CLFS Software Installation Guide

Contents

1 About Installing CLFS Software

ESF Software Release ESF XX-2.3.0

This publication (number S-2521-D) includes software installation procedures for Cray software release ESF XX-2.3.0 and supports Lustre® File System by Cray (CLFS) systems. This software applies to Cray XC30 and Cray XE6, Cray XE6m and Cray XK6 systems using a CLFS server.

- ESF XX-2.3.0 is installed on CLFS nodes and supports Lustre storage that provides connectivity to Cray supported storage devices.
- CLFS nodes are typically configured as Lustre MGS, MDS, or OSS nodes and are managed by Bright Cluster Manager 6.1 software (Bright).
- The Cray Integrated Management Services (CIMS) node runs Bright software and provisions the CLFS nodes with the CentOS 6.6 base operating system, Lustre software, Cray proprietary software and tools and utilities.
- `lustre_control stop` now puts a file system in suspend mode and `lustre_control start` removes a file system from suspend mode (via touch and removal of the appropriate file system files in `/var/esfsmon`)
- The `esfsmon` health check log location has moved from `/tmp/esfsmon` to /`tmp/esfsmon/`*`filesystem`* directory (`/var/esfsmon` has not changed)
- An email notification mechanism is available in `esfsmon` that is configured in `esfsmon.conf` as parameter `ESFSMON_MAILTO`. This requires that an email relay is configured for the CIMS node (`esms`).

This guide describes procedures for the following types of installations:

- An initial software installation involves installing and configuring the entire system and is generally performed for new hardware. Additional tasks included are configuring Bright categories for esfsmon 2.1.0, device-mapper multipath (DM Multipath), and Lustre are included.

  NOTE: An upgrade path for this software release is currently not available, an initial install is required. Sites must perform a fresh install of ESF XX-2.3.0 which includes Lustre 2.5 server and a new manual configuration of the ESF XX-2.3.0 image. The configuration settings from previous ESF software images are not transferred to the ESF XX-2.3.0 fresh install software image.

Typographic Conventions

`Monospace`	Indicates program code, reserved words, library functions, command-line prompts, screen output, file/path names, key strokes (e.g., `Enter` and `Alt-Ctrl-F`), and other software constructs.
`Monospaced Bold`	Indicates commands that must be entered 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 graphical user interface window or element.

\ (backslash)	At the end of a command line, indicates the Linux® shell line continuation character (lines joined by a backslash are parsed as a single line). Do not type anything after the backslash or the continuation feature will not work correctly.

Scope and Audience

This document is intended for experienced Linux system administrators.

Related Publications

The Bright administration guide and user guides are stored on the CIMS node (as Adobe™ PDF™ files) in the `/cm/shared/docs/cm` directory.

The following documents contain information that may be helpful:

- *Data Management Platform (DMP) Administrator's Guide (S-2327)*
- *Installing Cray Integrated Management Services (CIMS) Software (S-2522)*
- *Installing Cray Development and Login (CDL) Software (S-2520)*
- *Manage Lustre for the Cray Linux Environment (CLE) (S-0010)*

2 ESF XX-2.3.0 Software Introduction

This guide describes how to install and configure Lustre File System by Cray™ (CLFS) nodes with release ESF XX-2.3.0 software. ESF release ESF XX-2.3.0 includes the CentOS 6.6 base operating system.

A Cray Data Management Practices (DMP) system consists of a Cray management server (CIMS) node and one or more service nodes. Systems configured for high availability (HA) include a pair of CIMS nodes. The supported service nodes include login and development nodes (CDL) and shared Lustre™ file system (CLFS) nodes.

Each CLFS node type is configured with the same software image initially, then customized for its role in the system. CLFS nodes can be configured as metadata servers (MDS/MGS) or object storage servers (OSS) in a DMP system. MDS nodes typically have a connection to the site LDAP server whereas OSS servers have no external Ethernet connections to the customer network.

- A CLFS node is controlled and provisioned by the CIMS running Bright Cluster Manager™ 6.1 (Bright) software. The CIMS uses Bright to provision the software image and boot the CLFS node. Software to manage the Lustre™ file system, such as Cray `lustre_control` and esfsmon 2.1.0, is installed on the CIMS node, rather than on the CLFS node.
- An unmanaged node runs independently. Software is installed directly on the CLFS node. ESF XX-2.3.0 supports only managed CLFS nodes.

CLFS nodes provide a high-performance Lustre shared file system that operates independently of the Cray system. A DMP file system is comprised of metadata server nodes (MDS), object storage servers (OSS) and object storage target (OST) devices from DataDirect™ Networks (DDN) or NetApp™ within an InfiniBand® interconnect. Each CLFS node file system requires its own set of servers and block storage and are always configured with full redundancy in controllers, servers, and cabling for high-availability.

2.1 ESF Software Distribution Media

The release media contains the ESF and Lustre® software and is available as an ISO file named ESF XX-2.3.0-*DatestampVer*.iso. Cray uses the following convention for ISO file names:

PROD-ARCH-N.N.N-DatestampVer.iso

PROD	Product name, such as ESF, ESM, or ESL
ARCH	Supported architecture: XX for all architectures, XC for Cray XC series systems, and XE for Cray XE and XK series systems.
N.N.N	Release number, such as 2.3.0 for ESF XX-2.3.0.
Datestamp	Unique date stamp in the ISO name, in the format YYYYMMDDHHmm (such as `201509031326`)
Ver	Installer version, such as `a03`

2.2 ESF XX-2.3.0 Distribution Media ISOs

The following Cray ESF release media ISOs are required for a complete installation of ESF software.

ESF ISO

The ESF XX-2.3.0 ISO file name is similar to `ESF-XX-2.3.0-201509031326a06.iso`. The contents of this ISO file must be copied to `/root/isos` on the CIMS node when performing an initial software installation.

CentOS 6.6 ISO

The CentOS 6.6 release media contains the CentOS™ 6.6 base operating system. The CentOS 6.6 ISO file name is `CentOS-6.6-x86_64-bin-DVD1.iso`.

The ESF XX-2.3.0 software installation procedure requires that the `CentOS-6.6-x86_64-bin-DVD1.iso` ISO file exist in `/root/isos` on the CIMS node.

Bright 6.1 CentOS ISO

The ESF XX-2.3.0 software installation procedure requires Bright Cluster Manager updates for CentOS. This ISO file, `bright6.1-centos6u6.iso` must exist in `/root/isos` on the CIMS node when installing this software.

Lustre® 2.5.4

The ESF XX-2.3.0 release media ISO includes Lustre version 2.5.4.

3 Prepare to Install a Managed CLFS Node

Before you start the installation for a managed CLFS, Cray recommends the following actions:

- Read the *ESF Release Errata* and the *ESF README* provided with your ESF release package for any additional installation-related requirements, corrections to this installation guide, and other relevant information about the release package.
- Configure the BIOS settings for the node. Refer to *Configure BIOS Settings for a Dell R720 Managed CLFS Node* on page 11.
- Read the Field Notices (FNs) related to kernel security fixes to identify if there are any changes to this release package.
- Read this section and ensure that the prerequisites are satisfied before beginning an initial CLFS software installation; verify that the network connections are in place, and know the network configuration values. Refer to *CLFS Node Network Interfaces* on page 8.

  IMPORTANT: All existing data on the CLFS node will be lost on initial image installation. If any data should be saved, it must be archived before starting the installation.

3.1 CLFS Node Network Interfaces

The networks required by the CLFS node must be defined and configured in Bright on the CIMS node and assigned to the correct physical hardware interface on the CLFS node. The network interfaces are configured differently for a CLFS nodes in a Cray Data Management Practices (DMP) or Cray Tiered Adaptive Storage (TAS) system. All of the network settings and configuration information is configured using Bright Cluster Manager (Bright) software.

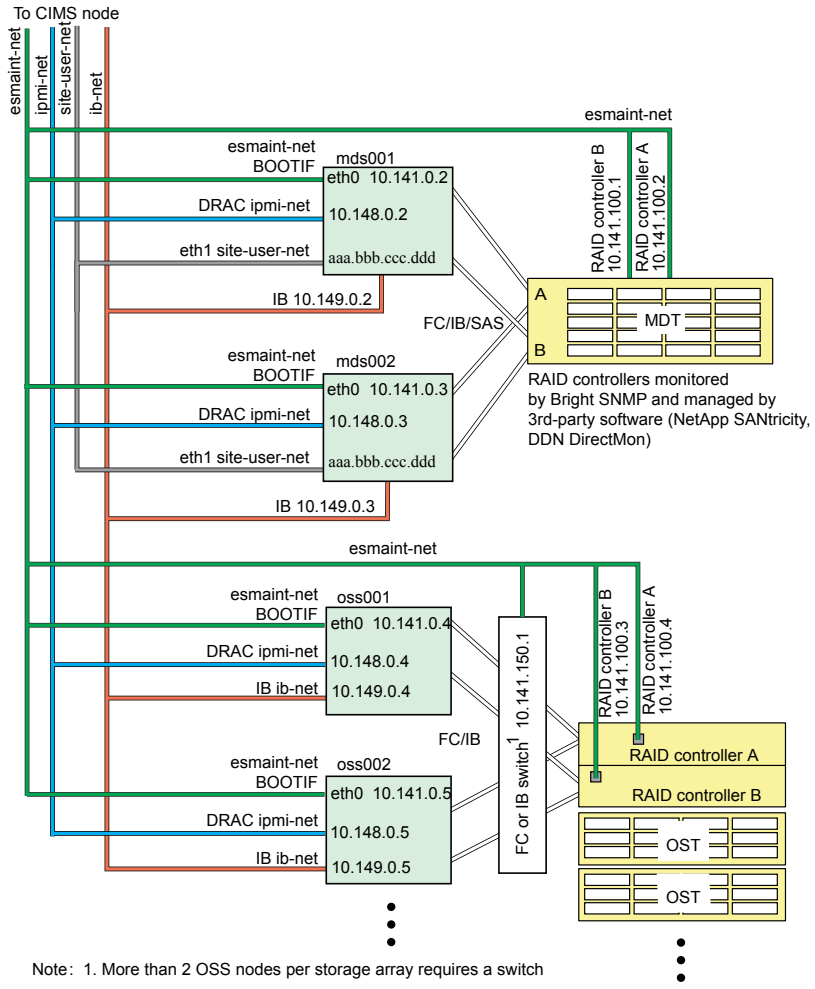
Other high-speed interfaces (InfiniBand® or 10GbE for example) adapters are also configured in Bright.

Table 1. CLFS Node Network Interfaces

Port	Description
`eth0`	Connection to the maintenance network (`esmaint-net`). This interface must be set to `BOOTIF` in Bright.
`eth1`	Connection to the external user (site) network for user access (`site-user-net`). This network may also support NIS or LDAP servers. For CIMS nodes, use `eth1` for `site-admin-net`.
`eth2`	Connection to the metadata network (`metadata-net`).
`eth3`	Connection to the failover network (`failover-net`).
`eth4` and `eth5`	Connection to the GigE customer network (`customer-net`).

Port	Description
ipmi0	Connection to the Integrated Dell Remote Access Controller (iDRAC) remote console and power management network (ipmi-net).
ib0	Connection to the InfiniBand network (ib-net). This interface card(s) must be in a slot that is capable of supporting the full bandwidth of the card. A CLFS node may also have additional FC or SAS interfaces for connections to the storage arrays or tape systems. These connections to storage must also be in a slot that is capable of supporting the full bandwidth of the card.

Figure 1. DMP CLFS Node Networks



3.2 CLFS Node System Configuration Values

The system's configuration values are site dependent. You may find it helpful to print the following table and record the actual values for your site.

Table 2. Site-dependent Configuration Values

Site Dependent Configuration Values Description	Example	Actual Value
MDS default gateway (usually on `site-user-net`)	*aaa.bbb.ccc.ddd*	
MDS domain name of customer network	*your.domain.com*	
MDS netmask of customer network	255.255.255.0	
MDS 1 host name	`esfs-mds001`	
MDS 1 IP address on maintenance network `BOOTIF` (`esmaint-net`)	10.141.0.2	
MDS 1 IPMI (iDRAC) IP address (`ipmi-net`)	10.148.0.2	
MDS 1 InfiniBand IP address (`ib-net`)	10.149.0.2	
MDS 1 `eth0` MAC address (check pull-out card on server)	01:02:03:04:05:06	
MDS 1 customer network IP address (`site-user-net`)	*aaa.bbb.ccc.ddd*	
MDS 2 host name	`esfs-mds002`	
MDS 2 IP address on maintenance network `BOOTIF` (`esmaint-net`)	10.141.0.3	
MDS 2 IPMI (iDRAC) IP address (`ipmi-net`)	10.148.0.3	
MDS 2 InfiniBand IP address (`ib-net`)	10.149.0.3	
MDS 2 `eth0` MAC address	01:02:03:04:05:06	
MDS 2 customer network IP address (`site-user-net`)	*aaa.bbb.ccc.ddd*	
OSS 1 host name	`esfs-oss001`	
OSS 1 IP address on maintenance network `BOOTIF` (`esmaint-net`)	10.141.0.4	
OSS 1 IPMI (iDRAC) IP address (`ipmi-net`)	10.148.0.4	

Site Dependent Configuration Values Description	Example	Actual Value
OSS 1 InfiniBand IP address (`ib-net`)	10.149.0.4	
OSS 1 `eth0` MAC address	01:02:03:04:05:06	
OSS 2 host name	`esfs-oss002`	
OSS 2 IP address on maintenance network `BOOTIF` (`esmaint-net`)	10.141.0.5	
OSS 2 IPMI (iDRAC) IP address (`ipmi-net`)	10.148.0.5	
OSS 2 InfiniBand IP address (`ib-net`)	10.149.0.5	
OSS 2 `eth0` MAC address	01:02:03:04:05:06	

3.3 ESF Software Installation Passwords

The following default account names and passwords are used throughout the ESF XX-2.3.0 software installation process. Cray recommends that sites change these default passwords after completing the installation.

Table 3. ESF XX-2.3.0 Default System Passwords

System or Software	Account Name	Password
CIMS node	`root`	`initial0`
CLFS node	`root`	`initial0`
Bright administrator	N/A	`initial0`

3.4 Configure BIOS Settings for a Dell R720 Managed CLFS Node

Prerequisites

Before configuring a managed CLFS node for ESF software installation and Bright Cluster Manager (Bright), configure the BIOS and Dell remote access controller (iDRAC) settings. The following items are required:

- The CLFS node must be cabled to both the administration network (`esmaint-net`) and IPMI network (`ipmi-net`).

- The CLFS node using a workload manager such as TORQUE or Moab® must be cabled directly to the SDB on the Cray system over the `wlm-net`.
- The CLFS node must be configured to provide IPMI Serial Over LAN (SOL) for remote console support.
- Processor hyper-threading (the logical processor setting) must be disabled on CLFS nodes.
- Configure DMP service nodes to remain powered off after a site power failure. Cray recommends that the CIMS node power on first and become operational before services nodes are powered on.

About this task

Use this procedure to configure the BIOS settings for a Dell R720 CLFS node before installing ESF software.

Procedure

1. Power up the service node. When the BIOS power-on self-test (POST) process begins, quickly press the **F2** key after the following menu displays in the upper-right corner of the console.

   ```
   F2 = System Setup
   F10 = System Services
   F11 = BIOS Boot Manager
   F12 = PXE Boot
   ```

   NOTE: Use the **Tab** key to move to different areas on the screen. To select an item, use the up-arrow and down-arrow keys to highlight the item, then press the **Enter** key. Press the **Esc** key to exit a submenu and return to the previous screen.

   When the **F2** key is recognized, the `F2 = System Setup` line changes to `Entering System Setup`. After the POST process completes and all disk and network controllers have been initialized, the `Dell System Setup` screen displays:

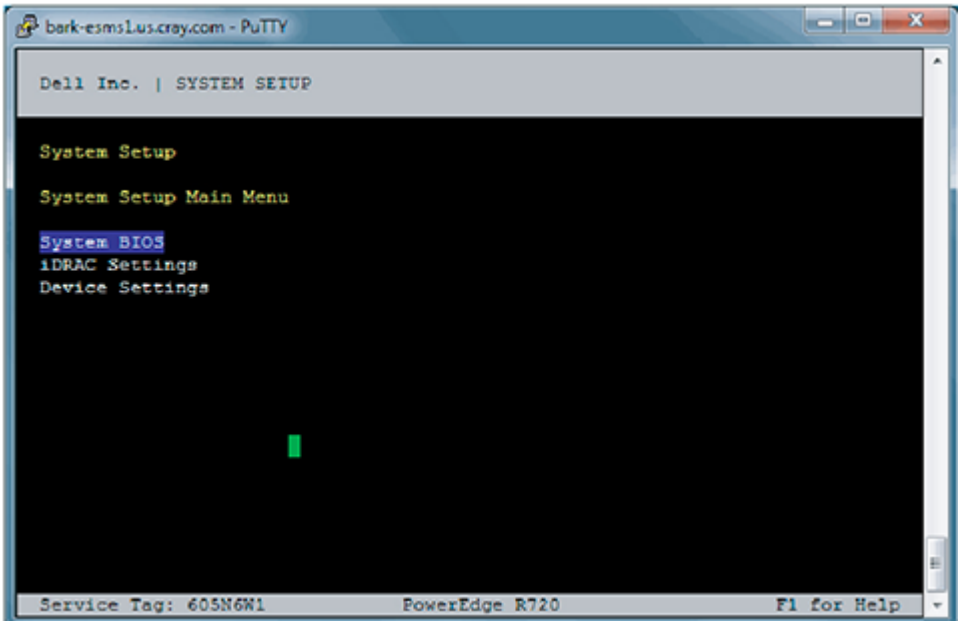
   ```
   System BIOS
   iDRAC Settings
   Device Settings
   ```

2. Change the system BIOS settings.

   a. Select **System BIOS**, then press **Enter**.

Figure 2. Dell System BIOS Settings

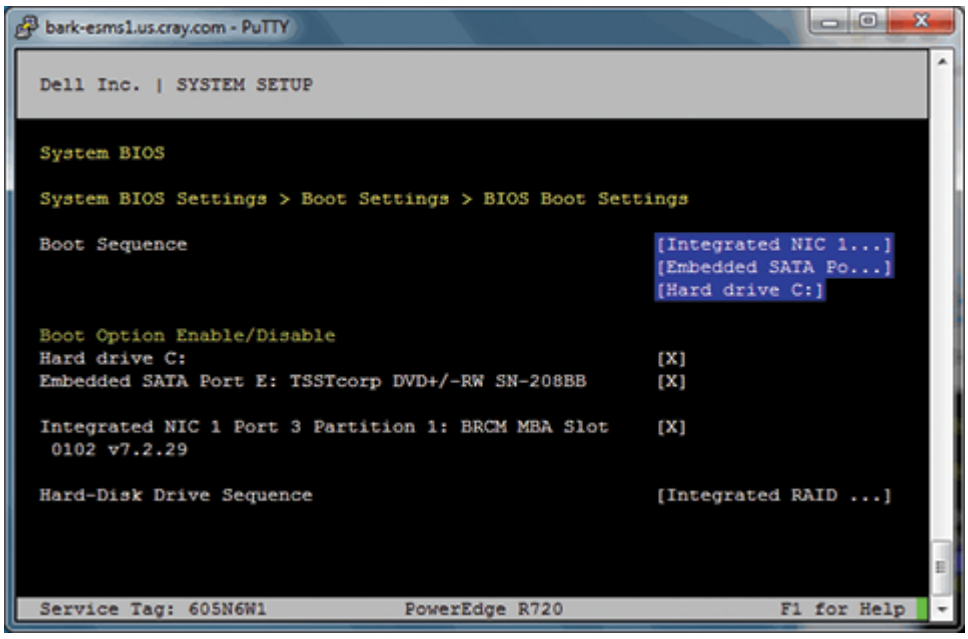
Dell R720 System BIOS System Setup menu



b. Select **Boot Settings**, then press **Enter**.

c. Select **BIOS Boot Settings**, then press **Enter**.

