterStor E1000 management switch configuration to add support for ClusterStor data services release 2.0. This procedure does not cover management switch configuration settings which are common to both ClusterStor data services release 2.0 and ClusterStor E1000. Such common settings include:

- Switch name
- Admin password

Prerequisites

- Verify that ClusterStor data services release 2.0 will connect to a ClusterStor E1000 (not L300) system.
- Configure the Aruba 8360 management switches for the ClusterStor E1000.
- Verify that site network meets the requirements specified in Installation Prerequisites.

Procedure

1. Log into the primary management switch of the pair using SSH or a console.
2. Execute the `configure terminal` command.

   The remaining steps require this context.
3. Create the VLANs required by ClusterStor data services release 2.0 if they do not exist.

   ```
   sw0#
   vlan 2 name NMN vsx-sync description Node Mgmt Network vlan 4 name HMN vsx-sync description H/W Mgmt Network vlan 7 name AN vsx-sync description Access Network
   ```
4. Modify the existing E1000 spanning tree configuration to support per-vlan spanning tree.

   E1000 configuration enables spanning tree and sets the priority to 4 for the primary switch and 8 for the secondary switch. E1000 does not use per-vlan spanning tree, but ClusterStor data services release 2.0 does.

   ```
   sw0# spanning-tree mode rpvst spanning-tree spanning-tree vlan 1,2,4,7
   ```
5. Create multichassis Link Aggregation Groups (MCLAGs) if there is not one already.

   These MCLAGS support active-active bonded interfaces to the ClusterStor data services nodes.

   ```
   sw0#
   interface lag 1 multi-chassis no shutdown description MCLAG to SMU no routing vlan trunk native 2 vlan trunk allowed 2,4,7 lacp mode active interface lag 2 multi-ch
   ```
6. Create a LAG for the inter-switch link (ISL) if one does not already exist.

   ```
   sw0# interface lag 256 no shutdown description ISL for VSX no routing vlan trunk native 1 tag vlan trunk allowed all lacp mode active
   ```

   The ClusterStor E1000 generally does not use multiple ISL connections between the switches in the pair for a LAG.

   This command simplifies the reconfiguration of the ISL on an active system by including the `vlan trunk` statements in a single command. Later steps use this `vlan trunk`.
7. Configure the VLANs on the first management switch in the pair by running the following command. Omit the `interface vlan 2` and `interface vlan 4` sections if these VLANs already exist. Replace *AN_NETWORK_ADDRESS* in the `vlan 7` section with the first available IP address reserved by the site for the Access Network (AN).

   This command configures the VLANs used for the private Node Management Network (NMN), Hardware Management Network (HMN), and the AN.

   ```
   sw0#
   interface vlan 2 vsx-sync active-gateways ip mtu 9198 ip address 10.252.0.2/16 active-gateway ip mac 12:01:00:00:01:00 active-gateway ip 10.252.0.1 ip helper-addres
   ```
8. Configure the VLANs on the second management switch in the pair by running the following command.

   The only differences between the commands for the first and second switches are the IP addresses used for each of the three VLANs.

   Again, omit `interface vlan 2` and `interface vlan 4` sections if these VLANs already exist. Replace *AN_NETWORK_ADDRESS_2* in the `vlan 7` section with the second IP address reserved by the site for the AN.

   ```
   sw1#
   interface vlan 2 vsx-sync active-gateways ip mtu 9198 ip address 10.252.0.3/16 active-gateway ip mac 12:01:00:00:01:00 active-gateway ip 10.252.0.1 ip helper-addres
   ```
9. Configure the SNMPv3 user if this user is not already configured by running the following command on both switches.

   This command assumes the default password.

   ```
   snmpv3 user testuser auth md5 auth-pass ciphertext \ AQBapf1TKYh28GLx4x7Bp5XyAT0j2jnm9fDMNei1tR+BTyrqCQAAAITcQ4YsQX2noQ== \ priv des priv-pass ciphertext \ AQBapaNP
   ```
10. Configure the VSX on the first management switch.

    ```
    sw0#
    vsx inter-switch-link lag 256 role primary linkup-delay-timer 5 vsx-sync bgp copp-policy dns mclag-interfaces ospf route-map snmp ssh static-routes stp-global time
    ```
11. Configure the VSX on the second management switch.

    ```
    sw1#
    vsx inter-switch-link lag 256 role secondary linkup-delay-timer 5 vsx-sync bgp copp-policy dns mclag-interfaces ospf route-map snmp ssh static-routes stp-global tim
    ```
12. Configure the routes on the first management switch by running the following command. Replace *SITE_GW_IP* with the IP address of a gateway on the local site network.

    ```
    sw0# ip route 0.0.0.0/0 SITE_GW_IP ip route 10.92.100.60/32 10.252.0.6 ip route 10.94.100.60/32 10.254.0.6
    ```
13. Configure the routes on the second management switch by running the following command. Replace *AN_NETWORK_ADDRESS_1* with the IP address assigned to VLAN7 on the first management switch in Step 6 if there is a single connection to the local site network.

    The IP address for the default route must be the first available one provided for the AN network.

    ```
    sw1# ip route 0.0.0.0/0 AN_NETWORK_ADDRESS_1 ip route 10.92.100.60/32 10.252.0.6 ip route 10.94.100.60/32 10.254.0.6
    ```
14. Configure BGP on the first management switch by modifying and then executing the following commands.

    Modify the following commands per site network configuration as follows:
    - Replace *AN_IP_SUBNET* with the AN IP address range provided by the site.
    - Replace the IP addresses for the `pl-bgp-an` entries with assigned IP addresses from the site-provided range used for the AN. These addresses are *PL_BGP_AN_IP_1*, *PL_BGP_AN_IP_2*, and *PL_BGP_AN_IP_3* in the example commands.
    - Replace the IP address used for the `next-hop` entries for `pl-bgp-an` with address that are offsets of +6, +17, and +28 from the base subnet IP used for `pl-bgp-an`.