Figure 3. Dell R720 BIOS Boot Settings

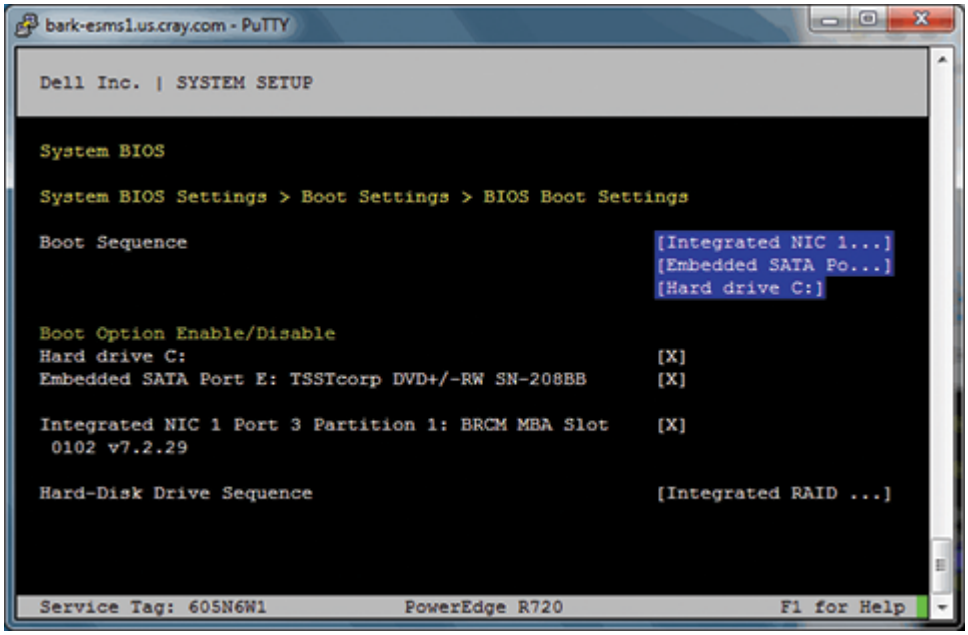


d. Select **Boot Sequence**, then press **Enter** to view the boot sequence.

e. In the pop-up window, change the boot order so that the integrated NIC appears first, before the optical (DVD) drive. The hard drive should be last on the list.Use the up-arrow or down-arrow key to highlight an item, then use the **+** and **-** keys to move the item up or down.

f. Be sure that **Hard Drive C:** is enabled **[X]** under the **Boot Option Enable/Disable** section.

Figure 4. Dell R720 BIOS Boot Sequence



g. Press **Enter** to return to the **BIOS Boot Settings** screen.

h. Press **Esc** to exit **BIOS Boot Settings**.

i. Press Escape to exit **Boot Settings** and return to the **System BIOS Settings** screen.

3. On the **System BIOS Settings** screen, select **Processor Settings** and press **Enter**.

a. Select **Logical Processor**, and press **Enter**. Verify that **Logical Processor** is set to **Disabled**.

b. Press **Esc** to exit **Processor Settings**.

4. Change the serial communication settings.

a. On the **System BIOS Settings** screen, select **Serial Communication**.

b. On the **Serial Communication** screen, select **Serial Communication**.

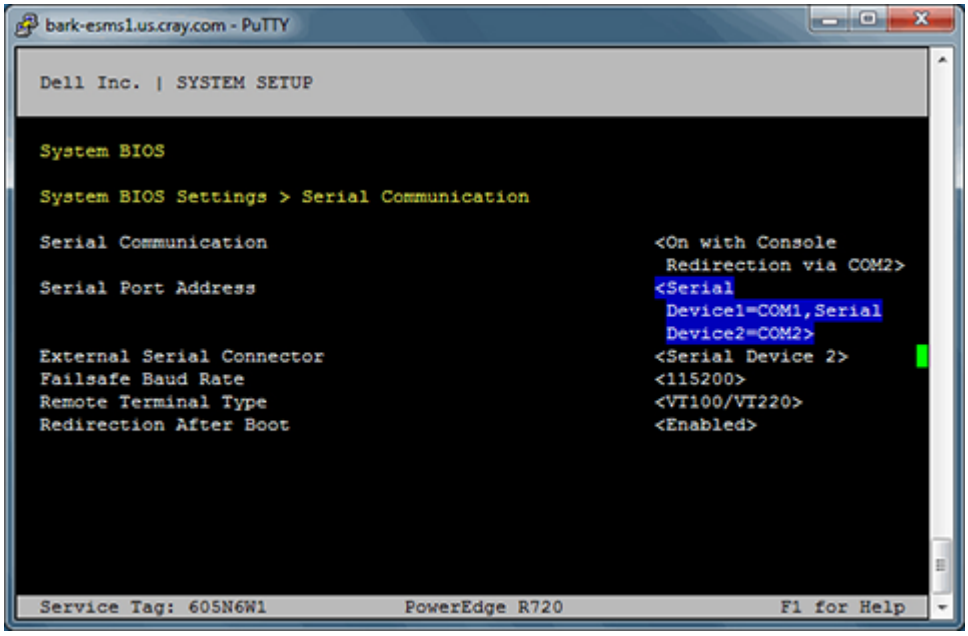
A pop-up window displays the available options.

c. Select **On with Console Redirection via COM2**, then press **Enter**.

d. Verify that **Serial Port Address** is set to **Serial Device1=COM1**, **Serial Device2=COM2**. This setting enables the remote console. If this setting is incorrect, remote console to access to the CLFS node. If necessary, press **Enter** to display the available options.

e. Change the setting to **Serial Device1=COM1, Serial Device2=COM2**.

Figure 5. Dell R720 BIOS Serial Communication Settings



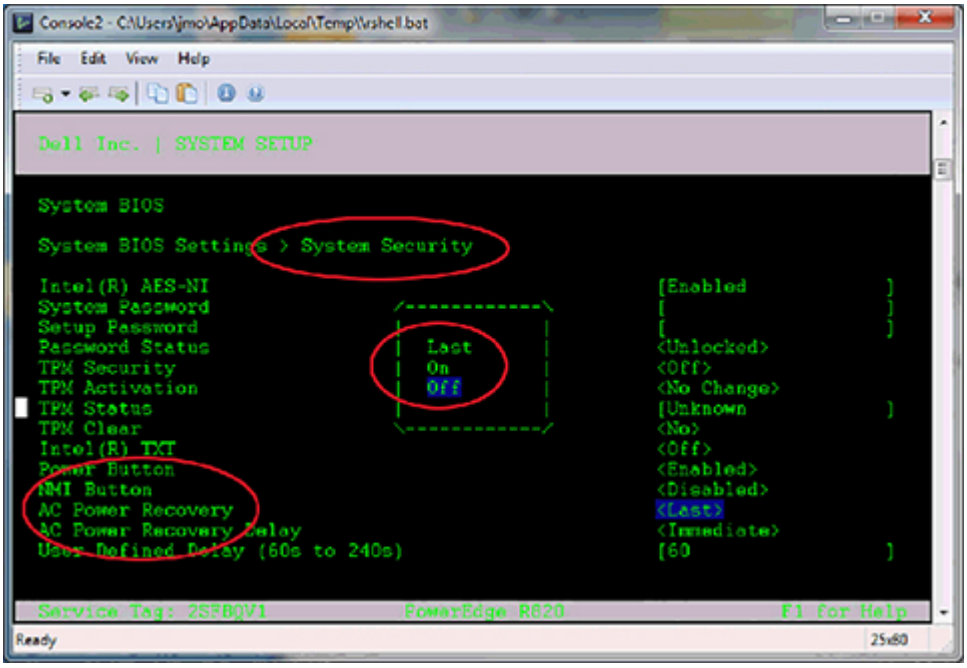
5. Press **Enter** to return to the **Serial Communication** screen.
6. Select **External Serial Connector**.

   A pop-up window displays the available options.
7. In the pop-up window, select **Remote Access Device**, then press **Enter** to return to the previous screen.
8. Select **Failsafe Baud Rate**.
9. In the pop-up window, select **115200**, then press **Enter** to return to the previous screen.
10. Press **Esc** to exit the **Serial Communication** screen.
11. Press the **Esc** key to exit the **System BIOS Settings** screen.
12. Press **Esc** to exit to the **BIOS Settings** screen.

    The `Settings have changed` message displays.
13. Select **Yes** to save your changes.

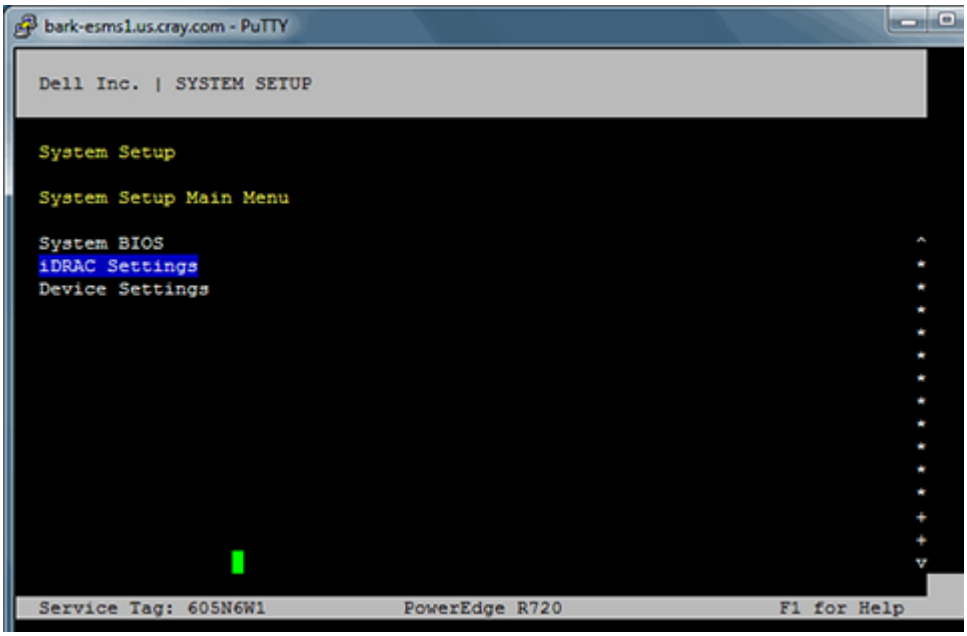
    The `Settings saved successfully` message displays.
14. Select **OK**.
15. From the **System BIOS** menu, select **System BIOS Settings**, then **System Security**, then **AC Power Recovery**. Set **AC Power Recovery** to **Off** to configure the CLFS node to remain powered off after a system power failure. Cray recommends that the CIMS node power on and become operational before service nodes power on.

Figure 6. Dell R720 BIOS AC Power Recovery Settings



16. Press Esc until you return to the **System Setup Main Menu** and select **iDRAC Settings**, then press **Enter**.

Figure 7. Dell R720 BIOS iDRAC Settings



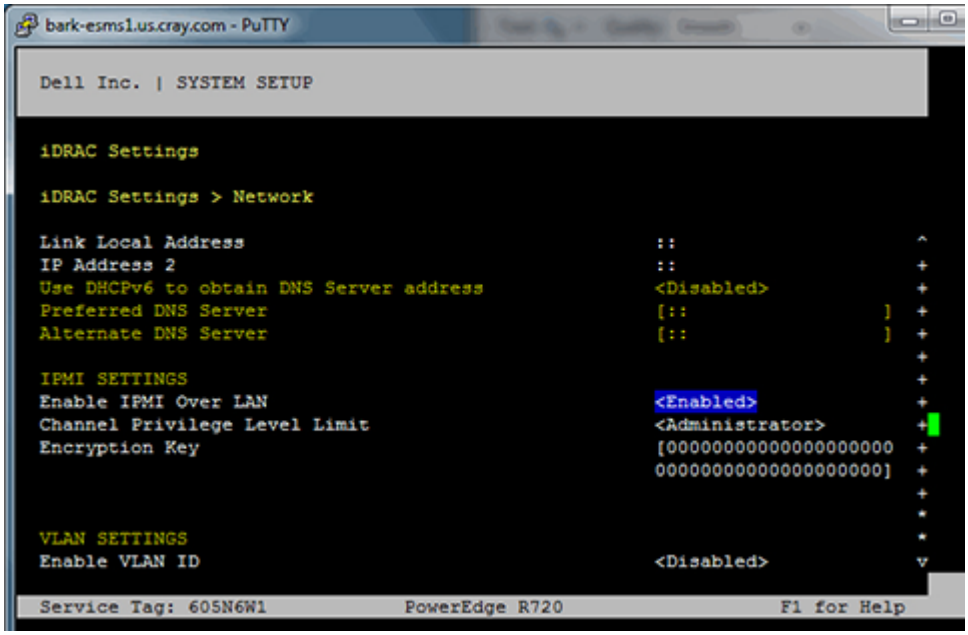
17. Select **Network**, then press **Enter**.

18. Change the IPMI settings to enable the Serial Over LAN (SOL) console.

    a. Use the down-arrow key to scroll to the **IPMI SETTINGS** list.

b. Select **Enable IPMI over LAN**, then press **Enter**.

Figure 8. Dell R720 BIOS iDRAC IPMI Over LAN Setting



c. In the pop-up window, select **<Enabled>**.

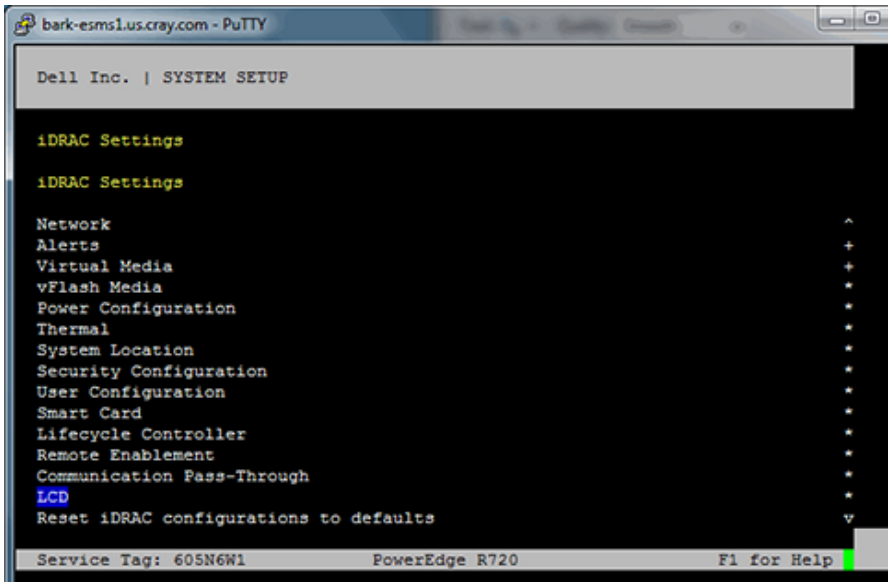
d. Press **Enter** to return to the previous screen.

19. Press **Esc** to exit the **Network** screen and return to the **iDRAC Settings** menu.

20. Change the LCD configuration to show the hostname in the LCD display.

a. On the **iDRAC Settings** screen, use the down-arrow key to highlight **LCD**, then press **Enter**.

Figure 9. Dell R720 BIOS iDRAC LCD Setting



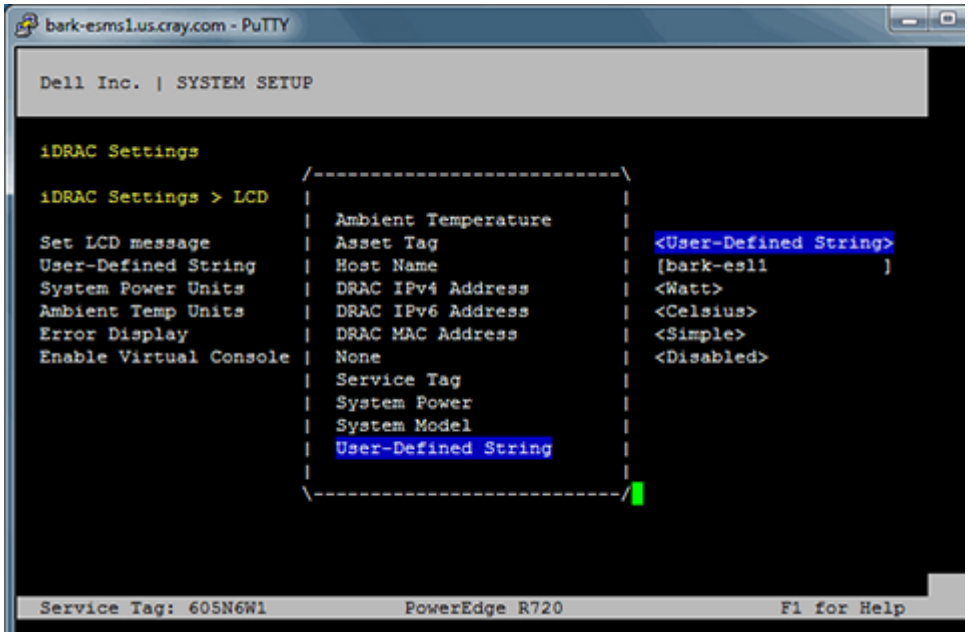
b. Select **Set LCD message**.

   A pop-up window opens.

c. In the pop-up window, select **User-Defined String**, then press **Enter**.

d. Select **User-Defined String** (again), then press **Enter**.

Figure 10. Dell R720 BIOS iDRAC LCD User-Defined String Setting



   A text pop-up window opens for entering a new string.

e. In the text pop-up window, enter the CLFS hostname (such as **clfs-mdc001**), then press **Enter**.

f. Press the **Esc** to exit the LCD screen.

g. Press **Esc** again to exit the **Network** screen.

h. Press **Esc** again to exit the **iDRAC Settings** screen.

i. The `Settings have changed` message displays. Select **Yes**, then press **Enter** to save the changes.