    ```
    sw0#
    ip prefix-list pl-bgp-an seq 10 permit AN_IP_SUBNET ge 26 ip prefix-list pl-bgp-hmn seq 20 permit 10.94.100.0/24 ge 24 ip prefix-list pl-bgp-nmn seq 30 permit 10.92
    ```
15. Configure BGP on the second management switch. Use the same BGP configuration as the first management switch, except for the `bgp router-id` value.

    The only configuration difference between the two switches is the `bgp router-id`.

    ```
    sw1#
    ip prefix-list pl-bgp-ean seq 10 permit AN_IP_SUBNET ge 26 ip prefix-list pl-bgp-hmn seq 20 permit 10.94.100.0/24 ge 24 ip prefix-list pl-bgp-nmn seq 30 permit 10.9
    ```

16. Configure the interface or interfaces used for the ISL if the ISL does not already exist.

    This step will disrupt traffic between nodes on the management network until the ISL configuration is completed.

    a. Verify that the port used for the ISL between the primary and secondary management switches is the same. Record the interface identifier.

    b. Run the following command on the secondary switch. Replace *ISL_INTERFACE_ID* with the interface identifier determined in the previous substep.

    ```
    sw1# interface ISL_INTERFACE_ID no shutdown mtu 9198 description ISL/VSX link lag 256
    ```

    The connection to the secondary switch may drop if you are using an SSH session for this procedure.

    c. Run the previous command on the primary management switch. Replace the interface identifier if needed. Use the switch serial console to complete configuration of ISL connection if you are unable to log into the first switch through SSH.

17. Identify three switch ports on each switch that are available for the DL325 host OS connections of the ClusterStor data services management nodes. Record the interface identifiers for use in the next step.

    The ClusterStor data services management nodes require three connections per switch. Two of these connections are for the OS and one is for the iLO. Each of these nodes must have a connection to each of the two management switches. The ClusterStor data services management nodes bond these two interfaces at the host end.

18. Configure ports for the management nodes by executing the following commands on the primary switch. Replace *CDS_INTERFACE_ID_1*, *CDS_INTERFACE_ID_2*, and *CDS_INTERFACE_ID_3* with the interface IDs connected to the first, second, and third ClusterStor data services nodes, respectively.

    ```
    sw0#
    interface CDS_INTERFACE_ID_1 no shutdown mtu 9198 description SMU MCLAG link lag 1 interface CDS_INTERFACE_ID_2 no shutdown mtu 9198 description SMU MCLAG link lag
    ```

19. Configure the ports for the DL325 iLO connections. Run the following command for each port on the primary switch connected to an iLO interface of ClusterStor data services node. Replace *ILO_INTERFACE_ID* with the interface identifier of the port connected to a different DL325 each time the you execute the command.

    ```
    sw0# interface ILO_INTERFACE_ID no shutdown mtu 9198 description HMN access link no routing vlan access 4
    ```

20. Repeat the previous step on the secondary management switch.

21. Configure the port that connects to the site network. Run the following command on the switch that connects to the site network.

    Modify the command as follows:

    - Replace *SITE_NET_INTERFACE_ID* with the identifier of the switch port that will connect to a switch on the site network. Port 48 is typically reserved for this connection. Use a different port if 48 is already used.
    - Replace *SITE_NETWORK_IP* with an IP address on the site network that can be used to route onto the AN network. This IP address is not on the AN subnet.
    - Replace `P2P AN to LOC01LABCANS1-Port1-4` with a more appropriate port description.

    ```
    default interface SITE_NET_INTERFACE_ID interface SITE_NET_INTERFACE_ID no shutdown mtu 9198 description P2P EAN to LOC01LABCANS1-Port1-4 ip address SITE_NETWORK_IP
    ```

22. Save these configuration changes by executing the `write memory` command on both switches.

Configure Aruba 6300M Switches

This procedure modifies an existing ClusterStor E1000 management switch configuration to add support for ClusterStor data services release 2.0. This procedure does not cover management switch configuration settings which are common to both ClusterStor data services release 2.0 and ClusterStor E1000. Such common settings include:

- Switch name
- Admin password

Prerequisites

- Verify that ClusterStor data services release 2.0 is paired to a ClusterStor E1000 (not L300) system.
- Verify that ClusterStor data services nodes will share the Aruba 6300M management switches with the E1000.
- Configure the Aruba 6300M management switches for the ClusterStor E1000.
- Verify that the site network meets the requirements specified in Installation Prerequisites.

Procedure

1. Log into the primary management switch of the pair using SSH or a console.
2. Execute the `configure terminal` command.

   The remaining steps require this context.
3. Create the VLANs required by ClusterStor data services release 2.0 if they do not exist.

   ```
   sw0# vlan 2 Name NMN description Node Mgmt Network vlan 4 name HMN description H/W Mgmt Network vlan 7 name AN description Access Network
   ```
4. Modify the existing E1000 Spanning Tree configuration to support per-vlan Spanning Tree.

   E1000 configuration enables Spanning Tree and sets the priority to 4 for the primary switch and 8 for the secondary switch. E1000 does not use per-vlan Spanning Tree, but ClusterStor data services release 2.0 does.

   ```
   sw0# spanning-tree mode rpvst spanning-tree spanning-tree vlan 1,2,4,7
   ```
5. Create a LAG for the inter-switch link (ISL) if one is not configured already.

   ```
   sw0# interface lag 256 no shutdown description ISL Link no routing vlan trunk native 1 tag vlan trunk allowed all
   ```

   The ClusterStor E1000 generally does not use multiple ISL connections between the switches in the pair for a LAG.

   This command simplifies the reconfiguration of the ISL on an active system by including the `vlan trunk` statements in a single command.
6. Configure the VLANs on the first management switch in the pair by running the following command. Omit the `interface vlan 2` and `interface vlan 4` sections if these VLANs are already configured. Replace *AN_NETWORK_ADDRESS* in the `vlan 7` section with the first available IP address reserved by the site for the Access Network (AN).

   This command configures the VLANs used for the private Node Management Network (NMN), Hardware Management Network (HMN), and the AN.

   ```
   sw0#
   interface vlan 2 ip mtu 9198 ip address 10.252.0.2/16 active-gateway ip mac 12:01:00:00:01:00 active-gateway ip 10.252.0.1 ip helper-address 10.92.100.222 interface
   ```
7. Configure the VLANs on the second management switch in the pair by running the following command.

   The only differences between the commands for the first and second switches are the IP addresses used for each of the three VLANs.

   Again, omit `interface vlan 2` and `interface vlan 4` sections if these VLANs are already configured. Replace *AN_NETWORK_ADDRESS_2* in the `vlan 7` section with the second IP address reserved by the site for the AN.

   ```
   sw1#
   interface vlan 2 ip mtu 9198 ip address 10.252.0.3/16 active-gateway ip mac 12:01:00:00:01:00 active-gateway ip 10.252.0.1 ip helper-address 10.92.100.222 interface
   ```
8. Configure the SNMPv3 user if this user is not already configured by running the following command on both switches.

   This command assumes the default password.

   ```
   snmpv3 user testuser auth md5 auth-pass ciphertext \ AQBapflTKYh28GLx4x7Bp5XyAT0j2jnm9fDMNeiltR+BTyrqCQAAAITcQ4YsQX2noQ== \ priv des priv-pass ciphertext \ AQBapaNP
   ```
9. Configure the routes on the first management switch by running the following command. Replace *SITE_GW_IP* with the IP address of a gateway on the local site network.

   ```
   sw0# ip route 0.0.0.0/0 SITE_GW_IP ip route 10.92.100.60/32 10.252.0.6 ip route 10.94.100.60/32 10.254.0.6
   ```
10. Configure the routes on the second management switch by running the following command. Replace *AN_NETWORK_ADDRESS_1* with the IP address assigned to VLAN7 on the first management switch in Step 6 if there is a single connection to the local site network.

    The IP address for the default route must be the first available one provided for the AN network.

    ```
    sw1# ip route 0.0.0.0/0 AN_NETWORK_ADDRESS_1 ip route 10.92.100.60/32 10.252.0.6 ip route 10.94.100.60/32 10.254.0.6
    ```
11. Configure BGP on the first management switch by modifying and then executing the following commands.

    Modify the following commands per site network configuration as follows:
    - Replace *AN_IP_SUBNET* with the AN IP address range provided by the site.
    - Replace the IP addresses for the `pl-bgp-an` entries with assigned IP addresses from the site-provided range used for the AN. These addresses are identified as *PL_BGP_AN_IP_1*, *PL_BGP_AN_IP_2*, and *PL_BGP_AN_IP_3* in the example commands.
    - Replace the IP address used for the `next-hop` entries for `pl-bgp-an` with address that are offsets of +6, +17, and +28 from the base subnet IP used for `pl-bgp-an`.

    ```
    sw0#
    ip prefix-list pl-bgp-an seq 10 permit AN_IP_SUBNET ge 26 ip prefix-list pl-bgp-hmn seq 20 permit 10.94.100.0/24 ge 24 ip prefix-list pl-bgp-nmn seq 30 permit 10.92
    ```
12. Configure BGP on the second management switch. Use the same BGP configuration as the first management switch, except for the `bgp router-id` value.

    The only configuration difference between the two switches is the `bgp router-id`.

    ```
    sw1#
    ip prefix-list pl-bgp-ean seq 10 permit AN_IP_SUBNET ge 26 ip prefix-list pl-bgp-hmn seq 20 permit 10.94.100.0/24 ge 24 ip prefix-list pl-bgp-nmn seq 30 permit 10.9
    ```
13. Configure the interface or interfaces used for the ISL if the ISL is not already configured.

    This step will disrupt traffic between nodes on the management network until the ISL configuration is completed.

    a. Verify that the port used for the ISL between the primary and secondary management switches is the same. Record the interface identifier.

    b. Run the following command on the secondary switch. Replace *ISL_INTERFACE_ID* with the interface identifier determined in the previous substep.

       ```
       sw1# interface ISL_INTERFACE_ID no shutdown mtu 9198 description ISL link lag 256
       ```

       The connection to the secondary switch may drop if you are using an SSH session for this procedure.

    c. Run the previous command on the primary management switch. Replace the interface identifier if needed. Use the switch serial console to complete configuration of ISL connection if you are unable to log into the first switch through SSH.
14. Identify three switch ports on each switch that are available for the DL325 host OS connections of the ClusterStor data services management nodes. Record the interface identifiers for use in the next step.

The ClusterStor data services management nodes require three connections per switch. Two of these connections are for the OS and one is for the iLO. Each of these nodes must have a connection to each of the two management switches. The ClusterStor data services management nodes bond these two interfaces at the host end.

15. Configure ports for the ClusterStor data services management nodes. Execute the following command one time on the primary switch for each of the three nodes. Replace *CDS_INTERFACE_ID* in the command with one of the interface identifiers determined in the previous step for the primary switch.

    ```
    sw0# interface CDS_INTERFACE_ID no shutdown mtu 9198 description CDS Access Link no routing vlan trunk native 2 vlan trunk allowed 2,4,7
    ```

16. Repeat the previous step for the secondary management switch.

17. Configure the ports for the DL325 iLO connections by running the following command for each port connected to an iLO interface. Replace *ILO_INTERFACE_ID* with the interface identifier of the port connected to a different DL325 each time the command is run.