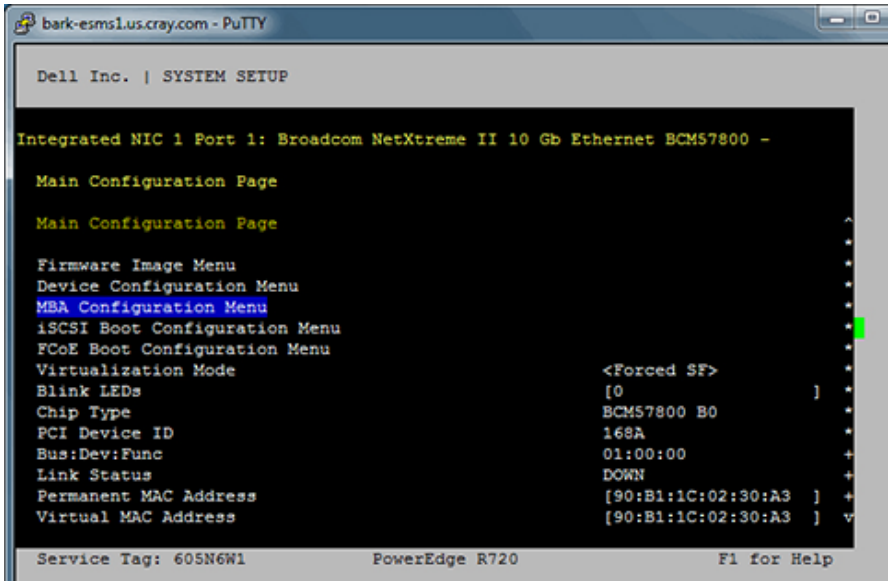
j. The `Settings saved successfully` message displays. Select **OK**, then press **Enter**.

21. Change the device settings so that the CLFS node can PXE boot on the CIMS administration network (`esmaint-net`).

a. On the **System Setup Main Menu**, select **Device Settings**, then press **Enter**.

b. In the **Device Settings** window, select **Integrated NIC 1 Port N ...**, then press **Enter**. The **Main Configuration Page** opens.

Figure 11. Dell R720 BIOS Integrated NIC Port Settings



Choose the NIC port number that corresponds to the Ethernet port for the `esmaint-net` network. If the `esmaint-net` network uses the first Ethernet port (`eth0`), select **Integrated NIC 1 Port 1 ....**. If the `esmaint-net` network uses the third Ethernet port (`eth2`), select **Integrated NIC 1 Port 3 ....**.

NOTE:

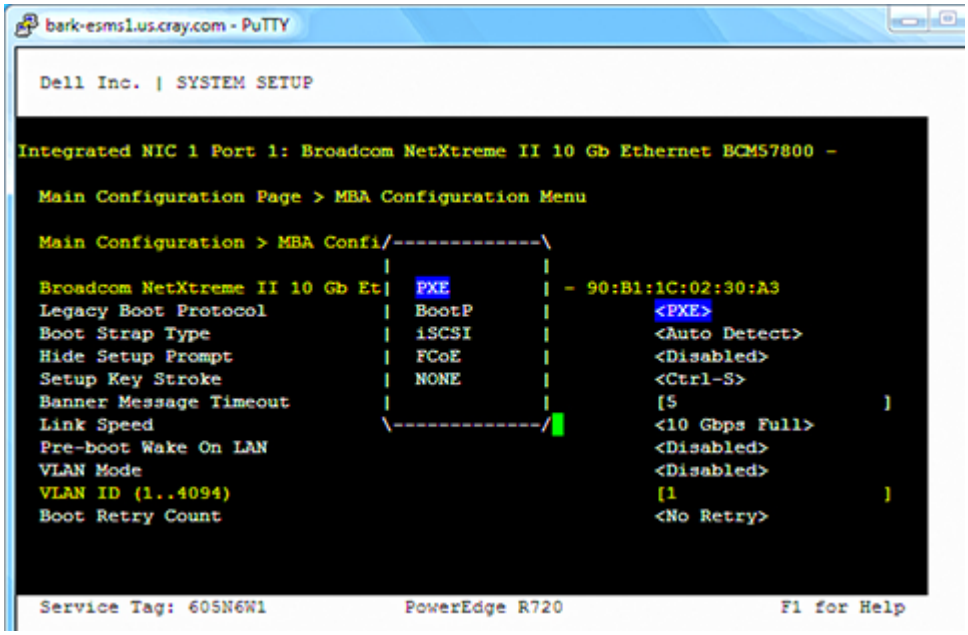
Disable PXE booting on the other three Ethernet ports.

22. On the **Main Configuration Page** screen, select **MBA Configuration Menu**, then press Enter.

a. On the **MBA Configuration Menu** screen, select **Legacy Boot Protocol**, then press **Enter**. A pop-up window displays the available options.

b. In the pop-up window, use the down-arrow key to select **PXE**, then press **Enter**.

Figure 12. Dell R720 BIOS Legacy Boot Protocol Settings



c. Press **Esc** to exit the **MBA Configuration Menu** screen.

d. Press **Esc** to exit the **Main Configuration Page** screen.

e. Verify that Legacy Boot Protocol is set to **None** for the other Ethernet ports.

f. If necessary, return to *22.a* on page 19 to change the settings for each port.

g. Press the **Esc** key to exit the **Device Settings** screen.

h. A "Settings have changed" message displays. Select Yes, then press Enter to save your changes.

i. A `Settings saved successfully` message displays. Select **OK**, then press **Enter**. The main screen (**System Setup Main Menu**) displays.

23. Press **Esc** to exit the **System Setup Main Menu**.

The utility displays the prompt `Are you sure you want to exit and reboot?`

24. Select **Yes** at the prompt.

4 Create a Managed CLFS Node in Bright

Use following procedure as a checklist to create a managed CLFS node. Detailed procedures for each step below are included in this document:

1. Create a managed CLFS node in Bright by cloning the default service node (`node001`) and configuring the network interfaces and settings.
2. Create the `site-user-network` or other required networks for the node.
3. Test boot (power cycle) the node so that it boots from the `default-image` and configures the iDRAC.
4. Install ESF XX-2.3.0 software to create a CLFS software image on the CIMS.
5. Configure a generic CLFS category (`esfs-generic`) in Bright.
6. Configure the finalize script (`esf_finalize.sh`) for the `esfs-generic` category.
7. Configure kdump on the `esfs-generic` category.
8. Assign the ESF XX-2.3.0 software image to the `esfs-generic` category.
9. Examine the default exclude lists and customize exclude lists for your site.
10. Test boot the CLFS node from the ESF image and `esfs-generic` category to verify the configuration.
11. After the `esfs-generic` category is customized and tested, clone `esfs-generic` to the categories required by esfsmon 2.1.0.
12. Assign even and odd CLFS nodes to the appropriate esfsmon 2.1.0 categories.
13. Perform other configuration procedures for each category as needed for the site.

    IMPORTANT: Cray strongly recommends that newly defined service nodes are power cycled (test-booted) from the default service node image (`default-image`). A boot cycle from `default-image` configures the iDRAC port, verifies node BIOS settings are set correctly, and verify Bright configuration. Resolve boot failures before you proceed.

    If the generic CLFS node boots successfully, install the CLFS software and configure the `esfs-generic` category.

    If your site has multiple CLFS nodes, perform this procedure first on a single generic node. After creating the full image and configuring the Bright settings for the generic node, you will return to this procedure to create the remaining CLFS nodes. Clone the generic node, and customize the node to be either an odd or even MDS or OSS node, then assign the node to the appropriate category in Bright.

4.1 Create a Generic CLFS Node in Bright

Prerequisites

The BIOS for a managed CLFS node must be configured before performing this procedure.

About this task

To create a CLFS node in Bright Cluster Manager® (Bright), clone an existing CLFS node that is fully configured, or the clone the default node (`node001`) and configure the node settings.

> **IMPORTANT:** Beginning with ESM release ESM XX-3.0.0, MDS node names must contain the string `mds`, and OSS node names must contain the string `oss` so that a regular expresssion in the finalize script (`site.esf_finalize.sh`) script can configure each node type.

If a functioning CLFS node does not exist, clone the default node (`node001`) created during the CIMS node installation process. CLFS nodes can then be specialized to function in a Data Management Practices (DMP) system as metadata server (MDS) or object storage server (OSS) nodes, or Cray Cray Tiered Adaptive Storage (TAS) systems as metadata controller (MDC) or data mover (DM) nodes.

Procedure

1. Log in to the CIMS node as `root` and run `cmsh`.

```
cims1#  cmsh
[cims1]%
```

2. Enter `device` mode.

```
[cims1]% device
[cims1->device]%
```

3. List the available devices.

```
[cims1->device]%  list
Type                 Hostname (key)   MAC                Category   Ip              Network
-------------------- ---------------- ------------------ ---------- --------------- --------------
EthernetSwitch       switch01         00:00:00:00:00:00             10.141.253.1    esmaint-net
MasterNode           esms1            78:2B:CB:40:CE:CA             10.141.255.254  esmaint-net
PhysicalNode         node001          00:00:00:00:00:00  default    10.141.0.1      esmaint-net
```

4. > **IMPORTANT:** Always make sure that newly cloned nodes boot from default software image (named `default-image`) without errors before attempting other configuration tasks. This configures the iDRAC port and verifies the managed node can boot successfully from Bright.

   Clone a node to create a new CLFS node.

   - If a fully configured CLFS node exists, clone the configured node to create another CLFS node and assign it to the `default` category.

```
[cims1->device]%  clone clfs_node clfs-mdc001
Base name mismatch, IP settings will not be modified!
[cims1->device*[clfs-mdc001*]]%  commit
[cims1->device[clfs-mdc001]]%  exit
[cims1->device]%  set clfs-mdc001 category default
```

   - If a fully configured CLFS does not exist, clone the default node (`node001`) to create a generic CLFS node and assign it to the `default` category.

```
[cims1->device]%  clone node001 clfs-mdc001
Base name mismatch, IP settings will not be modified!
[cims1->device*[clfs-mdc001*]]%  commit
[cims1->device[clfs-mdc001]]%  exit
cims1->device]%  set clfs-mdc001 category default
```

5. Commit the changes.

```
[cims1->device*]%  commit
[cims1->device]%
```

6. Change the interface settings for the new node. Select the new node (`use`) and switch to `interfaces` mode to list the network interfaces on `clfs-mdc001`.

```
[cims1->device]%  use clfs-mdc001
[cims1->device[clfs-mdc001]]% interfaces
[cims1->device[clfs-mdc001]->interfaces]%  list
Type         Network device name  IP               Network
------------ -------------------- ---------------- ----------------
bmc          ipmi0                10.148.0.1       ipmi-net
physical     BOOTIF [prov]        10.141.0.1       esmaint-net
```

7. Set the `BOOTIF` and `ipmi0` interface addresses for the new node. These addresses must be different from those used by the default or original node.

```
[cims1->device*[clfs-mdc001*]->interfaces]% set bootif ip 10.141.0.2
[cims1->device*[clfs-mdc001*]->interfaces*]% set ipmi0 ip 10.148.0.2
[cims1->device*[clfs-mdc001*]->interfaces*]% list
Type         Network device name  IP               Network
------------ -------------------- ---------------- ----------------
bmc          ipmi0                10.148.0.2       ipmi-net
physical     BOOTIF [prov]        10.141.0.2       esmaint-net
```

8. Configure the `ib-net` network and interface. Cray recommends that `ib0` on OSS or MDS nodes connect to the IB switch, `ib1` should not be used. Connect `ib2` and `ib3` to the storage array controllers.

```
[cims1->device*[clfs-mdc001*]->interfaces]% add physical ib0
[cims1->device*[clfs-mdc001*]->interfaces*[ib0*]]% set network ib-net
[cims1->device*[clfs-mdc001*]->interfaces*[ib0*]]% set ip 10.149.0.2
[cims1->device*[clfs-mdc001*]->interfaces*[ib0*]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Additional Hostnames
Card Type
DHCP                           no
IP                             10.149.0.2
MAC                            00:00:00:00:00:00
Network                        ib-net
Network device name            ib0
Revision
Speed
Type                           physical
```

9. Commit the changes.

```
[cims1->device*[clfs-mdc001*]->interfaces*[ib0*]]% commit
[cims1->device[clfs-mdc001]->interfaces[ib0]]%
```

10. Exit `ib0` mode.

```
[cims1->device[clfs-mdc001]->interfaces[ib0]]% exit
```

11. Check the `ib0` interface settings

```
[cims1->device[clfs-mdc001]->interfaces]% list
Type         Network device name  IP               Network
------------ -------------------- ---------------- ----------------
bmc          ipmi0                10.148.0.2       ipmi-net
physical     BOOTIF [prov]        10.141.0.2       esmaint-net
physical     ib0                  10.149.0.2       ib-net
```

12. Exit `interface` mode and return to `device` mode.

```
[cims1->device[clfs-mdc001]->interfaces]% exit
[cims1->device[clfs-mdc001]]%
```

13. Display the status of the new node.

```
[cims1->device[clfs-mdc001]]% status
cims1  ............... [  DOWN  ]  (Unassigned)
```

The node is unassigned because the MAC address has not been set in Bright.

14. Set the MAC address for the node (`eth0` on `esmaint-net`).

```
[cims1->device[clfs-mdc001]]% set mac MACaddress
cims1  ............... [  DOWN  ]  (Unassigned)
[cims1->device*[clfs-mdc001*]]%
```

The node status is `Unassigned` because the MAC address has not been set in Bright.

15. Set the management network to `esmaint-net`.

```
[cims1->device*[clfs-mdc001*]]% set managementnetwork esmaint-net
```

16. Set the rack information (device height, position in the rack, and rack index). This example assumes that the device height is 2U, its position in the rack is 0, and the rack index is 0. Substitute the actual values for the node.

```
[cims1->device*[clfs-mdc001*]]% set deviceheight 2
[cims1->device*[clfs-mdc001*]]% set deviceposition 0
[cims1->device*[clfs-mdc001*]]% set rack 1
```

17. Commit the changes and quit `cmsh`.

```
[cims1->device*[clfs-mdc001*]]% commit
[cims1->device*[clfs-mdc001]]% quit
cims1#
```

18. If this node is connected to a site user network (`site-user-net`), configure the network parameters for the `site-user-net` in Bright. Power cycle the node to boot it from the `default-image` software image to configure the iDRAC and verify that Bright can sucessfully boot the node.

The next task after configuring a generic CLFS node in Bright is typically to *Create site-user-net Network in Bright* on page 24 if users will need network access to this node.

4.2 Create `site-user-net` Network in Bright

Prerequisites

A fully configured CIMS node running Bright software and managed service nodes.

About this task

The external `site-user-net` network is created on the `eth1` interface if the Data Management Practices (DMP) service node requires access to a site LDAP or authentication service (in some CLFS nodes), or supports user logins for CDL nodes. User accounts on external (CDL nodes) enable users to access a Cray shared file system for software development. There are typically no user login accounts on CLFS nodes.

Procedure

1. Log in to the CIMS node and start `cmsh`.

```
cims1# cmsh
[cims1]%
```

2. Switch to `network` mode:

```
[cims1]% network
[cims1->network]%
```

3. Display the existing networks.

```
[cims1->network]% list
Name (key)       Type         Netmask bits     Base address     Domain name          IPv6
---------------- ------------ ---------------- ---------------- -------------------  ----
esmaint-net      Internal     16               10.141.0.0       esmaint-net.cluster  no
globalnet        Global       16               0.0.0.0          cm.cluster           no
ib-net           Internal     16               10.149.0.0       ib-net.cluster       no
ipmi-net         Internal     16               10.148.0.0       ipmi-net.cluster     no
site-admin-net   External     24               aaa.bbb.ccc.ddd  your.domain.com        no
```

4. Determine whether `site-user-net` exists on the CIMS node.

   a. If a `site-user-net` already exists on the CIMS node, proceed to step *5* on page 26.

   b. If there is no site user network, create `site-user-net` by cloning the `site-admin-net` network.

```
[cims1->network]% clone site-admin-net site-user-net
[cims1->network*[site-user-net*]]%
```

   c. The cloned network inherits the same settings as the original network (`site-admin-net`). You must change several settings for the new `site-user-net`: base address, broadcast address, domain name, gateway, and (if necessary) netmask bits. Display the existing settings.

```
[cims1->network*[site-user-net*]% show
Parameter                      Value
------------------------------ ----------------------
Base address                   aaa.bbb.ccc.ddd
Broadcast address              aaa.bbb.255.255
Domain Name                    your.domain.com
Dynamic range end              0.0.0.0
Dynamic range start            0.0.0.0
Gateway                        aaa.bbb.ccc.ddd
IPv6                           no
Lock down dhcpd                no
MTU                            1500
Management allowed             no
Netmask bits                   24
Node booting                   no
Notes                          <0 bytes>
Revision
Type                           External
name                           site-user-net
```

   d. Change the base address.

```
[cims1->network*[site-user-net*]% set baseaddress site-user-netBaseAddress
```

   e. Change the broadcast address.

```
[cims1->network*[site-user-net*]% set broadcastaddress site-user-netBroadcastAddress
```

   f. Change the domain name.

```
[cims1->network*[site-user-net*]% set domainname site-user-netDomainName
```

   g. Change the gateway.

```
[cims1->network*[site-user-net*]% set gateway site-user-netGateway
```

h. If necessary, change the netmask bits.

```
[cims1->network*[site-user-net*]% set netmaskbits NN
```

i. Commit your changes.

```
[cims1->network*[site-user-net*]% commit
```

5. Switch to `device` mode.

```
[cims1->network[site-user-net]% device
```

6. Add an interface to the `site-user-net` network. This example shows the hostname `servicenode2`, the Ethernet port `eth1`, and the example IP address *`aaa.bbb.ccc.ddd`*. Substitute the CLFS node's hostname and IP address when configuring the `eth1` interface. Repeat this step for each service node that is added to the `site-user-net` network.

```
[cims1->device]% addinterface -n servicenode2 physical eth1 site-user-net aaa.bbb.ccc.ddd
```

7. Commit the changes.

```
[cims1->device*]% commit
```

8. Show the interfaces on `servicenode2`.

```
[cims1->device]%  use servicenode2
[cims1->device[servicenode2]% interfaces; list
Type          Network device name  IP              Network
------------ -------------------- ---------------- -----------------
bmc           ipmi0                10.148.0.2       ipmi-net
physical      BOOTIF [prov]        10.141.0.2       esmaint-net
physical      eth1                 aaa.bbb.ccc.ddd  site-user-net
physical      ib0                  10.149.0.2       ib-net
```

9. Display the existing networks.

```
[cims1->device[servicenode2]->interfaces]% network list
Name (key)       Type         Netmask bits     Base address     Domain name          IPv6
---------------- ------------ ---------------- ---------------- -------------------- ----
esmaint-net      Internal     16               10.141.0.0       esmaint-net.cluster  no
globalnet        Global       16               0.0.0.0          cm.cluster           no
ib-net           Internal     16               10.149.0.0       ib-net.cluster       no
ipmi-net         Internal     16               10.148.0.0       ipmi-net.cluster     no
site-admin-net   External     24               aaa.bbb.0.0      your.domain.com         no
site-user-net    External     24               aaa.bbb.0.0      your.domain.com         no
```

10. Exit interface mode `cmsh`.

```
[cims1->device[servicenode2]->interfaces]% exit
[cims1->device[servicenode2]
```

4.3 Test Boot a Service Node from the Default Software Image

Prerequisites

Before you can boot the service node, you must have set up the BIOS and configured the node in Bright so that it boots from the `default-image` software image.