    ```
    sw0# interface ILO_INTERFACE_ID no shutdown mtu 9198 description HMN access link no routing vlan access 4
    ```

18. Repeat the previous step on the secondary management switch.

19. Configure the port that connects to the site network. Run the following command on the switch that connects to the site network.

    Modify the command as follows:

    - Replace *SITE_NET_INTERFACE_ID* with the identifier of the switch port that will be cabled to a switch on the site network. Port 48 is typically reserved for this connection. Use a different port if 48 is already used.
    - Replace *SITE_NETWORK_IP* with an IP address on the site network that can be used to route onto the AN network. This IP address is not on the AN subnet.
    - Replace `P2P AN to LOC01LABCANS1-Port1-4` with a more appropriate port description.

    ```
    default interface SITE_NET_INTERFACE_ID interface SITE_NET_INTERFACE_ID no shutdown mtu 9198 description P2P EAN to LOC01LABCANS1-Port1-4 ip address SITE_NETWORK_IP
    ```

20. Save these configuration changes by executing the `write memory` command on both switches.

Prepare for Installation

This procedure ensures that the site can proceed with the ClusterStor data services installation.

Prerequisites

- Perform either Configure Aruba 6300M Switches or Configure Aruba 8360 Switches depending on the model of the E1000 management switches.
- Verify that the customer site network matches the requirements described in Installation Prerequisites.

Procedure

1. Compile and record the system_install_env.list variable settings listed in Installation Prerequisites.

   Later steps in the installation process use these IP addresses.

2. Obtain or allocate a host that will install the ClusterStor data services software on the ClusterStor data services nodes.

   This installation system must use Linux or a similar OS and be able to run Podman.

3. Set a static IP address of 10.254.0.254 on the installation system.

4. Configure the iLO interfaces of the ClusterStor data services nodes. Assign these interfaces static IP addresses within the HMN VLAN (10.254.0.200 to 10.254.0.253).

   Refer to **https://support.hpe.com/hpesc/public/docDisplay?docId=a00029920en_us&docLocale=en_US** for instructions on configuring a static IP address for the iLO.

5. Connect the installation system to the ClusterStor management switch pair.

6. Install Podman on the installation system.

7. Download the ClusterStor data services release 2.0 product files onto the installation system.

   These files include:
   - **kjplat-airgaprelease-kjmgmt-1.0-202109030195057.iso:** the OpenSUSE image for the bare metal phase of the installation
   - **kjplat-airgaprelease-kjmgmt-1.0-20210930201040.qcow2:** the Virtual Machine (VM) installation image for the VMs that will run on the base OS. Kubernetes runs and manages ClusterStor data services microservices on these VMs.
   - **cds-2.0.15.sdp:** the Software Deployment Package (SDP) file that contains the ClusterStor data services software.
   - **installer-kjmgmt-1.0.0-20211005095900.tar:** An archive that contains the installer Podman containers and the system_install_env.list installation configuration file. These containers will run on the installation system and install the ClusterStor data services stack on the three nodes.

8. Start the NTP container.

   This container provides a time synchronization service during the installation. For more information on this container, refer to **https://hub.podman.com/r/cturra/ntp**.

   a. Download the NTP container.

   ```
   install$ podman pull cturra/ntp
   ```

   b. Start the NTP container.

   ```
   install$ podman run --name=ntp --restart=always --detach --publish=123:123/udp cturra/ntp
   ```

9. Launch the web server container.

   This container provides the access to the installation files over the network.

   a. Disable SELinux if the OS of the installation system is Linux and has SELinux enabled.

   ```
   install# setenforce 0
   ```

   b. Launch the web server container. Replace *INSTALL_FILE_PATH* with the full path to the directory that contains all the installation files except the container files.

   ```
   install#
   podman run -it --rm -d -p 8080:80 --name web -v \INSTALL_FILE_PATH:/usr/share/nginx/html:ro -d nginx
   ```

10. Configure the authoritative DNS server for the site domain to forward queries for the ClusterStor data services subdomain to the ClusterStor data services DNS server.

    The DNS instance at the customer site must forward the subdomain specified by the *MERCURY_SYSTEM_EXTERNAL_DOMAIN* variable to the IP address specified by *MERCURY_SYSTEM_DNS_ADDRESS*. Both of these values are stored in the system_install_env.lst file.

Edit the system_install_env.list file

The ClusterStor data services installation process uses the information stored in the system_install_env.list configuration file to:

- Install the base OS.
- Install the VMs.
- Configure the bond mode and VLANs on the two 10GbE interfaces.
- Install the base management platform.
- Deploy the ClusterStor data services software.

Skip this procedure if a customized system_install_env.list file already exists.

Prerequisites

Prepare for Installation

Procedure

Configure the installation process

1. Extract the installer archive file.

   ```
   install# tar xvf installer-kjmgmt-1.0.0*.tar
   ```

2. Change the current working directory to the newly extracted directory.

   ```
   install# cd installer-kjmgmt-1.0.0
   ```

   This directory includes a template system_install_env.list file to customize if a site-specific system_install_env.list does not exist at this point.