About this task

Always test boot newly configured service nodes from the `default-image` software image to configure the iDRAC port and configure the node properly in Bright.

Procedure

1. Connect a console keyboard, monitor, and mouse to the front panel USB connections on the service node.
2. Manually power cycle the service node to boot it for the first time.
3. **IMPORTANT:** Newly defined service nodes must boot from the `default-image` software image. A manual power cycle (reboot) using the `default-image` image configures the iDRAC port to enable remote access, verifies the service node BIOS settings, and verifies the node settings in Bright are configured properly. Verify that there are no failures during the test boot before you proceed.

   If the node boots successfully, continue configuration tasks for the service node. Otherwise, carefully review the BIOS and Bright configuration settings.

4.4 Install ESF Software on a Managed CLFS Node

Prerequisites

Configure a generic CLFS node in Bright (clone the default node (`node001`), configure network interfaces, and boot the CLFS node from the `default-image` software image to configure the node settings and iDRAC port before installing ESF software.

About this task

Use this procedure for initial ESF software installations only, and creates a CLFS node software image on the CIMS node `/cm/images` directory. The ESF installation software creates two default categories for CLFS nodes on the CIMS node, `esFS-MDS`, and `esFS-OSS` if they do not exist.

Procedure

1. Login to the CIMS node as `root`.
2. If necessary, create a `/media/cdrom` directory.

   ```
   cims1# mkdir -p /media/cdrom
   ```

3. Download or copy the CentOS 6.6 software ISO file (`CentOS-6.6-x86_64-bin-DVD1.iso`) to `/root/isos` on the CIMS node. This ISO file is required for the software installation process.

   a. Insert the DVD into the DVD drive and mount it to `/media/cdrom`.

   ```
   cims1# mount /dev/cdrom /media/cdrom
   ```

   b. Copy the contents of the CentOS 6.6 DVD to `/root/isos` on the CIMS node.

   ```
   cims1# dd if=/dev/cdrom of=/root/isos/CentOS-6.6-x86_64-bin-DVD1.iso bs=1M
   ```

4. Download or copy the Bright and CentOS software ISO file (`bright6.1-centos6u6.iso`) to `/root/isos`. This ISO file is required for ESF software installation.

   a. Insert the DVD into the DVD drive and mount it to `/media/cdrom`.

   ```
   cims1# mount /dev/cdrom /media/cdrom
   ```

   b. Copy the contents of the Bright and CentOS DVD to `/root/isos` on the CIMS node.

   ```
   cims1# dd if=/dev/cdrom of=/root/isos/bright6.1-centos6u6.iso bs=1M
   ```

5. Mount the ESF release media on the CIMS node.

   - If you have the release media on DVD, insert the ESF software DVD into the DVD drive and mount it to `/media/cdrom`.

   ```
   cims1# mount /dev/cdrom /media/cdrom
   ```

   - If you have an ISO file of the release media, mount the ESF ISO to `/media/cdrom`. The ISO file name depends on the supported architecture, release number, and installer version. This procedure shows the example ISO name `ESF-XX-2.3.0-201509031326a06.iso`, and assumes that the ISO file is in the `/root/isos/` directory.

   ```
   cims1# mount -o loop,ro /root/isos/ESF-XX-2.3.0-201509031326a06.iso /media/cdrom
   ```

6. Copy all files from the ESF release media to the CIMS node. Cray recommends storing the release files in a new subdirectory that uniquely identifies the release. The following examples use the directory `/root/release/ESF-XX-2.3.0-201509031326a06.iso`.

   ```
   cims1# mkdir -p /root/release/ESF-XX-2.3.0-201509031326
   cims1# cp -pr /media/cdrom/* /root/release/ESF-XX-2.3.0-201509031326
   ```

7. Unmount the ESF release media. If using a DVD, eject the DVD.

   ```
   cims1# umount /media/cdrom
   cims1# eject
   ```

8. Run the `ESFinstall` script to install ESF software on the CIMS node `/cm/images` directory.

   ```
   cims1# cd /root/release/ESF-XX-2.3.0-201509031326

   cims1# ./ESFinstall -v
   ./ESFinstall: exec to temp script: /tmp/ESFinstall.PID
   Installation output will be captured in /var/adm/cray/logs/ESFinstall.PID.log.ESF-XX-2.3.0-201509031326
   Tue July 3 15:56:48 CST 2015: ESFinstall.PID Starting, Version ESF-XX-2.3.0-201509031326-a03.
   ```

   By default, `ESFinstall` uses the release version string for the image directory name and creates the image in `/cm/images`. For example, the previous command creates the image in `/cm/images/ESF-XX-2.3.0-201509031326`. To specify a different image directory name, use the `-s` *`imagename`* option, where *`imagename`* is the full path to the ESF software image file. For more information, run `ESFinstall -h` to see the usage statement.

9. While the installation runs, examine the initial output and note the process ID (PID) of the `ESFinstall` process. `ESFinstall` creates a log file in `/var/adm/cray/logs/ESFinstall.`*`PID`*`.log.`*`imagename`* using the PID.

10. When the installation completes, locate the new CLFS node software image in `/cm/images` and note the directory (image) name (`ESF-XX-2.3.0-201509031326` in this example). You will use this directory in the following sections to customize the software image. You will also use the directory name as the name of the image when configuring the Bright category for CLFS nodes.

```
cims1# ls /cm/images
ESF-XX-2.3.0-201509031326  ESL-XC-1.0.2-201302211318  ESL-XE-1.1.1-kdump   default-image
ESL-TEST-4102013           ESL-XE-1.1.1-201211150916  ESL-XE-1.1.1_CLE4.1  default-image.previous
```

11. When creating, adding, or modifying a a software image in Bright, enter relevant information in the `Notes` property for the image using either the `cmsh` or `cmgui`. Use `cmsh softwareimage` mode `set notes` command to launch a `vim` editor session to enter relevant information about the software image. Enter `:help` in the editor to display help text.

```
cims1# cmsh
[cims1]% softwareimage
[cims1->softwareimage]% list
Name (key)                          Path
----------------------------------- ----------------------------------------
ESF-XX-2.3.0-201509031326           /cm/images/
default-image                       /cm/images/default-image

[cims1->softwareimage]% use ESF-XX-2.3.0-201509031326
[cims1->softwareimage[ESF-XX-2.3.0-201509031326]]% set notes
(in vim editor) created new ESF-XX-2.3.0-201509031326 image - Jane Doe, 6-15-2015.
[cims1->softwareimage*[ESF-XX-2.3.0-201509031326*]]% commit
[cims1->softwareimage[cims1]]% quit
```

12. Verify the `lustre` service is running for the ESF software image. Use the `chroot` shell to open a shell in the ESF software image and verify the `lustre` service is enabled.

```
cims1# chroot /cm/images/cims1
[root@cims1 /]# chkconfig --list lustre
lustre 0:off 1:off 2:on 3:on 4:on 5:on 6:off
```

- If the lustre service is enabled, exit the `chroot` shell:

```
[root@cims1 /]# exit
cims1#
```

- If the `lustre` service is not started, start the `lustre` service and exit the `chroot` shell.

```
[root@cims1 /]# chkconfig lustre on
[root@cims1 /]# exit
```

13. Optional: If the CLFS node supports Fibre Channel host bus adapters, set the software image kernel parameter `rdloaddriver` to `scsi_dh_rdac`.

```
cims1# cmsh
[cims1]% softwareimage
[cims1->softwareimage]% use
[cims1->softwareimage[]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Boot FSPart                    98784247966
Creation time                  Thu, 23 Jan 2015 12:43:46 CST
Enable SOL                     yes
FSPart                         98784247966
Kernel modules                 <37 in submode>
Kernel parameters
Kernel version                 2.6.32-358.18.1.el6.x86_64
Locked                         no
Name
Notes                          <0 bytes>
Path                           /cm/images/
Revision
SOL Flow Control               no
SOL Port                       ttyS1
SOL Speed                      115200
```

```
[cims1->softwareimage[]]% set kernelparameters rdloaddriver=scsi_dh_rdac
[cims1->softwareimage*[*]]% commit
```

```
[cims1->softwareimage[]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Boot FSPart                    98784247966
Creation time                  Thu, 23 Jan 2015 12:43:46 CST
Enable SOL                     yes
FSPart                         98784247966
Kernel modules                 <37 in submode>
Kernel parameters              rdloaddriver=scsi_dh_rdac
Kernel version                 2.6.32-358.18.1.el6.x86_64
Locked                         no
Name
Notes                          <0 bytes>
Path                           /cm/images/
Revision
SOL Flow Control               no
SOL Port                       ttyS1
SOL Speed                      115200
```

14. Exit `cmsh`.

```
[cims1->softwareimage]% quit
```

5 Configure a CLFS Node in Bright

The ESF software installation script (`ESFinstall`) creates an `esFS-MDS` category and an `esFS-OSS` category in Bright. The `esFS-MDS` category is cloned and then configured into a customized `esfs-generic` category for the site. After all configuration tasks are complete for the `esfs-generic` category, that category is cloned to the following Bright categories to support the high-availability features in esfsmon 2.1.0: `esfs-even-`*`filesystem`*, `esfs-odd_`*`filesystem`*, and `esfs-failed-`*`filesystem`*. Each CLFS node is then placed into the appropriate category based on node hostname (`device` name in Bright) to support esfsmon 2.1.0.

Bright categories associate a specific software image, finalize script, disk drive configuration, and other information with a specific node type. The ESF software image (for example, `ESF-XX-2.3.0-201509031326` is assigned to each CLFS category, yet each node category can have different configuration settings.

The `ESM XX-3.0.0` release provides a single finalize script, `site.esf_finalize.sh`, that configures both MDS and OSS nodes. The node names (`$HOSTNAME`) must contain the string `mds` or `oss` so that the `site.esf_finalize.sh` script can configure both node types. Also, CLFS node names must be numbered so that they can be placed in either an odd or even category for esfsmon 2.1.0.

5.1 Configure a Generic CLFS Category in Bright

Prerequisites

The CLFS BIOS is configured, network interfaces for `esmaint-net` and `site-user-net` are configured in Bright, the node has been power cycled to boot from the `default-image` software image in Bright, and an ESF software (for example `ESF-XX-2.3.0-201509031326`) is installed in `/cm/images` on the CIMS node.

About this task

This procedure clones the default `esFS-MDS` category to `esfs-generic`, then customizes the `esfs-generic` category network settings, default disk partitions, exclude lists, finalize script (`site.esf_finalize.sh`), kdump, etc.

Procedure

1. Log in to the CIMS node and start `cmsh`.

```
cims1# cmsh
[cims1]%
```

2. Switch to `category` mode and list categories and associated software images.

```
[esms1]% category
[esms1->category]% list
Name (key)               Software image
```

```
------------------------ ------------------------
default                  default-image
default-diskless         default-image
esFS-MDS                 default-image
esFS-OSS                 default-image
esLogin-XC               ESL-XC-1.0.2-2013022113+
esLogin-XE               ESL-XE-1.1.1_CLE4.1
```

3. Clone the `esFS-MDS` category to `esfs-generic` and commit the change.

```
[cims1->category]% clone esFS-MDS esfs-generic
[cims1->category*]% commit
[cims1->category]%
```

4. Assign the CLFS node software image (ESF-XX-2.3.0-201509031326) to the `esfs-generic` category in Bright.

```
[cims1->category]% use esfs-generic
[cims1->category[esfs-generic]]% set softwareimage ESF-XX-2.3.0-201509031326
[cims1->category*[esfs-generic*]]%
```

5. If required, set the default gateway so that service nodes can contact an external LDAP server on `site-user-net`. For gateway, use the IP address of your site's gateway (on `site-user-net`).

```
[cims1->category*[esfs-generic*]]% set defaultgateway gateway
```

6. Commit the changes.

```
[cims1->category*[esfs-generic*]]% commit
[cims1->category*[esfs-generic]]%
```

7. Configure the `esfs-generic` category with the proper disk partition sizes and configuration. A custom disk partition layout can be applied using an XML schema to a category of nodes. A disk partition layout that is applied to an individual node within a category overrides the category setting. Add `<blockdev>` XML entries to the `site.esfs-diskfull.xml` file if there are several LUNs on a SAS RAID that are used for MDTs. The XML file currently defines `/dev/sda` through `/dev/sdz`. Changes made to the `esfs-diskfull.xml` file do not occur until they are saved in `cmgui`, or committed in `cmsh`, and the node is rebooted. Bright detects that the partitioning has changed and invokes a FULL install. (Bright also ignores the NOSYNC install mode and will repartition the drives on other nodes in the category as well.) Older model CLFS nodes use `esfs-small-diskfull.xml` for disk partitioning to accommodate smaller hard drives.

   IMPORTANT: If the default disk setup XML files are updated in a ESM release or if the site disk setup XML files have been customized, system administrators must compare the newly released disk setup XML files with the current production disk setup XML files, and merge the changes manually. After the changes have been merged, you must load the updated disk setup file into the Bright database for the node category and reboot all the nodes that use that category.

   a. Quit `cmsh` and copy the default XML configuration to the `etc` directory to prevent it from being overwritten during software updates.

```
[cims1->category[esfs-generic]]% quit
cims1# cd /opt/cray/esms/cray-es-diskpartitions-XX
cims1# mkdir -p etc
cims1# cp -p default/esfs-diskfull.xml etc/site.esfs-diskfull.xml
```

   b. Edit the `site.esfs-diskfull.xml` file to change partition sizes or configuration.

```
cims1# vi etc/site.esfs-diskfull.xml
```

   c. Save the changes and return to `cmsh`.

8. Set the disk setup parameter for the `esfs-generic` category and commit the changes.

```
cims1# cmsh
[cims1]% category use esfs-generic
[cims1->category[esfs-generic]]% set disksetup /opt/cray/esms/cray-es-diskpartitions-XX\
/etc/site.esfs-diskfull.xml
[cims1->category*[esfs-generic*]]% commit
```

9. Confirm your changes to the `site.esfs-diskfull.xml` script.

```
[cims1->category[esfs-generic]]% get disksetup
<?xml version="1.0" encoding="UTF-8"?><diskSetup xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
<device>
<blockdev>/dev/sdablockdev>/dev/sda></blockdev>
<blockdev>/dev/sdablockdev>/dev/sdb></blockdev>
<blockdev>/dev/sdablockdev>/dev/sdc></blockdev>
...
```

5.2 Node Finalize Scripts for CLFS Nodes

The node finalize script is configured for the default `esFS-MDS` category during ESF software installation. The default `esFS-MDS` category is cloned to a generic CLFS category `esfs-generic` that is further customized and cloned into the required esfsmon 2.1.0 categories.

The finalize script runs before `init` and is used to set a file configuration or to initialize special hardware, sometimes after a hardware check. It is run in order to make software or hardware work before, or during the later `init` stage of boot. Use a finalize script to execute commands before `init`, when the commands cannot be stored persistently anywhere else, or when it is needed because a choice between (otherwise non-persistent) configuration files must be made based on the hardware before `init` starts.

A single node finalize script, `site.esf_finalize.sh`, is configured to support both MDS or OSS nodes beginning with release ESM XX-3.0.0. The `site.esf_finalize.sh` script differentiates between MDS and OSS nodes by checking the node name (*`$HOSTNAME`*), which must include either the string `mds` or `oss`.

> **IMPORTANT:** Files created or modified by a finalize script must be listed in the `excludelistupdate` exclude list for the category. Software updates will overwrite customized files if the files are not specified in an exclude list for the category. Customized files must also be specified in the `excludelistgrab`, and `excludelistgrabnew` exclude lists to prevent customized files from being copied to the CIMS node.

5.3 Configure Exclude Lists

You can accept the default exclude lists or customize them for the `esfs-generic` category before you clone the category to make `esfs-odd-`*`filesystem_name`*, `esfs-even-`*`filesystem_name`*, and `esfs-failed-`*`filesystem_name`*.

Refer to the *Data Management Platform (DMP) Administrator's Guide (S-2327)* for more information about Bright exclude lists.

5.4 Configure kdump on CLFS Nodes

Prerequisites

The CLFS node must be fully configured in Bright.

About this task

The kdump utility is configured on CLFS nodes (CentOS 6.6) by modifying the configuration files for the software image and configuring the Bright category. Dump files from service nodes can be stored either on the CIMS node using NFS®, or on a persistent partition (`/var/crash`) on service node local disk. Configure kdump for a generic CLFS category (such as `esfs-generic`) before the generic category is cloned into the categories required for esfsmon 2.1.0.

Procedure

1. Log in to the CIMS node as `root`.

2. Clone the CLFS node (in this example clfs-mdc001) node software image to a test image (this example clones `ESF-XX-2.3.0-201509031326` to `ESF-XX-2.3.0-kdump`.

   IMPORTANT: The time required to clone a software image in Bright depends on the image size. Cloning a minimal image (operating system only) completes in 5 to 10 minutes. A fully configured image can require between 30 to 75 minutes. The Bright clone operation spawns a background process that does not prevent you from rebooting a node or performing other configuration changes to software image before the image is fully cloned. The best practice is to copy the software image in `/cm/images` from the UNIX prompt, wait for the prompt to return, then clone the software image in Bright before proceeding.

```
cims1# cd /cm/images
cims1# cp -pr ESF-XX-2.3.0-201509031326 ESF-XX-2.3.0-kdump
cims1# cmsh
[cims1%] softwareimage
[cims1->softwareimage]% clone ESF-XX-2.3.0-201509031326 ESF-XX-2.3.0-kdump
```

3. Commit the changes.

```
[cims1->softwareimage*[ESF-XX-2.3.0-kdump]]% commit
cims1->softwareimage[ESF-XX-2.3.0-kdump]]%
```

4. Switch to `category` mode, and clone a new test category, or configure kdump for the `esfs-generic` category.

   - If performing an initial install of a CLFS node, use the `esfs-generic` category and enter:

```
[cims1->category]% use esfs-generic
[cims1->category*[esfs-generic*]%
```

   - To clone an existing CLFS node category, enter:

```
[cims1->category]% clone production_category test_category
[cims1->category*[test_category*]%
```

5. Assign the kdump software image (`ESF-XX-2.3.0-kdump`) to the esfs-generic category.

```
[cims1->category]% use esfs-generic
[cims1->category*[esfs-generic*]% set softwareimage ESF-XX-2.3.0-kdump
```

6. Add `/var/crash` to the exclude lists for the `ESF-XX-2.3.0-kdump` image. The set command launches the `vim` editor enables each exclude list to be edited. Add `- /var/crash/*` to each of the exclude list, list of excluded files:

```
[cims1->category*[esfs-generic*]% set excludelistsyncinstall
[cims1->category*[esfs-generic*]% set excludelistupdate
[cims1->category*[esfs-generic*]% set excludelistgrab
[cims1->category*[esfs-generic*]% set excludelistgrabnew
```

7. Save each file and commit the changes.

```
[cims1->category*[esfs-generic*]% commit
[cims1->category[esfs-generic]%
```

8. Assign the `esfs-generic` category to a CLFS node (clfs-mdc001) and commit the changes.

```
[cims1->category[esfs-generic]% device use clfs-mdc001
[esms1->device[clfs-mdc001]]% set category esfs-generic
[esms1->device*[clfs-mdc001*]]% commit
[esms1->device[clfs-mdc001]]%
```

9. Set the storage locations for crash dumps:

   - If you are saving kdump crash files to the service node local disk, add the following lines to the finalize script for the `esfs-generic` category. This command opens the `vim` editor. Scroll down and add the lines below before `exit 0` and save the finalize script, commit your changes, then continue at step *13* on page 36.

```
[cims1->device[clfs-mdc001]]% category use esfs-generic
[cims1->category[esfs-generic]]% set finalizescript
DEV=$( awk -- '{ if ($2 == "/localdisk/var/crash") { print $1; exit 0 } }' < /proc/mounts )[ -n "$DEV" ] && e2label $DEV crash


[cims1->category*[esfs-generic*]]% commit
[cims1->category[esfs-generic]]%
```

   - If crash dumps will be saved to the CIMS node, go to step *10* on page 35.