3. Open the system_install_env.list file in a text editor on the installation system.
4. Set `NUMBER_NODES` to `3`.
5. Set the values for the iLO access.
   a. Set `BMC_IP_1` to the iLO IP address for the first ClusterStor data services node.
   b. Set `BMC_USERNAME1` to the user name for the iLO for the first ClusterStor data services node.
   c. Set `BMC_PASSWORD1` to the password for the iLO for the first ClusterStor data services node.
   d. Repeat the previous three substeps for the other two ClusterStor data services nodes.
6. Set `MERCURY_HOSTNAME_1`, `MERCURY_HOSTNAME_2`, and `MERCURY_HOSTNAME_3` to the values determined in Installation Prerequisites.
7. Set `INSTALLATION_ISO_URL` to the following value: `http://10.254.0.254:8080/`*`ISO_IMAGE_NAME`*`.ISO`. Replace *ISO_IMAGE_NAME* with the full name of the installation ISO image file.
8. Set `TF_VAR_BASE_IMAGE_URL` to `http://10.254.0.254:8080/`*`QCOW2_IMAGE_NAME`*`.qcow2`. Replace *QCOW2_IMAGE_NAME* with the full name of the VM qcow2 image file.
9. Set `CSPLAT_ARG` to `http://10.254.0.254:8080/`*`SDP_FILE_NAME`*`.sdp`. Replace *SDP_FILE_NAME* with the full name of the SDP file.

Configure ClusterStor data services cluster

10. Set `MERCURY_DEFAULT_GATEWAY` to the value determined in Installation Prerequisites.
11. Set both `MERCURY_DOMAIN` and `TF_VAR_DOMAIN` to the wanted name for the ClusterStor data services cluster, appended with `.local`.

    Both of these variables must be set to the same value. They set a local, internal domain used only by the underlying platform.
12. Set `MERCURY_CAN_CIDR_1` and `MERCURY_CAN_ADDRESSES_1` to the values determined in Installation Prerequisites.
13. Repeat the previous step for the `MERCURY_CAN_CIDR` and `MERCURY_CAN_ADDRESSES` variables for the other two nodes.

14. Set `MERCURY_CAN_METALLBADDRESSES_DYNAMIC` to the IP address range determined in Installation Prerequisites.
15. Set `MERCURY_CAN_METALLBADDRESSES` to the IP address range determined in Installation Prerequisites.
16. Set `MERCURY_SYSTEM_EXTERNAL_DOMAIN` to the value determined in Installation Prerequisites.
17. Set `MERCURY_SYSTEM_DNS_ADDRESS` to the address determined in Installation Prerequisites.
18. Set `MERCURY_NTP_SERVERS` to a comma-separated list of the IP addresses accessible to the three ClusterStor data services nodes, starting with `10.254.0.254`. Do not insert any spaces in this list.

    This list must start with `10.254.0.254`, the HMN IP address of the installation system. The ClusterStor data services nodes use the NTP service running on the installation system before the installation software fully configures the network. Including the installation system ensures that the node times are in sync during initial configuration.
19. Set `SWITCH_6300M` to `true` if the management switches are Aruba model 6300M. Set this variable to `false` if the management switches are Aruba model 8360.
20. Set `MERCURY_SRN_ADDRESSES_1` to the range of 11 IP addresses on the ClusterStor HSN assigned to the first ClusterStor data services node.
21. Repeat the previous step for `MERCURY_SRN_ADDRESSES_2` and `MERCURY_SRN_ADDRESSES_3` for the second and third nodes, respectively.

    The range for the second node must begin immediately after the range for the first node. Likewise, the range for the third node must begin immediately after the range for the second node.
22. Set `MERCURY_SRN_TYPE` to either `ib` or `eth`, depending on the ClusterStor HSN type.
23. Set `FS_TYPE` to `lustre`.
24. Set `FS_NAME` to the name of the Lustre file system.
25. Set `FS_MGS` to the Lustre Network Identifiers (NIDs) of the ClusterStor MGS nodes.

    This value uses the same format as the `mount -t lustre` command. For example,

    ```
    FS_MGS=10.10.100.4@o2ib:10.10.100.5@o2ib
    ```
26. Set `FS_MOUNT_PT` to `/lus`.
27. Set `FS_MOUNT_OPTS` to `_netdev,noatime,lazystatfs,flock,noexec,nodev,suid`
28. Set `FS_LNET_ID` to the LNet id of the ClusterStor HSN.

    This value is derived from the LNet label. This value can usually be set to `@o2ib` since ClusterStor E1000 uses RDMA over Converged Ethernet (RoCE) by default. Customer sites that do not use RoCE can use other options such as `@tcp`.

Set BIOS Clock on ClusterStor data services Nodes

The system (BIOS) clocks for all three ClusterStor data services nodes must be in sync for the installation process to succeed.

Prerequisites

Procedure

1. Begin booting the first ClusterStor data services node.
2. Press the F9 key when that option first appears during the boot process.
3. Navigate to System Utilities > System Configuration > BIOS/Platform Configuration (RBSU) > Date and Time.
4. Enter the correct time.
5. Press the F12 key.

   The BIOS will save the changes and resume the boot process on the node.
6. Repeat the previous five steps on the other two ClusterStor data services node.

Run `b_splat_bmc`

The b_splat_bmc installer container uses the iLO interfaces of the ClusterStor data services nodes to install the bare metal OS onto those nodes.

Prerequisites

- Edit the system_install_env.list file
- Set BIOS Clock on ClusterStor data services Nodes

Procedure

Prepare to run the b_splat_bmc container

1. Change the current directory on the installation system to the extracted installer archive directory. Skip this step if this directory is the current one.

   ```
   install# cd installer-kjmgmt-1.0.0
   ```

   This directory includes a template system_install_env.list file to customize if a site-specific system_install_env.list does not exist at this point.

2. Load the three container *.tar* files into the local Podman repository.

   ```
   install# podman load -i cray-b_splat_bmc.tar
   install# podman load -i cray-b_splat_airgap_cloudinit.tar
   install# podman load -i cray-c_splat.tar
   ```

3. Confirm that the container images loaded properly and note the `TAG` version number for each container.

   ```
   install# podman image ls
   REPOSITORY                              TAG    IMAGE ID     CREATED     SIZE
   release/cray-b_splat_bmc                1.0.13 ad43dcd073b0 10 days ago 289MB
   release/cray-b_splat_airgap_cloudinit 1.0.13 e64f3ea6e506 10 days ago 281MB
   release/cray-c_splat                    1.0.2  e1d88810da6f 4 weeks ago 596MB
   ```

4. Set the boot order of the first ClusterStor data services management node.

   a. Open the HTTP web address of the iLO in a web browser.

   b. Click Administration > Boot Order.

   c. Verify that the Server Boot Order list has iLO Virtual CD-ROM and Generic USB Boot at the bottom. Set these two options as the last ones if they are not already.

   d. Repeat the previous three substeps on the other two ClusterStor data services management nodes.

Run the b_splat_bmc container

5. Run the b_splat_bmc installer container.

   RELEASE_VERSION_TAG is the version number for that container reported in Step 5.

   ```
   install# podman run -t --env-file ./system_install_env.list \ release/cray-b_splat_bmc:RELEASE_VERSION_TAG
   ```

6. Wait for the b_splat_bmc container to finish.

7. Verify that the system times for the ClusterStor data services nodes are in sync.

   In the following command:
   - *CDS_NODE_1* is the IP address or host name for the first ClusterStor data services node.
   - *CDS_NODE_2* is the IP address or host name for the second ClusterStor data services node.
   - *CDS_NODE_3* is the IP address or host name for the third ClusterStor data services node.

   The nodes use Pacific timezone (PST or PDT) in the output of the `date` command.

   ```
   install# pdsh -w root@CDS_NODE_1,root@CDS_NODE_2,root@CDS_NODE_3 date
   ```

Run the Final Installation Containers

To complete the ClusterStor data services installation, run the final two installation containers.

Prerequisites

Run `b_splat_bmc`

Procedure

1. Run the b_splat_airgap_cloudinit installer container.

   RELEASE_VERSION_TAG is the version number of the b_splat_airgap_cloudinit container reported during Run `b_splat_bmc`.

   ```
   install#
   podman run -t --env-file ./system_install_env.list \ release/cray-b_splat_airgap_cloudinit:RELEASE_VERSION_TAG
   ```

2. Run the c_splat container.

   RELEASE_VERSION_TAG is the version number of the c_splat container reported during Run `b_splat_bmc`.

   ```
   install# podman run -t --env-file ./system_install_env.list \ release/cray-c_splat:RELEASE_VERSION_TAG
   ```

Add or Remove Users in Keycloak

This procedure details how to use Keycloak to:

- Add one or more administrator accounts during or after installation of ClusterStor data services.
- Remove an account after ClusterStor data services installation.

Do not use this procedure to configure the account required by the Emitter service. Refer to Configure the Emitter User in Keycloak for those instructions.

Viewing ClusterStor data services dashboards requires an administrator-level account. Each system administrator must have their own administrator-level account.

Procedure

1. **Optional:** Obtain the `admin` password for the Keycloak web interface. Skip this step if the ClusterStor data services administrator already knows the Keycloak `admin` password.
   a. Log into the base operating system (OpenSUSE) of a ClusterStor data services node.
   b. Query Kubernetes for the Keycloak `admin` password.

   ```
   ext-adm#
   kubectl get secret -n services keycloak-master-admin-auth \ --template={{.data.password}} | base64 --decode; ech
   ```

2. Open the Keycloak Administration Console in a web browser.

   The web address for this console is https://auth. *SYSTEM_DOMAIN*/keycloak/admin/master/console. *SYSTEM_DOMAIN* is the ClusterStor data services cluster domain name set by the `MERCURY_SYSTEM_EXTERNAL_DOMAIN` variable during installation.

3. Log into Keycloak using the `admin` user name and the password.
4. Click Manage > Users in the pane on the left side of the window.
5. **Optional:** Remove a user. Refer to **https://www.keycloak.org/docs/latest/server_admin/index.html#proc-deleting-user_server_administration_guide** for instructions.
6. Add another user account.
   a. Click Add user in the main pane of the window.
   b. Click Save.
   c. Click the Details tab.
   d. Enter a user name for the account into the Username box.
   e. Set User Enabled button to ON.
   f. Click the Save button.
   g. Click the Credentials tab.
   h. Enter a password for the account into the Password and Password Confirmation boxes.
   i. Set the Temporary button to OFF.
   j. Click the Set Password button.
   k. Click the Role Mappings tab.
   l. Start typing **`shasta`** into the Client Roles box.
   m. Select `shasta` from the Client Roles drop-down list after it appears.
   n. Add the `admin` role to the Assigned Roles box.
7. **Optional:** Verify that Administration Console > Manage Users reflects the new or deleted user account.

Configure an External Administration System

HPE does not support logging into ClusterStor data services nodes after installation. This procedure configures an external system to run `kubectl` commands as necessary for administration tasks after installation is complete.

See **https://kubernetes.io/docs/concepts/configuration/organize-cluster-access-kubeconfig/** for more information about kubeconfig files.

Prerequisites

Run the Final Installation Containers

Procedure

1. Install `kubectl` on an external system allocated for ClusterStor data services administration.

   This system must connect to the local site network and be able to route to the ClusterStor data services cluster.

2. Copy the /root/.kube/config file from a ClusterStor data services node to the external management system.

   ```
   ext-adm# export CDSHOST=CDS_NODE_FQDN
   scp root@$CDSHOST:/root/.kube/config $CDSHOST.config
   ```

   CDS_NODE_FQDN is the external Fully Qualified Domain Name (FQDN) of a ClusterStor data services node.