10. Use `fsexports` to determine whether the CIMS node exporting `/var/crash`.

```
[cims1->category[esfs-generic]]% device use cims1
[cims1->device[cims1]]% fsexports
[cims1->device[cims1]->fsexports]% list
Name (key)                                    Path
--------------------------------------------- -------------------------------
/cm/shared@esmaint-net                        /cm/shared
/home@esmaint-net                             /home
/var/spool/burn@esmaint-net                   /var/spool/burn
/cm/node-installer/certificates@esmaint-net   /cm/node-installer/certificates
/cm/node-installer@esmaint-net                /cm/node-installer
/var/crash@esmaint-net                        /var/crash
```

   - If `/var/crash` is not exported from the CIMS node, then configure it to export the file system to service nodes and commit the changes.

```
[cims1->device[cims1]->fsexports]% add /var/crash
[cims1->device*[cims1*]->fsexports*[/var/crash*]]% set name /var/crash@esmaint-net
[cims1->device*[cims1*]->fsexports*[/var/crash*]]% set extraoptions no_subtree_check
[cims1->device*[cims1*]->fsexports*[/var/crash*]]% set hosts esmaint-net
[cims1->device*[cims1*]->fsexports*[/var/crash*]]% set write yes
[cims1->device*[cims1*]->fsexports*[/var/crash*]]% commit
```

     Exit the `/var/crash` submode.

```
[cims1->device[cims1]->fsexports[/var/crash]]% exit
[cims1->device[cims1]->fsexports]%
```

11. Enter `quit` to exit `cmsh`.

12. Update the exported file systems.

```
cims1# exportfs -a
```

13. Use the `chroot` edit the `/boot/pxelinux.cfg/default` file in kdump test image (`ESF-XX-2.3.0-kdump`).

```
cims1# chroot /cm/images/ESF-XX-2.3.0-kdump
cims1:/> vi /boot/pxelinux.cfg/default
```

14. Scroll down to locate the following line:

```
# End of documentation, configuration follows:
```

15. Enter the following lines in the `default` configuration file:

```
LABEL kdump
KERNEL vmlinuz
IPAPPEND 3
APPEND initrd=initrd crashkernel=512M CMDS console=tty0 console=ttyS1,115200n8 CMDE
MENU LABEL ^KDUMP      - Normal boot mode with kdump
MENU DEFAULT
```

16. Examine the other `LABEL` entries in the `default` configuration file and remove the line `MENU DEFAULT`.

17. Exit and save the file.

18. Verify that `/var/crash` exists in the image and is a directory. If necessary, enter:

```
cims1:/> mkdir /var/crash
cims1:/> ls -l /var/crash
```

19. Edit the `/etc/kdump.conf` file and add or modify the following lines at the end of the file:

- To store crash dump files on the CIMS node add these lines.

  ```
  cims1:/> vi /etc/kdump.conf

  path /var/crash
  core_collector makedumpfile -c --message-level 1 -d 27
  link_delay 60
  default reboot
  ```

- To store crash dump files on the service node local disk, add these lines. Specify either an `ext4` or `ext3` file system.

  ```
  path /var/crash
  core_collector makedumpfile -c --message-level 1 -d 27
  link_delay 60
  default reboot
  ext3 LABEL=crash
  ```

20. Create a persistent partition (`/var/crash`) in the disk setup XML file for the kdump test category `esfs-generic`. A separate crash partition for crash dumps on the service node software image prevents `/var` from filling up and causing errors in the operating system.

21. Exit the editor and save the file.

22. Enable the `kdump` service in the `ESF-XX-2.3.0-kdump` software image.

```
cims1:/> chkconfig kdump on
```

23. Enter **exit** to exit the chroot shell.

24. Start a console window for the clfs-mdc001 node, reboot the node, and test kdump.

    a. Start a console window on the test service node (clfs-mdc001).

    ```
    cims1# cmsh
    [cims1]% device; use clfs-mdc001
    [cims1->device[clfs-mdc001]]% rconsole
    ```

    b. In the console window, reboot the test node.

    ```
    clfs-mdc001: reboot
    clfs-mdc001: Reboot in progress ...
    ```

    c. When the node reboots, log in, and initiate kdump.

    ```
    cims1#  ssh clfs-mdc001
    clfs-mdc001#  echo c > /proc/sysrq-trigger
    ```

    If dumping over NFS® to the , the dump file is created in /var/crash on the node. If dumping to the service node's local disk, the dump file is created in /var/crash on the local disk.

25. If necessary, assign the kdump software image (ESF-XX-2.3.0-kdump to the esfs-generic category, or switch to category mode and configure the production category to use the kdump software image.

```
cims1->category[esfs-generic]]% device
[cims1->device]% reboot -c esfs-generic
clfs-mdc001: Reboot in progress ...
```

26. Enter **quit** to exit cmsh.

5.5 Configure the Node Finalize Script for CLFS Nodes

Procedure

1. Log in to the CIMS node as root.

2. Make a copy of the default finalize script to an etc directory to prevent it from being overwritten during software updates.

```
esms1# cd /opt/cray/esms/cray-es-finalize-scripts-XX
esms1# mkdir -p etc
esms1# cp -p default/esf_finalize.sh etc/site.esf_finalize.sh
```

3. Edit the etc/site.esf_finalize.sh script.

```
esms1# vi etc/site.efs_finalize.sh
```

4. MDS configuration - Edit the IB subnet manager section. Cray recommends that the subnet manager for the IB fabric should be run on ib0 of the MDS node(s) for system configurations that have many CLFS file system nodes connected to an IB switch. Uncomment the following command line if you want to start the IB subnet manager on ib0 for MDS nodes.

```
# echo "/usr/sbin/opensm --daemon -g \`/usr/sbin/ibstat mlx4_0 1 | grep GUID | awk '{print \$3}'\`" >> /localdisk/etc/rc.d/rc.local
```

5. Uncomment the following lines if the MDS node must be configured for LDAP.

```
# chroot /localdisk chkconfig nslcd on
# chroot /localdisk chkconfig nscd on
```

6. If you want to set up multipath name mappings to configure multipath for MDS nodes, uncomment the commands in the next section. Enter the actual WWID's and aliases here. This was configured in previous releases in the `/etc/multipath/bindings` file. Configuring multipath name mappings in a `site.esf_finalize.sh` finalize script is the best practice.

```
#cat << EOMP >> /localdisk/etc/multipath.conf
#multipaths {
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias mgt
#   }
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias mdt
#   }
#}
#EOMP
```

7. OSS configuration - Uncomment the following lines to disable IB cards not present.

```
#sed -i -e "s/MTHCA_LOAD=yes/MTHCA_LOAD=no/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/QIB_LOAD=yes/QIB_LOAD=no/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/MLX4_EN_LOAD=yes/MLX4_EN_LOAD=no/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/CXGB3_LOAD=yes/CXGB3_LOAD=no/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/NES_LOAD=yes/NES_LOAD=no/" /localdisk/etc/infiniband/openib.conf
```

8. Uncomment the following lines to configure SRP for IB connected RAID.

```
## These lines are to configure SRP for IB-connected RAID
#sed -i -e "s/SDP_LOAD=yes/SDP_LOAD=no/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/SRP_LOAD=no/SRP_LOAD=yes/" /localdisk/etc/infiniband/openib.conf
#sed -i -e "s/SRP_DAEMON_ENABLE=no/SRP_DAEMON_ENABLE=yes/" /localdisk/etc/infiniband/openib.conf
```

9. Uncomment one of the following lines to start an IB subnet manager on `ib2` for OSS nodes. Cray recommends that `ib0` on OSS nodes be connected to the IB switch, that `ib1` should not be used, and that `ib2` and `ib3` connect to the storage array controllers.

   a. If one port from the IB card (`mlx4_1` in this example) is connected to storage, uncomment the following line to start the IB subnet manager on `ib2` (the IB interface connected to the storage array). Change the `2` to designate the IB interface port connected to the storage array.

   ```
   #echo "/usr/sbin/opensm --daemon -g \`/usr/sbin/ibstat mlx4_1 2 | grep GUID | awk '{print \$3}'\`" >> /localdisk/etc/rc.d/
   rc.local
   ```

   b. If both ports of the IB card are connected to storage, each line will activate one port. Uncomment one or both lines according to your site's configuration.

   ```
   #echo "/usr/sbin/opensm --daemon -g \`/usr/sbin/ibstat mlx4_1 1 | grep GUID | awk '{print \$3}'\`" >> /localdisk/etc/rc.d/
   rc.local
   #echo "/usr/sbin/opensm --daemon -g \`/usr/sbin/ibstat mlx4_1 2 | grep GUID | awk '{print \$3}'\`" >> /localdisk/etc/rc.d/
   rc.local
   ```

10. Uncomment the following lines to increase IB write performance.

```
#echo "### esFS max_sect setting" >> /localdisk/etc/srp_daemon.conf
#echo "max_sect=65535" >> /localdisk/etc/srp_daemon.conf
```

11. If you want to set up multipath name mappings to configure multipath for MDS nodes, uncomment the commands in the next section. Enter the actual WWID's and aliases here. This was configured in previous releases in the `/etc/multipath/bindings` file. Configuring multipath name mappings in a `site.esf_finalize.sh` finalize script is the best practice.

```
#cat << EOMP >> /localdisk/etc/multipath.conf
#
#multipaths {
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias ost0
#   }
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias ost1
#   }
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias ost2
#   }
#   multipath {
#      wwid  360080000000000000000000000000000
#      alias ost3
#   }
#}
#
#EOMP
```

12. Run the `cmsh` and set the finalize script for the `esfs-generic` category.

```
esms1# cmsh
[esms1]% category
[esms1->category]% use esfs-generic
[esms1->category[esfs-generic]]% set finalizescript/opt/cray/esms/cray-es-finalize-scripts-XX/etc/
site.esf_finalize.sh
[esms1->category*[esfs-generic*]]% commit
[esms1->category[esfs-generic]]%
```

13. Verify the `site.esf_finalize.sh` finalize script is correct.

```
[esms1->category[esfs-generic]]% get finalizescript
[esms1->category[esfs-generic]]% quit
```

5.6 Enable Boot Record in Software Image

Prerequisites

The service node must be fully configured in Bright.

About this task

Enable the boot record setting for software images so that service nodes can boot from their local hard drive if the node is not available.

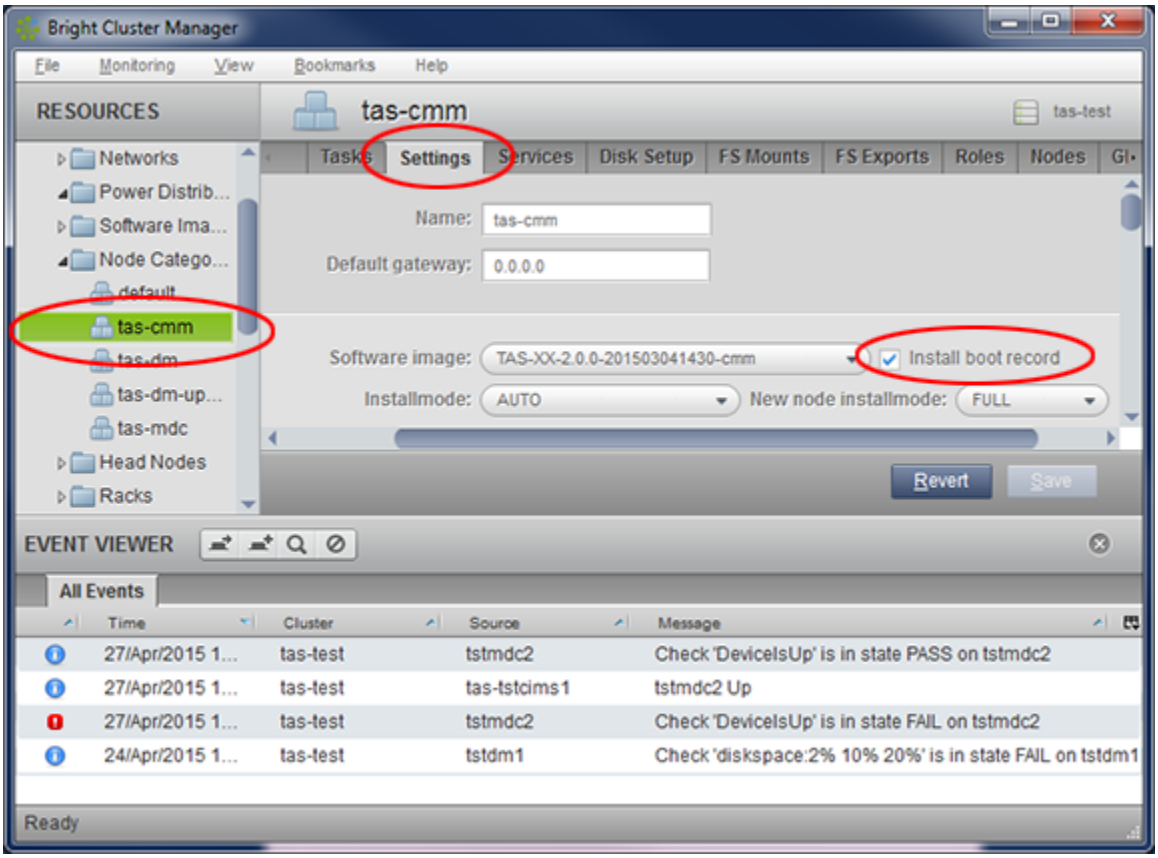
Set service node software images configured in Bright so that the node-installer installs a boot record on the local drive. This enables service nodes to reboot without access to the node.

Set the `installbootrecord` property for the service node settings and service node category to `on` so that the node can boot from the local drive. Booting from the hard drive must be set to a higher priority than network (PXE) booting in the BIOS settings for the node. Otherwise PXE booting occurs, despite the value set for `installbootrecord`.

Procedure

1. Use `cmgui` to select the service node category or device in the **RESOURCES** pane.
2. Select the **Settings** tab for the category or device.
3. Set the `Install boot record` checkbox and save the setting in the service node device or category as shown in the figure.
4. Unplug the cable for the `esmaint-net` network (the provisioning network or BOOTIF network) on the service node and reboot the node to determine if it can boot successfully without access to the node.

Figure 13. Install Boot Record



Using `cmsh`:

5. Log in to the node as `root`.
6. Enable the `installbootrecord` setting for either the device, or category:
   - To set the `installbootrecord` setting for a single service node, enter:

     ```
     cims1 # cmsh -c "device use device_name; set installbootrecord yes; commit"
     ```

   - To set the `installbootrecord` setting for a service node category, enter:

     ```
     cims1 # cmsh -c "category use category_name; set installbootrecord yes; commit"
     ```

7. Unplug the cable for the `esmaint-net` network (the provisioning network or BOOTIF network) on the service node and reboot the node to determine if it can boot successfully without access to the node.

5.7 Clone esfs-generic Category to esfsmon 2.1.0 Categories

Prerequisites

The CLFS node has booted the ESF software image when assigned to the `esfs-generic` category.

About this task

After the `esfs-generic` category is fully configured to support generic CLFS nodes, then clone the category to create `esfs-odd-`*`filesystem`*, `esfs-even-`*`filesystem`*, and `esfs-failed-`*`filesystem`* categories for esfsmon 2.1.0 and assign odd and even CLFS nodes to the appropriate category.

Procedure

1. Log in to the CIMS node as `root`.

2. Start `cmsh` and switch to `category` mode.

```
cims1# cmsh
[cims1]% category
```

3. Clone the `esfs-generic` category to even, odd, and failed categories required for esfsmon 2.1.0.

```
[cims1->category]% clone esfs-generic esfs-even-filesystem
[cims1->category*[esfs-even-filesystem*]]% clone esfs-generic esfs-odd-filesystem
[cims1->category*[esfs-odd-filesystem*]]% clone esfs-generic esfs-failed-filesystem
[cims1->category*[esfs-failed-filesystem*]]% commit
[cims1->category[esfs-failed-filesystem]]% list

Name (key)                 Software image
-------------------------  ------------------------
default                    default-image
esFS-MDS                   ESF-XX-2.3.0-201509031326+
esfs-generic              ESF-XX-2.3.0-kdump
esFS-OSS                   ESF-XX-2.3.0-201509031326
esLogin-XC                 ESL-XC-2.2.0-20140116037
esfs-even-filesystem       ESF-XX-2.3.0-kdump
esfs-failed-filesystem     ESF-XX-2.3.0-kdump
esfs-odd-filesystem        ESF-XX-2.3.0-kdump

[esms1->category*[esfs-failed-filesystem]]%
```

4. Enter **`quit`** to exit `cmsh`.

5.8 Verify the ESF Software Image and `esfs-generic` Category

Prerequisites

A fully configured CLFS node that has sucessfully booted from the `default` category and `default-image` software image.

About this task

This procedure boots the `clfs-mdc001` CLFS node from the ESF XX-2.3.0 software image and `esfs-generic` category to verify the complete configuration. If the CLFS node is configured with Fibre Channel (FC) host bus adapters (HBAs), configure SCSI RDAC driver kernel parameters for Fibre Channel storage and configure a separate ESF software image for service nodes using FC HBAs. Refer to *Configure RDAC Kernel Parameter for the ESF Software Image* on page 63.

Procedure

1. Log in to the CIMS node and start `cmsh`.

```
cims1# cmsh
[cims1]%
```

2. Switch to `device` mode.

```
cims1->device]%
```

3. List the devices to show their assigned category. Make sure the node being configured is assigned to the `esfs-generic` category.

```
cims1->device]% list
Type                 Hostname (key)   MAC                Category      Ip              Network
-------------------- ---------------- ------------------ ----------    --------------- --------------
EthernetSwitch       switch01         00:00:00:00:00:00                10.141.253.1    esmaint-net
MasterNode           cims1            78:2B:CB:40:CE:CA                10.141.255.254  esmaint-net
PhysicalNode         node001          00:00:00:00:00:00  default       10.141.0.1      esmaint-net
PhysicalNode         clfs_mds001      00:00:00:00:00:00  esfs-generic  10.141.0.1      esmaint-net
PhysicalNode         clfs_mdc01      78:2B:CB:50:E9:A3   esfs-generic  10.141.0.2      esmaint-net
```

4. Reboot the `clfs-mdc001` node.

```
[cims1->device]% use clfs-mdc001
[cims1->device[clfs-mdc001]]% reboot
```

From device mode, you can reboot all the nodes in a category using the `-c` command line option:

```
[cims1->device[clfs-mdc001]]% reboot -c catagory
```

Use the following syntax to specify a range of nodes:

```
[cims1->device[clfs-mdc001]]%  reboot -n esfs-mds[1,2],esfs-oss[1-4]
```

5. Open a console window to monitor the the `clfs-mdc001` boot messages.

- Use the `rconsole` command from `cmsh device` mode to start a remote console session. Press **Enter** in the console window to display output.

```
[cims1->device[clfs-mdc001]]% rconsole
 ========================================================================
To exit IPMI SOL <Enter> "&" "." (an ampersand followed by a period)
========================================================================
```

- To open a console window using `cmgui`, right-click on the node in the resource tree and select **Remote Console**, or select the node in the resource tree and use the **Tasks** tab **Remote Console** button.

6. Verify that the `esfs-generic` category and CLFS node (in this example clfs-mdc001) boot successfully and are configured properly before creating the categories required for esfsmon 2.1.0.

5.9 Add CLFS Nodes to esfsmon 2.1.0 Categories

Prerequisites

CLFS nodes have booted ESF software and are configured under the `esfs-generic` category, and the ESF software image has been customized and configured to support kdump, and `esfs-generic` category has been cloned into the categories required to support esfsmon 2.1.0 (`esfs-odd-`*`filesystem`*, `esfs-even-`*`filesystem`*, and `esfs-failed-`*`filesystem`*.

About this task

After the CLFS nodes have been configured to boot ESF from the `esfs-generic` category, the ESF software image has been customized and support kdump, and the `esfs-generic` category has been cloned into the odd, even, and failed, categories, each CLFS node is then assigned to the even or odd category based on the node name (`$HOSTNAME`).

Procedure

1. Log in to the CIMS node and start `cmsh`.

```
cims1# cmsh
[cims1]%
```

2. Switch to `device` mode.

```
[cims1]% device
```

3. Add a node to the `esfs-even-`*`filesystem`* category.

```
[cims1->device]% use clfs-mdc001
[cims1->device[clfs-mdc001]]% set category esfs-odd-filesystem
[cims1->device[clfs-mdc001*]]%
```

4. Set the rack information (device height, position in the rack, and rack number). This example indicates that the device height is 2U, its position in the rack is 12, and the rack number is 1. To display a list of all the possible settings, enter a question mark (`?`) and press **Enter**, or enter set and press the **Tab** key.

```
[cims1->device[clfs-mdc001*]]% set deviceheight 2
[cims1->device[clfs-mdc001*]]% set deviceposition 12
[cims1->device[clfs-mdc001*]]% set rack 1
```

5. Commit the changes.

```
[cims1->device[clfs-mdc001*]]% commit
[cims1->device[clfs-mdc001]]%
```

6. Repeat steps *3* on page 43 through *5* on page 43 to add each CLFS node to an odd or even category.

7. Check the CLFS categories.

```
[->device[clfs-mdc001*]]% commit
[cims1->device[clfs-mdc001]]%


[cims1->device[clfs-mdc001]]% category
```

```
[cims1->category]]% list
Name (key)                    Software image
------------------------      ------------------------
default                       default-image
default-diskless              default-image
esfs-generic                  ESF-XX-2.3.0-kdump
esFS-MDS                     ESF-XX-2.3.0-kdump
esFS-OSS                     ESF-XX-2.3.0-kdump
esfs-even-filesystem         ESF-XX-2.3.0-kdump
esfs-even-filesystem          ESF-XX-2.3.0-kdump
esfs-failed-filesystem        ESF-XX-2.3.0-kdump
esfs-odd-filesystem           ESF-XX-2.3.0-kdump
esLogin-XC                    ESL-XC-1.0.2-2013022113+
esLogin-XE                    ESL-XE-1.1.1_CLE4.1
```

```
[cims1->category]% usedby esfs-even-filesystem
Category used by the following:
Type             Name                      Parameter                 Autochange
---------------- ------------------------- ------------------------- ------------
Device           esfs-mds002               category                  no
Device           esfs-mds004               category                  no
Device           esfs-oss002               category                  no
Device           esfs-oss004               category                  no

[esms1->category]% usedby esfs-odd-filesystem
Category used by the following:
Type             Name                      Parameter                 Autochange
---------------- ------------------------- ------------------------- ------------
Device           esfs-mds001               category                  no
Device           esfs-mds003               category                  no
Device           esfs-oss001               category                  no
Device           esfs-oss003               category                  no
```

8. Enter **quit** to exit `cmsh`.

5.10 Backup the CLFS Software Image

About this task

Create a backup of Data Management Practices (DMP) production software images, images that are newly installed, upgraded, or customized for a specific configuration.

Procedure

1. Log in to the CIMS node as `root`.

2. Clone the CLFS software image to a backup image (this example clones `ESF-XX-2.3.0-kdump` to `ESF-XX-2.3.0-kdump-backup`.

   IMPORTANT: The time required to clone a software image in Bright depends on the image size. Cloning a minimal image (operating system only) completes in 5 to 10 minutes. A fully configured image can require between 30 to 75 minutes. The Bright clone operation spawns a background process that does not prevent you from rebooting a node or performing other configuration changes to software image before the image is fully cloned. The best practice is to copy the software image in `/cm/images` from the UNIX prompt, wait for the prompt to return, then clone the software image in Bright before proceeding.

```
cims1# cd /cm/images
cims1# cp -pr ESF-XX-2.3.0-kdump ESF-XX-2.3.0-kdump-backup
cims1# cmsh
```

```
[cims1%] softwareimage
[cims1->softwareimage]% clone ESF-XX-2.3.0-kdump ESF-XX-2.3.0-kdump-backup
```

3. Commit the changes.

```
[cims1->softwareimage*[ESF-XX-2.3.0-kdump]]% commit
cims1->softwareimage[ESF-XX-2.3.0-kdump]]%
```

4. Enter **quit** to exit `cmsh`.

6 Configure RAID Devices on CLFS Nodes

To configure RAID devices a on CLFS nodes, LUNs must configured on the storage arrays. Configure storage array LUNs using the OEM software provided for the storage array. NetApp storage devices may display errors when implementing multipath host mappings (`Volume not on preferred path due to AVT/RDAC failover`). Set the host mappings parameter to **Linux (DM-MP)** or **Linux** (on older firmware). Use the SANtricity™ Storage Manager Software command **Host Mappings->Default Group->Change Default Host Operating System** to configure the host mappings setting.

Verify LUNs are Available to CLFS Nodes

Make sure the CLFS node is fully configured in Bright and can boot from the ESF XX-2.3.0 software image.

Reboot the CLFS node or probe the SCSI bus to verify the LUNs are available. Two paths to LUN 5 in the example below (`/dev/sdg` and `/dev/sdm`) indicate multipath is enabled. Lustre® file systems can be configured after LUNs are available to the nodes. From the CLFS node, use the `lsscsi` command to probe the SCSI bus.

```
[root@esfs-oss001 ~]# lsscsi
[0:2:0:0] disk DELL PERC H710P 3.13 /dev/sda
[5:0:0:0] cd/dvd PLDS DVD+-RW DS-8A8SH KD51 /dev/sr0
[7:0:0:0] disk LSI INF-01-00 0780 /dev/sdb
[7:0:0:1] disk LSI INF-01-00 0780 /dev/sdc
[7:0:0:2] disk LSI INF-01-00 0780 /dev/sdd
[7:0:0:3] disk LSI INF-01-00 0780 /dev/sde
[7:0:0:4] disk LSI INF-01-00 0780 /dev/sdf
[7:0:0:5] disk LSI INF-01-00 0780 /dev/sdg
[8:0:0:0] disk LSI INF-01-00 0780 /dev/sdh
[8:0:0:1] disk LSI INF-01-00 0780 /dev/sdi
[8:0:0:2] disk LSI INF-01-00 0780 /dev/sdj
[8:0:0:3] disk LSI INF-01-00 0780 /dev/sdk
[8:0:0:4] disk LSI INF-01-00 0780 /dev/sdl
[8:0:0:5] disk LSI INF-01-00 0780 /dev/sdm
```

7 Configure the LSI MegaCLI RAID Utility

CIMS nodes that use Avago Technologies MegaRAID™ controllers, also the `megaraid_sas` kernel module with PERC710P, PERC 6/i RAID or other hardware modules. To manage, monitor, and configure the local RAID systems, install the LSI MegaCLI RAID utility on the CIMS. The MegaCLI utility also enables you to configure monitoring metrics, healthchecks, and administrator alerts in Bright that monitor the CIMS local RAID systems.

- Use *Configure MegaCLI in Bright for the CIMS Node* on page 47 to install and run the MegaCLI utility. The utility installs in `/opt/MegaRAID/MegaCli`. Verify the utility is installed and running correctly, then configure the Bright `healthcheck` feature to monitor CIMS node RAID devices.
- Use *Configure MegaCLI in Bright for Service Nodes* on page 51 to configure the MegaCLI utility for a service node.

7.1 Configure MegaCLI in Bright for the CIMS Node

Prerequisites

The CIMS node must be configured with Avago Technologies RAID devices using PERC710P, PERC 6/i RAID or other hardware modules.

About this task

The Avago Technologies MegaCLI RAID utility enables the CIMS node and Bright Cluster Manager (Bright) to monitor metrics, perform healthchecks, and initiate alerts. Data Management Practices (DMP) nodes also use the `megaraid_sas` kernel module with PERC710P, PERC 6/i RAID or other hardware modules. To manage, monitor, and configure the local RAID systems, install the MegaCLI RAID utility on the CIMS node.

Procedure

1. Open a web browser and access the download agreement at *http://www.lsi.com/Pages/user/eula.aspx?file=http://www.lsi.com/* (Avago Technologies). Click **Accept** to accept the software license and download agreement.

2. Navigate to the storage downloads area of the website and search for "MegaCLI Linux".

3. Download the latest MegaCLI archive for Linux®. Use a web browser to navigate to the **Downloads** area of *www.lsi.com*. Save the ZIP file locally, (for example, `MegaCLI.zip`).

4. Log in to CIMS node as `root`, copy the downloaded MegaCLI archive to the CIMS node, and decompress the archive.

```
cims1# mkdir /root/MegaCLI
cims1# cd /root/MegaCLI
```

```
cims1# scp user@remotesystem:/user/MegaCLI.zip /root/MegaCLI
cims1# unzip MegaCLI.zip
Archive: ./MegaCli_Linux.zip
creating: MegaCli_Linux/
inflating: MegaCli_Linux/MegaCli-8.07.08-1.i386.rpm
inflating: MegaCli_Linux/megacli_8.07.08-1_all.deb
```

5. Extract the Linux MegaCLI RPM from the archive.

```
cims1# pwd
/root/MegaCLI
cims1# cd MegaCli
cims1# rpm -ivh MegaCli-Revision.noarch.rpm
```

6. Run MegaCLI to verify it functions.

```
cims1# cd /opt/MegaRAID/MegaCli
cims1# ./MegaCli64 -AdpAllInfo -aAll |more
Adapter #0
==============================================================
                    Versions
                ================
Product Name    : PERC 6/i Integrated
Serial No       : 1122334455667788
FW Package Build: 6.0.2-0002

                    Mfg. Data
                ================
Mfg. Date       : 06/08/07
Rework Date     : 06/08/07
Revision No     :
Battery FRU     : N/A

                Image Versions in Flash:
                ================
FW Version         : 1.11.52-0396
BIOS Version       : NT13-2
WebBIOS Version    : 1.1-32-e_11-Rel
Ctrl-R Version     : 1.01-010B
Boot Block Version : 1.00.00.01-0008
...
```

7. Enter **cmsh** to start the Bright Cluster Manager shell (`cmsh`).

8. Switch to `monitoring healthchecks` mode and list the current health checks that are monitored in Bright.

```
[cims1]% monitoring healthchecks
[cims1->monitoring->healthchecks]% list
Name (key)                   Command
---------------------------- ------------------------------------------------
DeviceIsUp                   <built-in>
ManagedServicesOk            <built-in>
chrootprocess                /cm/local/apps/cmd/scripts/healthchecks/chrootp+
cmsh                         /cm/local/apps/cmd/scripts/healthchecks/cmsh
diskspace                    /cm/local/apps/cmd/scripts/healthchecks/diskspa+
dmesg                        /cm/local/apps/cmd/scripts/healthchecks/dmesg
exports                      /cm/local/apps/cmd/scripts/healthchecks/exports
failedprejob                 /cm/local/apps/cmd/scripts/healthchecks/failedp+
failover                     /cm/local/apps/cmd/scripts/healthchecks/failover
gpuhealth_quick              /cm/local/apps/cmd/scripts/healthchecks/gpuheal+
hardware-profile             /cm/local/apps/cmd/scripts/healthchecks/node-ha+
hpraid                       /cm/local/apps/cmd/scripts/healthchecks/hpraid
ib                           /cm/local/apps/cmd/scripts/healthchecks/ib
interfaces                   /cm/local/apps/cmd/scripts/healthchecks/interfa+
ldap                         /cm/local/apps/cmd/scripts/healthchecks/ldap
lustre                       /cm/local/apps/cmd/scripts/healthchecks/lustre
megaraid                     /cm/local/apps/cmd/scripts/healthchecks/megaraid
mounts                       /cm/local/apps/cmd/scripts/healthchecks/mounts
mysql                        /cm/local/apps/cmd/scripts/healthchecks/mysql
ntp                          /cm/local/apps/cmd/scripts/healthchecks/ntp
oomkiller                    /cm/local/apps/cmd/scripts/healthchecks/oomkill+
portchecker                  /cm/local/apps/cmd/scripts/healthchecks/portche+
```

```
rogueprocess                /cm/local/apps/cmd/scripts/healthchecks/roguepr+
schedulers                  /cm/local/apps/cmd/scripts/healthchecks/schedul+
smart                       /cm/local/apps/cmd/scripts/healthchecks/smart
ssh2node                    /cm/local/apps/cmd/scripts/healthchecks/ssh2node
swraid                      /cm/local/apps/cmd/scripts/healthchecks/swraid
testhealthcheck             /cm/local/apps/cmd/scripts/healthchecks/testhea+
```

- If `megaraid` health check is **not** present, add it, configure it, and save the changes for the CIMS node.

```
[cims1->monitoring->healthchecks]% add megaraid
[cims1->monitoring->healthchecks*[megaraid*]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Class of healthcheck           misc
Command
Description
Disabled                       no
Extended environment           no
Name                           megaraid
Notes                          <0 bytes>
Only when idle                 no
Parameter permissions          optional
Revision
Sampling method                samplingonnode
State flapping count           7
Timeout                        5
Valid for                      node,headnode
```

```
[cims1->monitoring->healthchecks*[megaraid*]]% set classofhealthcheck disk
[cims1->monitoring->healthchecks*[megaraid*]]% set command /cm/local/apps/cmd/scripts/
healthchecks/megaraid
[cims1->monitoring->healthchecks*[megaraid*]]% commit
[cims1->monitoring->healthchecks[megaraid]]%
```

- If the `megaraid` heath check is present, complete the configuration.

```
[cims1->monitoring->healthchecks[megaraid]]% monitoring setup healthconf headnode
[cims1->monitoring->setup*[HeadNode*]->healthconf*]% add megaraid
[cims1->monitoring->setup*[HeadNode*]->healthconf*]% add megaraid
[cims1->monitoring->setup*[HeadNode*]->healthconf*[megaraid*]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Check Interval                 120
Disabled                       no
Fail Actions
Fail severity                  10
GapThreshold                   2
HealthCheck                    megaraid
HealthCheckParam
LogLength                      3000
Only when idle                 no
Pass Actions
Revision
Stateflapping Actions
Store                          yes
ThresholdDuration              1
Unknown Actions
Unknown severity               10
```

- To configure this health check for a service node category, specify the service node category name on the command line. For example, to specify a service node category of `esfs-generic` use the following commands:

```
[cims1->monitoring->healthchecks[megaraid]]% monitoring setup healthconf esfs-generic
[cims1->monitoring->setup[esfs-generic]->healthconf]% add megaraid
[cims1->monitoring->setup*[esfs-generic*]->healthconf*[megaraid*]]% show
Parameter                      Value
------------------------------ ------------------------------------------------
Check Interval                 120
Disabled                       no
Fail Actions
Fail severity                  10
GapThreshold                   2
HealthCheck                    megaraid
```

```
HealthCheckParam
LogLength                       3000
Only when idle                  no
Pass Actions
Revision
Stateflapping Actions
Store                           yes
ThresholdDuration               1
Unknown Actions
Unknown severity                10
```

9. Enter **`commit`** from the `cmsh` command line to commit the changes.

10. Wait a few minutes, and switch to `device` mode to show the health data for the CIMS node. Verify that the `megaraid` health data shows `PASS`. To show the health data for a service node (`clfs-mdc001` for example), enter.

```
[cims1->monitoring->setup[esfs-generic]->healthconf[megaraid]]% device use clfs-mdc001
[cims1->device[clfs-mdc001]]% latesthealthdata
Health Check                 Severity Value            Age (sec.) Info Message
---------------------------- -------- ---------------- ---------- ------------------------
DeviceIsUp                   0        PASS             1
ManagedServicesOk            0        PASS             19
mounts                       0        PASS             19
exports                      0        PASS             19
smart                        0        PASS             19         sda: Smart command failed
ldap                         0        PASS             19
failover                     0        PASS             19
interfaces                   40       FAIL             19         eth4 not up
oomkiller                    0        PASS             19
cmsh                         0        PASS             19
mysql                        0        PASS             19
failedprejob                 0        PASS             19
diskspace:2% 10% 20%         0        PASS             19
ntp                          0        PASS             19
schedulers                   0        PASS             19
chrootprocess                0        PASS             19
megaraid                     0        PASS             19
```

11. Enter **`quit`** to exit `cmsh`.

12. Use the following command to run the `megaraid` health check from a UNIX command line.

```
[cims1->device[cims1]]% quit

cims1# /cm/local/apps/cmd/scripts/healthchecks/megaraid
PASS

cims1# /cm/local/apps/cmd/scripts/healthchecks/megaraid -d 3>&1
megacli command path: /opt/MegaRAID/MegaCli/MegaCli64
cmd:     /opt/MegaRAID/MegaCli/MegaCli64 -LdPdInfo -aALL -NoLog
line:    Adapter #0
adapter: 0
line:    Virtual Drive: 0 (Target Id: 0)
vdrive:  0
line:    State               : Optimal
vstate:  Optimal
line:    Span: 0 - Number of PDs: 3
span:    0
line:    PD: 0 Information
pdisk:   0
line:    Enclosure Device ID: 32
enc:     32
line:    Slot Number: 0
encslot: 0
line:    Firmware state: Online, Spun Up
pstate:  Online, Spun Up
line:    Drive has flagged a S.M.A.R.T alert : No
psmart:  No
line:    PD: 1 Information
...
```

7.2 Configure MegaCLI in Bright for Service Nodes

Prerequisites

The service node must be configured with LSI RAID devices using PERC710P, PERC 6/i RAID or other hardware modules.