3. Modify the configuration file.

   ```
   ext-adm# sed -i -E 's/10\(\.[0-9]\{1,3\}\)\{3\}:8443/'$CDSHOST':6443/' $CDSHOST.config
   ext-adm#
   sed -i -E 's/certificate-authority-data.*/insecure-skip-tls-verify: true/' $CDSHOST.config
   ```

   These commands replace the endpoint IP address with `$CDSHOST` and disable TLS host verification.

4. Copy this modified configuration file to the default location expected by `kubectl`.

   ```
   ext-adm# mv $CDSHOST.config ~/.kube/config
   ```

5. Verify that the external system can run `kubectl` commands for the ClusterStor data services system.

   ```
   ext-adm# kubectl get nodes
   NAME                          STATUS   ROLES    AGE   VERSION
   k8s-master-cdsxmp101-m3dl     Ready    master   18h   v1.18.6
   ...
   ```

6. Install the CSMS CLI on this external administration system.

   See Install CSMS CLI for instructions.

Install CSMS CLI

This procedure downloads and installs the CSMS CLI on an external Linux system. After completing this procedure, ClusterStor administrators can use CSMS CLI commands on that system to administer ClusterStor data services and move file data.

The CSMS CLI is a command-line interface for managing and monitoring ClusterStor data services clusters. This CLI can also manage data movement requests. ClusterStor data services administrators can use the CSMS CLI to perform software upgrades and other administrative operations.

Prerequisites

- Run the Final Installation Containers
- Perform Add or Remove Users in Keycloak to create the administrator-level account in Keycloak for this procedure.
- Obtain an external Linux system that either already has, or can have, Python 3 installed.

Procedure

1. Log into a Linux system that is not one of the ClusterStor data services nodes.

   Run all the commands in this procedure from this external system.

2. Save the Keycloak access token required to download the CSMS `tar` file.

   ADMIN is the user name of the administrator-level account and *ADMIN_PASSWORD* is the password for this account. *DOMAIN_NAME* is the external-facing domain name for the ClusterStor data services cluster set during installation by the `MERCURY_SYSTEM_EXTERNAL_DOMAIN` variable.

   ```
   ext-adm$
   export ACCESS_TOKEN=$(curl -k -s -d grant_type=password -d client_id=shasta \ -d username=ADMIN -d password=ADMIN_PASSWORD \ https://api.DOMAIN_NAME/keycloak/realms/s
   ```

3. Download the most recent CSMS CLI tarball.

   The following command is the preferred one and downloads the csms-latest.tar file from one of the ClusterStor data services nodes. The `wget` command uses the API gateway, but the load balancing performed by BGP selects a specific node to provide the file.

   The following example removes most of the output for brevity.

   ```
   ext-adm$ wget --no-check-certificate -d --header="Authorization: Bearer $ACCESS_TOKEN" \ https://api.DOMAIN_NAME/apis/csms/base/v1/cli/download --content-disposition
   . . .
   Saving to: 'csms-latest.tar'100%[=======================================================>] 140,713     --.-K/s   in 0.003s2021-11-18 16:14:46 (45.5 MB/s)
    - 'csms-latest.tar' saved [140713/140713]
   csms-latest.tar
   ```

4. Extract the downloaded CSMS CLI tarball.

   ```
   ext-adm$ tar -xvf csms-latest.tar
   x csms-latest/
   x csms-latest/csms-guide.pdf
   x csms-latest/csms-latest-py3-none-any.whl
   x csms-latest/csms.man
   x csms-latest/INSTALLATIONGUIDE.md
   ```

5. Install the CSMS CLI from the Python .whl file.

   The following commands install the csms-latest.tar file and assume that `pip` is a soft link to `pip3`.

   ```
   ext-adm$ python3 -m venv ~/myvenv
   ext-adm$ source ~/myvenv/bin/activate
   ext-adm$ pip install --upgrade pip
   ext-adm$ pip install --upgrade setuptools
   ext-adm$ pip install csms-latest/csms-latest-py3-none-any.whl
   ```

6. Initialize CSMS CLI to verify that the installation completed successfully.

   The `Cray Hostname` value must start with **`api`**.

   ```
   ext-adm$ csms init
   Cray Hostname: api.DOMAIN_NAME
   Username: ADMIN_USERNAME
   Password: ADMIN_PASSWORD

   Success!
   Initialization complete.
   ```

Configure the Emitter User in Keycloak

Starting with ClusterStor data services release 2.0, the Emitter service uses a separate account to request operations on behalf of Lustre clients. Configure this Emitter account in Keycloak after installing ClusterStor data services to allow Lustre users to manually request migration of particular files.

Prerequisites

Configure an External Administration System

Procedure

1. Obtain the `admin` password for the Keycloak web interface.

   The software installation process automatically sets this password.

   ```
   ext-adm#
   kubectl get secret -n services keycloak-master-admin-auth \ --template={{.data.password}} | base64 --decode
   ```

2. Open the following URL in a web browser to access the Keycloak Administration Console: https://auth.*CDS_DOMAIN*/keycloak/admin/master/console

   CDS_DOMAIN is the domain set by `MERCURY_SYSTEM_EXTERNAL_DOMAIN`.

3. Log into Keycloak using the `admin` user name and the password obtained in Step 1.
4. Click Manage > Users in the pane on the left side of the screen.
5. Click Add user in the main pane of the screen.
6. Enter the required values on the Details tab.
   a. Enter `emitter` into the Username box.
   b. Set User Enabled button to ON.
   c. Click the Save button.
7. Click the Credentials tab.
8. Enter the required values on the Credentials tab.
   a. Enter a password for the Emitter account into the Password and Password Confirmation boxes.
   b. Set the Temporary button to OFF.
   c. Click the Set Password button.
9. Click the Role Mappings tab.
10. Configure the roles for the Emitter user.
    a. Start typing `shasta` into the Client Roles box.
    b. Select `shasta` from the Client Roles drop-down list after it appears.
    c. Add the `admin` role to the Assigned Roles box.
11. Perform one of the following to verify successful `emitter` user setup:
    - Configure the Emitter Service if the Emitter service is not configured on the ClusterStor system.
    - Restart the Emitter service if this service was previously configured.

      ```
      MDS# systemctl restart cds_emitter
      ● cds_emitter.service - CDS emitter service
         Loaded: loaded (/usr/lib/systemd/system/cds_emitter.service; disabled; vendor preset: disabled)
         Active: active (running) since Thu 2021-07-22 22:54:43 UTC; 1s ago
        Process: 8187 ExecStop=/usr/sbin/lctl set_param mdt.*.hsm_control=shutdown (code=exited, status=0/SUCCESS)
        Process: 8191 ExecStartPost=/usr/sbin/lctl set_param mdt.*.hsm_control=external (code=exited, status=0/SUCCESS)
       Main PID: 8190 (cds_emitter)
         CGroup: /system.slice/cds_emitter.service
                 └─8190 /usr/sbin/cds_emitter -v

      Jul 22 22:54:43 cslmo4802 systemd[1]: Starting CDS emitter service...
      Jul 22 22:54:43 cslmo4802 lctl[8191]: mdt.testfs-MDT0000.hsm_control=external
      Jul 22 22:54:43 cslmo4802 systemd[1]: Started CDS emitter service.
      Jul 22 22:54:43 cslmo4802 cds_emitter[8190]: Establish NetLink connection
      Jul 22 22:54:44 cslmo4802 cds_emitter[8190]: NetLink connection complete
      Jul 22 22:54:44 cslmo4802 cds_emitter[8190]: Waiting for message from kernel
      ```

Configure the Emitter Service

The ClusterStor data services Emitter service must run on every ClusterStor MDS to provide ClusterStor data services data movement services directly to Lustre users through `lfs` commands. Data movement requests initiated by policy or through CSMS CLI commands do not require the Emitter service.

Configure the Emitter using `cscli` and `lctl` commands in ClusterStor.

ClusterStor will restart the Emitter service automatically when it terminates. Also, the Emitter is:

- **Started** when the ClusterStor data services IP address or secret are set.
- **Restarted** when either of the two are reset.
- **Stopped** when either of the two are cleared.

Use the `cscli cds` commands for all three of these actions.
Do not manually edit the emitter service configuration file, /etc/cds_emitter.conf. Puppet maintains this file on the ClusterStor nodes.

Prerequisites

Configure the Emitter User in Keycloak
Procedure

Set all MDTs to use an external HSM coordinator

1. Log into the ClusterStor primary management node as `admin`.
2. Set `hsm_control` on all MDTs to `external`, replacing *MGS* in the following command with the hostname of the active MGS node.

   This step requires the `sudo` command because the `lctl set_param` command requires `root` access on the MGS. The MGS controls the configuration of the Lustre file system and propagates this setting change to all the MDTs.

   This command does not return any output when successful. The next step confirms the change.

   ```
   MGMT0$ sudo pdsh -w MGS lctl set_param -P mdt.*.hsm_control=external
   ```

3. Confirm that all the MDTs will use an external HSM coordinator.

   The following command queries all the MDS nodes in the ClusterStor system to confirm the setting change the MDTs.

   MDS1_hostname and *MDS2_hostname* in this command output are the host names for the ClusterStor MDS nodes.

   ```
   MGMT0$ pdsh -g mds lctl get_param mdt.*.hsm_control
   MDS2_hostname: mdt.cls12345-MDT0001.hsm_control=external
   MDS1_hostname: mdt.cls12345-MDT0000.hsm_control=external
   ```

Configure Emitter parameters on the SMU

4. Set the host name that the Emitter will use to communicate with ClusterStor data services.

   CDS_DOMAIN in the following example is the value of `MERCURY_SYSTEM_EXTERNAL_DOMAIN` from the system_install_env.list file used during installation.

   ```
   MGMT0$ cscli cds ip_address -s api.CDS_DOMAIN
   cds: Updating of CDS IP Address is successfully
   cds: CDS ip address api.CDS_DOMAIN is registered
   ```

5. Set the password used by the Emitter to the `emitter` user password defined in Keycloak.

   The following command copies the password specified as *SECRET_STRING* to every MDS node in the ClusterStor system. This password enables the Emitter service running on the ClusterStor system to authenticate to the ClusterStor data services API Agent.

   ```
   MGMT0$ cscli cds secret -s SECRET_STRING
   cds: Updating of CDS Secret is successful
   cds: CDS secret key SECRET_STRING is registered
   ```

6. Confirm the status of the ClusterStor data services configuration.

   ```
   MGMT0$ cscli cds show
   Cds info :
   ```

```
==========================================

CDS IP Address is : api.CDS_DOMAIN

CDS secret key is : SECRET_STRING

cds ca cert is : /mnt/mgmt/var/lib/puppet/files/etc/cds_emitter.crt
```

This command confirms the correct Emitter service configuration and that the service started on all MDS nodes.

Scalable Search Service Index Directories

The Scalable Search Service creates and maintains a distributed file system metadata database within the Lustre file system. This database consists of directories named ._dbindex_ placed at known locations within the file system tree. These directories contain the sqlite database files and other files used by the Policy Engine and several ClusterStor data services tools.

Do not delete the ._dbindex_ directories, nor any of the files within them. Policy Engine and several Scalable Search tools may produce errors or incorrect results if they are missing.

The Scalable Search Service will, however, recreate any missing files during the next file system re-indexing.

Initial File System Indexing

ClusterStor data services will begin creating the initial indexes for the mounted Lustre file system within three hours after installation finishes. The number of files and directories in that file system will determine how much time this initial indexing will require. The larger the file system, the longer it will take to complete initial indexing.