About this task

The LSI® MegaCLI RAID utility enables Bright Cluster Manager (Bright) to monitor metrics, perform healthchecks, and initiate alerts. Data Management Practices (DMP) nodes that use LSI® Corporation MegaRAID™ controllers, also the `megaraid_sas` kernel module with PERC710P, PERC 6/i RAID or other hardware modules. To manage, monitor, and configure the local RAID systems on service nodes, install the LSI MegaCLI RAID utility on the CIMS node, then configure it for service nodes.

Procedure

1. If you have already downloaded MegaCLI and configured it for the CIMS node, go to step *8* on page 52.

2. Open a web browser and access the download agreement at *http://www.lsi.com/Pages/user/eula.aspx?file=http://www.lsi.com/* (Avago Technologies). Click **Accept** to accept the software license and download agreement.

3. Navigate to the storage downloads area of the website and search for "MegaCLI Linux".

4. Download the latest MegaCLI archive for Linux®. Use a web brower to navigate to the **Downloads** area of *www.lsi.com*. Save the ZIP file locally, (for example, `MegaCLI.zip`).

5. Log in to CIMS node as root, copy the downloaded MegaCLI archive to the CIMS node, and decompress the archive.

```
cims1# mkdir /root/MegaCLI
cims1# cd /root/MegaCLI
cims1# scp user@remotesystem:/user/MegaCLI.zip /root/MegaCLI
cims1# unzip MegaCLI.zip
Archive: ./MegaCli_Linux.zip
creating: MegaCli_Linux/
inflating: MegaCli_Linux/MegaCli-8.07.08-1.i386.rpm
inflating: MegaCli_Linux/megacli_8.07.08-1_all.deb
```

6. Extract the Linux MegaCLI RPM from the archive.

```
cims1# pwd
/root/MegaCLI
cims1# cd MegaCli
cims1# rpm -ivh MegaCli-Revision.noarch.rpm
```

7. Run MegaCLI on the CIMS node to verify it functions.

```
cims1# cd /opt/MegaRAID/MegaCli
cims1# ./MegaCli64 -AdpAllInfo -aAll |more
Adapter #0
==============================================================
                    Versions
                ================
Product Name    : PERC 6/i Integrated
```

```
Serial No       : 1122334455667788
FW Package Build: 6.0.2-0002

                    Mfg. Data
                ================
Mfg. Date       : 06/08/07
Rework Date     : 06/08/07
Revision No     :
Battery FRU     : N/A

                Image Versions in Flash:
                ================
FW Version         : 1.11.52-0396
BIOS Version       : NT13-2
WebBIOS Version    : 1.1-32-e_11-Rel
Ctrl-R Version     : 1.01-010B
Boot Block Version : 1.00.00.01-0008
...
```

8. Copy the current working service node software image in `/cm/images/` to a new image name that identifies it as supporting the MegaCLI utility.

> **IMPORTANT:** The time required to clone a software image in Bright depends on the image size. Cloning a minimal image (operating system only) completes in 5 to 10 minutes. A fully configured image can require between 30 to 75 minutes. The Bright clone operation spawns a background process that does not prevent you from rebooting a node or performing other configuration changes to software image before the image is fully cloned. The best practice is to copy the software image in `/cm/images` from the UNIX prompt, wait for the prompt to return, then clone the software image in Bright before proceeding.

```
cims1# cp -pr /cm/images/ESF-XX-2.3.0-kdump /cm/images/ESF-XX-2.3.0-kdump-megacli
```

9. Start `cmsh`, clone the service node software image `ESF-XX-2.3.0-kdump` to`ESF-XX-2.3.0-kdump-megacli`.

```
cims1# cmsh
[cims1]% softwareimage
[cims1->softwareimage]% clone ESF-XX-2.3.0-kdump ESF-XX-2.3.0-kdump-megacli
[cims1->softwareimage*]% commit
```

10. Switch to `category` mode, and clone the service node category (in this example `esfs-generic` to a new category named `esfs-generic-megacli` and assign the software image (`ESF-XX-2.3.0-kdump-megacli`) to the `esfs-generic-megacli` category.

```
[cims1->softwareimage]% category
[cims1->category]% clone esfs-generic esfs-generic-megacli
[cims1->category*[esfs-generic-megacli*]% commit
[cims1->category[esfs-generic-megacli]% set softwareimage ESF-XX-2.3.0-kdump-megacli
[cims1->category*[ESF-XX-2.3.0-kdump-megacli*]% commit
[cims1->category[ESF-XX-2.3.0-kdump-megacli]% quit
esms1#
```

11. Bind mount `/root/MegaCLI_Linux` to `/cm/images/ESF-XX-2.3.0-kdump-megacli/tmp/MegaCLI` or copy the RPM to a directory in the `ESF-XX-2.3.0-kdump-megacli` software image. This example shows the bind mount method.

```
cims1# mkdir -p /cm/images/ESF-XX-2.3.0-kdump-megacli/tmp/MegaCLI
cims1# mount --bind /root/MegaCLI_Linux /cm/images/ESF-XX-2.3.0-kdump-megacli/tmp/MegaCLI
```

12. Use the `chroot` shell to install the MegaCLI software on the `ESF-XX-2.3.0-kdump-megacli` software image, then exit the `chroot` shell.

```
cims1# chroot /cm/images/ESF-XX-2.3.0-kdump-megacli/
cims1:> rpm -ivh MegaCli-Revision.noarch.rpm
```

```
reparing... ########################################## [100%]
 1:MegaCli ########################################## [100%] esms1:> exit
```

13. Remove the bind mount.

```
cims1# umount /cm/images/ESF-XX-2.3.0-kdump-megacli/tmp/MegaCLI
cims1# rm -f /cm/images/ESF-XX-2.3.0-kdump-megacli/tmp/MegaCLI
```

14. Use `cmsh` to assign a service node (in this example, `esfs-mds1`) to the`esfs-generic-megacli` category.

```
cims1# cmsh
[cims1]% device
[cims1->device]% use clfs-mdc001
[cims1->device[clfs-mdc001]]% set category esfs-generic-megacli
[cims1->device*[clfs-mdc001*]]% commit
```

15. Reboot the `clfs-mdc001` node (or from `cmsh` use `imageupdate`) to test the `ESF-XX-2.3.0-kdump-megacli` software image and exit `cmsh`.

```
[cims1->device[clfs-mdc001]]% reboot clfs-mdc001
[cims1->device[clfs-mdc001]]% quit
```

16. Open a remote console on the `clfs-mdc001` node and verify that the `ESF-XX-2.3.0-kdump-megacli` software image boots without errors.

17. Use SSH to log in to `clfs-mdc001` and run the MegaCLI utility to verify that it is functioning on the service node.

```
cims1# ssh clfs-mdc001
Last login: Thu Nov  7 12:56:51 2013 from esms1.cm.cluster
[root@clfs-mdc001 ~]# cd /opt/MegaRAID/MegaCli
[root@clfs-mdc001 ~]# ./MegaCli64 -AdpAllInfo -aAll |more
Adapter #0
==============================================================
                    Versions
                ================
Product Name    : PERC 6/i Integrated
Serial No       : 1122334455667788
FW Package Build: 6.0.2-0002

                    Mfg. Data
                ================
Mfg. Date       : 06/08/11
Rework Date     : 06/08/11
Revision No     :
Battery FRU     : N/A

                Image Versions in Flash:
                ================
FW Version         : 1.11.52-0396
BIOS Version       : NT13-2
WebBIOS Version    : 1.1-32-e_11-Rel
Ctrl-R Version     : 1.01-010B
Boot Block Version : 1.00.00.01-0008
...
```

18. Assign other service nodes to the `esfs-generic-megacli` category.

```
cims1# cmsh
[cims1]% device
[cims1->device]% use servicenode
[cims1->device[servicenode]% set category esfs-generic-megacli
[cims1->device*[servicenode*]% commit
```

19. Switch to `monitoring healthchecks` mode and list the current health checks that are monitored in Bright.

```
[cims1->device[servicenode]% monitoring healthchecks
[cims1->monitoring->healthchecks]% list
Name (key)                  Command
--------------------------- ------------------------------------------------
DeviceIsUp                  <built-in>
ManagedServicesOk           <built-in>
chrootprocess               /cm/local/apps/cmd/scripts/healthchecks/chrootp+
cmsh                        /cm/local/apps/cmd/scripts/healthchecks/cmsh
diskspace                   /cm/local/apps/cmd/scripts/healthchecks/diskspa+
dmesg                       /cm/local/apps/cmd/scripts/healthchecks/dmesg
exports                     /cm/local/apps/cmd/scripts/healthchecks/exports
failedprejob                /cm/local/apps/cmd/scripts/healthchecks/failedp+
failover                    /cm/local/apps/cmd/scripts/healthchecks/failover
gpuhealth_quick             /cm/local/apps/cmd/scripts/healthchecks/gpuheal+
hardware-profile            /cm/local/apps/cmd/scripts/healthchecks/node-ha+
hpraid                      /cm/local/apps/cmd/scripts/healthchecks/hpraid
ib                          /cm/local/apps/cmd/scripts/healthchecks/ib
interfaces                  /cm/local/apps/cmd/scripts/healthchecks/interfa+
ldap                        /cm/local/apps/cmd/scripts/healthchecks/ldap
lustre                      /cm/local/apps/cmd/scripts/healthchecks/lustre
megaraid                    /cm/local/apps/cmd/scripts/healthchecks/megaraid
mounts                      /cm/local/apps/cmd/scripts/healthchecks/mounts
mysql                       /cm/local/apps/cmd/scripts/healthchecks/mysql
ntp                         /cm/local/apps/cmd/scripts/healthchecks/ntp
oomkiller                   /cm/local/apps/cmd/scripts/healthchecks/oomkill+
portchecker                 /cm/local/apps/cmd/scripts/healthchecks/portche+
rogueprocess                /cm/local/apps/cmd/scripts/healthchecks/roguepr+
schedulers                  /cm/local/apps/cmd/scripts/healthchecks/schedul+
smart                       /cm/local/apps/cmd/scripts/healthchecks/smart
ssh2node                    /cm/local/apps/cmd/scripts/healthchecks/ssh2node
swraid                      /cm/local/apps/cmd/scripts/healthchecks/swraid
testhealthcheck             /cm/local/apps/cmd/scripts/healthchecks/testhea+
```

- If `megaraid` health check is **not** present, add it, configure it, and save the changes for the CIMS node.

```
[cims1->monitoring->healthchecks]% add megaraid
[cims1->monitoring->healthchecks*[megaraid*]]% show
Parameter                   Value
--------------------------- ------------------------------------------------
Class of healthcheck        misc
Command
Description
Disabled                    no
Extended environment        no
Name                        megaraid
Notes                       <0 bytes>
Only when idle              no
Parameter permissions       optional
Revision
Sampling method             samplingonnode
State flapping count        7
Timeout                     5
Valid for                   node,headnode
```

```
[cims1->monitoring->healthchecks*[megaraid*]]% set classofhealthcheck disk
[cims1->monitoring->healthchecks*[megaraid*]]% set command /cm/local/apps/cmd/scripts/
healthchecks/megaraid
[cims1->monitoring->healthchecks*[megaraid*]]% commit
[cims1->monitoring->healthchecks[megaraid]]%
```

- If the `megaraid` heath check is present, configure the health check for the service node category (in this example, `esfs-generic-megacli`.

```
[cims1->monitoring->healthchecks[megaraid]]% monitoring setup healthconf esfs-generic
[cims1->monitoring->setup*[esfs-generic*]->healthconf*]% add megaraid
[cims1->monitoring->setup*[esfs-generic*]->healthconf*]% add megaraid
[cims1->monitoring->setup*[esfs-generic*]->healthconf*[megaraid*]]% show
Parameter                   Value
--------------------------- ------------------------------------------------
Check Interval              120
Disabled                    no
Fail Actions
Fail severity               10
GapThreshold                2
HealthCheck                 megaraid
```

```
HealthCheckParam
LogLength                   3000
Only when idle              no
Pass Actions
Revision
Stateflapping Actions
Store                       yes
ThresholdDuration           1
Unknown Actions
Unknown severity            10
```

20. Enter **`commit`** from the `cmsh` command line to commit the changes.

21. Wait a few minutes, and switch to `device` mode to show the health data for the CIMS node. Verify that the `megaraid` health data shows `PASS`. To show the health data for a service node (`clfs-mdc001` for example), enter:

```
[cims1->monitoring->setup[esfs-generic]->healthconf[megaraid]]% device use clfs-mdc001
[cims1->device[clfs-mdc001]]% latesthealthdata
Health Check                Severity Value            Age (sec.) Info Message
--------------------------- -------- ---------------- ---------- ------------------------
DeviceIsUp                  0        PASS             1
ManagedServicesOk           0        PASS             19
mounts                      0        PASS             19
exports                     0        PASS             19
smart                       0        PASS             19         sda: Smart command failed
ldap                        0        PASS             19
failover                    0        PASS             19
interfaces                  40       FAIL             19         eth4 not up
oomkiller                   0        PASS             19
cmsh                        0        PASS             19
mysql                       0        PASS             19
failedprejob                0        PASS             19
diskspace:2% 10% 20%        0        PASS             19
ntp                         0        PASS             19
schedulers                  0        PASS             19
chrootprocess               0        PASS             19
megaraid                    0        PASS             19
```

22. Enter **`quit`** to exit `cmsh`.

23. Use the following command to run the `megaraid` health check from a UNIX command line.

```
[cims1->device[cims1]]% quitcims1# /cm/local/apps/cmd/scripts/healthchecks/megaraid
PASS
cims1# /cm/local/apps/cmd/scripts/healthchecks/megaraid -d 3>&1
megacli command path: /opt/MegaRAID/MegaCli/MegaCli64
cmd:     /opt/MegaRAID/MegaCli/MegaCli64 -LdPdInfo -aALL -NoLog
line:    Adapter #0
adapter: 0
line:    Virtual Drive: 0 (Target Id: 0)
vdrive:  0
line:    State               : Optimal
vstate:  Optimal
line:    Span: 0 - Number of PDs: 3
span:    0
line:    PD: 0 Information
pdisk:   0
line:    Enclosure Device ID: 32
enc:     32
line:    Slot Number: 0
encslot: 0
line:    Firmware state: Online, Spun Up
pstate:  Online, Spun Up
line:    Drive has flagged a S.M.A.R.T alert : No
psmart:  No
line:    PD: 1 Information
...
```

8 Configure DM Multipath on CLFS Nodes

Prerequisites

All CLFS nodes configured for DM multipath have a connection to each storage array controller. Specific WWID values must be added to the `multipaths {}` section of the `/etc/multipath.conf` file so that devices have persistent names each time the node boots.

If you are configuring multipath, and are experiencing errors from NetApp™ storage devices such as `Volume not on preferred path due to AVT/RDAC failover`, be sure to set the host mappings parameter to Linux (DM-MP) or Linux (on older firmware). Use the SANtricity™ Storage Manager Software command Host Mappings->Default Group->Change Default Host Operating System to configure the host mappings setting.

About this task

Set up multipath name mappings to configure multipath for CLFS nodes in the `site.esf_finalize.sh` finalize script for the CLFS node category.

> **NOTE:** If you are configuring multipath, and are experiencing errors from NetApp™ storage devices such as Volume not on preferred path due to AVT/RDAC failover, be sure to set the host mappings parameter to **Linux (DM-MP)** or **Linux** (on older firmware). Use the SANtricity™ Storage Manager Software command **Host Mappings->Default Group->Change Default Host Operating System** to configure the host mappings setting.

Procedure

1. After `ESFinstall` has completed, the `multipathd` service must be enabled for the ESF software image (ESF-XX-2.3.0-201509031326 in this example).

```
cims1# chroot /cm/images/ESF-XX-2.3.0-201509031326
[root@cims1/]# chkconfig multipathd on
```

2. Copy `multipath.conf.cray` to `/etc` in the ESF software image.

```
[root@cims1/]# cp -p /opt/cray/esfs/cray-esf-multipath-XX/default/etc/multipath.conf.cray\
/etc/multipath.conf
```

3. Exit the `chroot` shell.

```
[root@cims1/]# exit
cims1#
```

4. Identify the disk names for DM multipath on the MDS nodes. Boot the first MDS node (clfs-mdc001and use SSH to login.

```
cims1# ssh clfs-mdc001
Last login: Tue Jun 11 15:42:52 2014 from cims1.cm.cluster
[root@clfs-mdc001 ~]#
```

5. Use the `multipath -ll` command to display the disk devices.

```
clfs-mdc001# multipath -ll
mpatha (360080e50002f82ca000002ca5102bcc4) dm-0 LSI,INF-01-00
size=2.2T features='2 pg_init_retries 50' hwhandler='1 rdac' wp=rw
|-+- policy='round-robin 0' prio=6 status=active
| `- 0:0:0:0 sda 8:0 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
`- 0:0:1:0 sdb 8:16 active ghost running
```

   In the previous step, the WWID (`360080e50002f82ca000002ca5102bcc4`) is the device labeled as `mdt0` instead of `mpatha`.

6. On some systems, unwanted disk device partitions may need to be removed using the `parted` command to identify which disk device (in `/dev/sda` form) corresponds to this WWID. Use the following commands to remove unwanted partitions.

```
clfs-mdc001# ls -l /dev/disk/by-id | grep scsi-360080e50002f82ca000002ca5102bcc4
lrwxrwxrwx 1 root root 9 Jun 10 09:28 scsi-360080e50002f82ca000002ca5102bcc4 -> ../../sda
[root@clfs-mdc001 ~]# parted /dev/sda
GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.

(parted) print
Model: LSI INF-01-00 (scsi)
Disk /dev/sda: 2398GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number  Start   End     Size    File system     Name       Flags
 1      17.4kB  20.5GB  20.5GB  ext3            /
 2      20.5GB  22.5GB  2048MB  ext3            /var
 3      22.5GB  24.6GB  2048MB  ext3            /tmp
 4      24.6GB  41.0GB  16.4GB  linux-swap(v1)  /dev/sda4
 5      41.0GB  2398GB  2357GB  ext3            /local
```

7. Remove all partitions so that Lustre® can use the entire device. This example removes partitions 2 through 5 and shows that only one partition is left (`/dev/sda`).

```
(parted) rm 5
(parted) rm 4
(parted) rm 3
(parted) rm 2
(parted) p
Model: LSI INF-01-00 (scsi)
Disk /dev/sda: 2398GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number  Start   End     Size    File system  Name  Flags
 1      17.4kB  20.5GB  20.5GB  ext3         /

(parted) rm 1
(parted) p
Error: /dev/sda: unrecognised disk label
(parted) quit
Information: You may need to update /etc/fstab.
```

8. Exit the `clfs-mdc001` node and SSH to the the next MDS node. Repeat these steps to identify all the disk names on each MDS for multipath.

9. Identify the disk names for multipath on the OSS nodes. Boot each OSS node and use SSH to login.

```
cims1# ssh clfs-oss001
Last login: Tue Jun 11 15:45:32 2013 from cims1.cm.cluster
[root@clfs-oss001 ~]#
```

10. Use the `multipath -ll` command to display the disk devices.

```
[root@clfs-oss001 ~]# multipath -ll
mpatha (360080e50001f8c64000000b4513098b2) dm-2 LSI,INF-01-00
size=7.3T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 rdac' wp=rw
|-+- policy='round-robin 0' prio=6 status=active
| `- 7:0:0:5 sdg 8:96 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
`- 8:0:0:5 sdm 8:192 active ghost running
mpathb (360080e50001f87c8000000b951309893) dm-3 LSI,INF-01-00
size=7.3T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 rdac' wp=rw
|-+- policy='round-robin 0' prio=6 status=active
| `- 8:0:0:4 sdl 8:176 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
`- 7:0:0:4 sdf 8:80 active ghost running
mpathc (360080e50001f8c64000000ae51309849) dm-1 LSI,INF-01-00
size=7.3T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 rdac' wp=rw
|-+- policy='round-robin 0' prio=6 status=active
| `- 7:0:0:3 sde 8:64 active ready running
`-+- policy='round-robin 0' prio=1 status=enabled
`- 8:0:0:3 sdk 8:160 active ghost running
.
.
.
```

11. On some systems, the disk devices may need to be modified with the `parted` command to remove unneeded partitions. Identify which disk device (`/dev/sda` form) relates to this WWID using the following comand and remove unwanted partitions.

```
[root@clfs-oss001 ~]# ls -l /dev/disk/by-id | grep scsi-360080e50001f8c64000000b4513098b2
lrwxrwxrwx 1 root root  9 Jun  3 14:40 scsi-360080e50001f8c64000000b4513098b2 -> ../../sdm

[root@clfs-oss001 ~]# parted /dev/sdm
GNU Parted 2.1
Using /dev/sda
Welcome to GNU Parted! Type 'help' to view a list of commands.

(parted) print
Model: LSI INF-01-00 (scsi)
Disk /dev/sdm: 2398GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number  Start   End     Size    File system     Name       Flags
 1      17.4kB  20.5GB  20.5GB  ext3            /
 2      20.5GB  22.5GB  2048MB  ext3            /var
 3      22.5GB  24.6GB  2048MB  ext3            /tmp
 4      24.6GB  41.0GB  16.4GB  linux-swap(v1)  /dev/sda4
 5      41.0GB  2398GB  2357GB  ext3            /local
```

12. Remove all partitions so that Lustre® can use the entire device. This example removespartitions 2 through 5 and shows that only one partition is left (`/dev/sda`).

```
(parted) rm 5
(parted) rm 4
(parted) rm 3
(parted) rm 2
(parted) p
Model: LSI INF-01-00 (scsi)
Disk /dev/sda: 2398GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt

Number  Start   End     Size    File system  Name  Flags
 1      17.4kB  20.5GB  20.5GB  ext3         /

(parted) rm 1
(parted) p
Error: /dev/sda: unrecognised disk label
(parted) quit
Information: You may need to update /etc/fstab.
```

13. Exit the `clfs-oss001` node and repeat these steps to identify the disk names for DM multipath for the other OSS nodes.

14. After the disk names for multipath have been identified, edit finalize script (modify the `site.esf_finalize.sh` script for the node category in `/opt/cray/esms/cray-es-finalize-scripts-XX/etc/`). Add specific WWID values to the `multipaths {}` section of the `/etc/multipaths.conf` file the make sure that the devices have persistent names each time the node boots. Specify a label for each disk, such as `/dev/mapper/ost5` or `/dev/mapper/mdt0`, then use that label for the Lustre configuration file.

    MGT and MDT aliases and WWID settings:

    ```
    cat << EOMP >> /localdisk/etc/multipath.conf
    multipaths {
    multipath {
    wwid 360080000000000000000000000000000
    alias mgt
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias mdt0
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias mdt1
    }
    }
    EOMP
    ```

    OST aliases and WWID settings:

    ```
    cat << EOMP >> /localdisk/etc/multipath.conf
    multipaths {
    multipath {
    wwid 360080000000000000000000000000000
    alias ost0
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias ost1
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias ost2
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias ost3
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias ost4
    }
    multipath {
    wwid 360080000000000000000000000000000
    alias ost5
    }
    }
    EOMP
    ```

15. Reboot all MDS and OSS nodes that are configured with multipath.

16. Confirm that the MDT device `/dev/mapper/mdt` is available on the MDS nodes.

    ```
    cims1# ssh clfs-mdc001 fdisk -l /mapper/mdt0
    ```

17. Confirm that the OST devices `/dev/mapper/mdt` are available on the OSS nodes.

```
cims1# ssh clfs-oss001  fdisk -l /mapper/ost0
cims1# ssh clfs-oss001  fdisk -l /mapper/ost1
cims1# ssh clfs-oss001  fdisk -l /mapper/ost2
cims1# ssh clfs-oss001  fdisk -l /mapper/ost3
cims1# ssh clfs-oss001  fdisk -l /mapper/ost4
cims1# ssh clfs-oss001  fdisk -l /mapper/ost5
```

18. The disk device name `/dev/mapper/mdt0` can be used for the MDT and MGT and `/dev/mapper/ost0` can be used for OST0 when configuring the Lustre `fs_defs` file and similarly for the other OST devices. The example below shows how to configure `/dev/mapper/ost[0-5]` to reference the appropriate devices with even numbered OST devices on `oss002` and odd numbered OST devices on `oss001`.

```
## MDT
## MetaData Target
mdt: node=esfs-mds001
dev=/dev/mapper/mdt0
fo_node=clfs-mdc002
index=0
## MGT
## Management Target
mgt: node=esfs-mds001
dev=/dev/mapper/mdt0
fo_node=clfs-mdc002
index=0
## Object Storage Target(s)
ost: node=esfs-oss00[2,1]
dev=/dev/mapper/ost[0-5]
fo_node=esfs-oss00[1,2]
index=0
70
```

9 Configure QLogic Switch Fibre Channel Utility

Prerequisites

The CLFS must be fully configured in Bright and boot from an ESF software image.

```
Usage: qaucli [options]
Legends:
   <hba instance>    - The HBA instance number of an HBA port.
   <hba wwpn>        - The World Wide Port Name of an HBA port.
                       (xx-xx-xx-xx-xx-xx-xx-xx or xxxxxxxxxxxxxx).
   <all>             - All available adapters
   <target wwnn>     - The World Wide Node Name of a target device.
                       (xx-xx-xx-xx-xx-xx-xx-xx or xxxxxxxxxxxxxx).
   <target wwpn>     - The World Wide Port Name of a target device.
                       (xx-xx-xx-xx-xx-xx-xx-xx or xxxxxxxxxxxxxx).
   <target portid>   - The Port ID of a target  device.
                       (xx-xx-xx or xxxxxx).
   <target id>       - The ID of a target device.
   <lun id>          - The Logical Unit Number of a LUN.
   <lunname>         - The udev persistent device name to be assigned to
                       a designated LUN (Linux only).
Options:
   [int]             - Interactive mode (Menu driven).
```

Procedure

1. Log in to the CIMS as `root`.

2. Use SSH to log in to the CLFS node.

```
cims1# ssh clfs-mdc001
Last login: Fri Jul 25 14:51:19 2014 from cims1.cm.cluster
clfs-mdc001#
```

3. Enter `qaucli` to start the CLI in interactive mode:

```
clfs-mdc001# /opt/QLogic_Corporation/QConvergeConsoleCLI/qaucli
Loading iSCSI Data ...
       QConvergeConsole
       CLI - Version 1.1.0 (Build 51)

   Main Menu

   1: Adapter Information
   2: Adapter Configuration
   3: Adapter Updates
   4: Adapter Diagnostics
   5: Adapter Statistics
   6: Refresh
   7: Help
   8: Exit
      Please Enter Selection: 1
```

4. Enter a number to select a menu item. Select 1 to display FC adapter information.

```
 FC Adapter Information
```

```
   1:  FC Adapter Information
   2:  FC Port Information
   3:  FC VPD Information
   4:  FC Target/LUN Information
(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)
       Please Enter Selection: 1

   Adapter Information

    1: HBA Model: QLE2562 SN: BFD1315J98783
         Port   1 WWPN: 21-00-00-24-FF-4E-30-FC Online
         Port   2 WWPN: 21-00-00-24-FF-4E-30-FD Online
    2: HBA Model: QLE2562 SN: BFD1336P39125
         Port   1 WWPN: 21-00-00-24-FF-57-8F-7E Online
         Port   2 WWPN: 21-00-00-24-FF-57-8F-7F Online
    3: HBA Model: QLE2562 SN: LFD1108L27163
         Port   1 WWPN: 21-00-00-24-FF-26-67-20 Online
         Port   2 WWPN: 21-00-00-24-FF-26-67-21 Online
```

10 Configure RDAC Kernel Parameter for the ESF Software Image

Prerequisites

The CLFS must be fully configured in Bright, boot from an ESF software image and include FC channel adapters.

About this task

When an OSS or MDS node boots from an ESF software image, the `scsi_dh_rdac` driver is not loaded at the correct time. This causes nodes that are attached to storage via FC channel adapters to encounter I/O errors, which (if enough LUNs are present) can significantly slow boot times. This condition can be corrected by preloading the RDAC module using a kernel parameter before the system starts the `qla2xxx` FC module.

Procedure

1. Log in to the CIMS node as `root`.

2. Clone the production ESF image to a new software image (named `ESF-XX-2.3.0` in this example.

```
cims1# cd /cm/images
cims1# cp -pr ESF-XX-2.3.0 ESF-XX-2.3.0-FC
```

> **IMPORTANT:** The time required to clone a software image in Bright depends on the image size. Cloning a minimal image (operating system only) completes in 5 to 10 minutes. A fully configured image can require between 30 to 75 minutes. The Bright clone operation spawns a background process that does not prevent you from rebooting a node or performing other configuration changes to software image before the image is fully cloned. The best practice is to copy the software image in `/cm/images` from the UNIX prompt, wait for the prompt to return, then clone the software image in Bright before proceeding.

3. Start `cmsh` and clone the `ESF-XX-2.3.0` software image to `ESF-XX-2.3.0-FC`.

```
cims1# cmsh
[cims1]% softwareimage
[cims1->softwareimage]% list
Name (key)                  Path                                      Kernel Version
--------------------------  ----------------------------------------- -------------------------
ESF-XX-2.3.0                 /cm/images/ESF-XX-2.3.0-201509031326     2.6.32-358.18.1.el6.x86_64
ESL-XC-1.3.0                /cm/images/ESL-XC-1.3.0                   3.0.74-0.6.8-default
ESL-XE-2.1.0-201309042109   /cm/images/ESL-XE-2.1.0-201309042109      2.6.32.59-0.7-default
default-image               /cm/images/default-image                  3.0.80-0.5-default
default-image.previous      /cm/images/default-image.previous         3.0.80-0.5-default

[cims1->softwareimage]% clone ESF-XX-2.3.0-201509031326 ESF XX-2.3.0-FC
[cims1->softwareimage*[ESF-XX-2.3.0-FC*]]% commit
[cims1->softwareimage[ESF-XX-2.3.0-FC]]%
```

4. Set kernel parameters to `rdloaddriver=scsi_dh_rdac`.

```
[cims1->softwareimage[ESF-XX-2.3.0-FC]]% set kernelparameters rdloaddriver=scsi_dh_rdac
[cims1->softwareimage*[ESF-XX-2.3.0-FC*]]% commit
```

5. To use the `cmgui` to set kernel parameters, select the software image from the **RESOURCES** tree, then select **Settings**, and enter the kernel parameter `rdloaddriver=scsi_dh_rdac` in the **Kernel Parameters** field.

11 Configure Lustre on CLFS Nodes

Cray recommends an MGT volume size of 1GiB.

Configuring Lustre on CLFS nodes involves the following tasks:

11.1 Configure Lustre `fs_defs` File for CLFS Nodes

Prerequisites

The CLFS node must be fully configured in Bright and boot from an ESF software image.

About this task

After the CLFS node is fully configured in Bright and boots and ESF software image, configure Lustre file systems using the `fs_defs` file and `luster_control` command.

Procedure

1. Log in to the CIMS node as `root`.

2. Copy the `example.fs_defs` file in `/opt/cray/esms/cray-lustre-control-XX/default/etc/` to `/opt/cray/esms/cray-lustre-control-XX/etc/scratch.fs_defs` and prevent it from being overwritten during software updates. Lustre file system names must be eight characters or less. This procedure creates a file system called `scratch`.

```
cims1# mkdir -p /opt/cray/esms/cray-lustre-control-XX/etc
cims1# cd /opt/cray/esms/cray-lustre-control-XX/etc
cims1# cp -p /opt/cray/esms/cray-lustre-control-XX/default/etc/example.fs_defs scratch.fs_defs
```

3. Display the Lustre network identifier (NID) map information to include in the `scratch.fs_defs` file.

```
cims1# lustre_control dump_nid_map -w esfs-mds00[1-2],esfs-oss00[1-4]
Performing 'dump_nid_map' from cims1 at Thu Jan 31 14:47:49 CST 2014

Hostname to LNET nid mapping for the nodes:
esfs-mds001,esfs-mds002,esfs-oss001,esfs-oss002,esfs-oss003,esfs-oss004
nid_map: nodes=esfs-mds00[1-2] nids=10.149.0.[1-2]@o2ib
nid_map: nodes=esfs-oss00[1-4] nids=10.149.0.[3-6]@o2ib
```

4. Configure device names in the `scratch.fs_defs` file.

```
cims1# ssh clfs-mdc001
Last login: Fri Apr 19 11:08:03 2014 from cims1.cm.cluster
[root@clfs-mdc001 by-id]# ls -l /dev/disk/by-id
total 0
lrwxrwxrwx 1 root root 9 Apr 18 21:14 scsi-3600508e000000000faef8330f46c9c09 -> ../../sda
lrwxrwxrwx 1 root root 10 Apr 18 21:14 scsi-3600508e000000000faef8330f46c9c09-part1 -> ../../sda1
lrwxrwxrwx 1 root root 10 Apr 18 21:14 scsi-3600508e000000000faef8330f46c9c09-part2 -> ../../sda2
lrwxrwxrwx 1 root root 9 Apr 18 16:43 scsi-3600a0b800026cfe400002dfd4becf544 -> ../../sdb
lrwxrwxrwx 1 root root 9 Apr 18 21:14 wwn-0x600508e000000000faef8330f46c9c09 -> ../../sda
lrwxrwxrwx 1 root root 10 Apr 18 21:14 wwn-0x600508e000000000faef8330f46c9c09-part1 -> ../../sda1
lrwxrwxrwx 1 root root 10 Apr 18 21:14 wwn-0x600508e000000000faef8330f46c9c09-part2 -> ../../sda2
lrwxrwxrwx 1 root root 9 Apr 18 16:43 wwn-0x600a0b800026cfe400002dfd4becf544 -> ../../sdb
```

   Use either the disk device names or the persistent device names (recommended). For a multipath configuration, device names should link to a logical alias under `/dev/mapper/`*`alias`*. Using the `/dev/mapper/`*`alias`* allows the target definition to be condensed considerably for large file systems using multipath. For example, a set of 12 LUNs whose alias contains 12 separate `ost: dev=xxx node=xxx...` lines can be represented as: `ost: node=esfs-oss[1,2] dev=/dev/mapper/lun[0-11] index=[0-11]`.

   This method assigns the even numbered LUNs to `esfs-oss1` and the odd numbered LUNs to `esfs-oss2`. This is configured in the node finalize script (`site.esf_finalize.sh`) either for the node, or the node category (such as `esfs-even-`*`filesystem`* for use with esfsmon 2.1.0).

   Multipath configuration device names can also be configured as a `dm-uuid` path name such as `/dev/disk/by-id/dm-uuid-mpath-wwid`. To configure device names in a non-multipath configuration, use a device mapper to assign names to the disk devices. The following command lists disks by ID. The persistent device name for `sdb` is `scsi-3600a0b800026cfe400002dfd4becf544`. Enter the value in the device configuration section of `scratch.fs_defs`. Either option is a link that points to the `/dev/dm-`*`x`* device.

5. Edit the `scratch.fs_defs` file to set values.

```
cims1# vi /opt/cray/esms/cray-lustre-control-XX/etc/scratch.fs_defs
```

6. Modify the `fs_name`, `nid_map`, `mdt`, `mgt`, `ost`, and `stripe_count` settings in the `scratch.fs_defs` file.

```
# file system name - must be 8 characters or less
fs_name: scratch
```

7. Enter the NID map information obtained in *3* on page 65.

8. Enter the `mdt` and `mgt` definitions and disk device name(s) for your system obtained from *4* on page 66. Include the failover node definition (`fo_node`) if configuring a failover system. If there is no failover node definition, remove this line. In this example, the metadata target (MDT) and management target (MGT) functions share the same CLFS node (`clfs-mdc001`).

```
## MDT
## MetaData Target
mdt: node=clfs-mdc001
dev=/dev/mapper/mdt0
fo_node=clfs-mdc002
## MGT
## Management Target
mgt: node=clfs-mdc001
dev=/dev/mapper/mdt0
fo_node=clfs-mdc002
```

9. Enter the `ost` definitions and disk device name(s) for your system obtained from *this step*. Include the failover node definition (`fo_node`) if configuring a failover system. If there is no failover node definition, remove this line. Include the target index value to simplify management of a large number of targets.

```
## OST
## Object Storage Target(s)
ost: node=clfs-oss001
dev=/dev/mapper/ost0
fo_node=clfs-oss002
index=0
ost: node=clfs-oss002
dev=/dev/mapper/ost1
fo_node=clfs-oss001
index=1
ost: node=clfs-oss001
dev=/dev/mapper/ost2
fo_node=clfs-oss002
index=2
ost: node=clfs-oss002
dev=/dev/mapper/ost3
fo_node=clfs-oss001
index=3
ost: node=clfs-oss001
dev=/dev/mapper/ost4
fo_node=clfs-oss002
index=4
ost: node=clfs-oss002
dev=/dev/mapper/ost5
fo_node=clfs-oss001
index=5
```

10. Exit and save the `scratch.fs_defs` file.

11.2 Configure Lustre File Systems on CLFS Nodes

Prerequisites

Managed CLFS node in Bright.

Procedure

1. Log in to the CIMS node as `root` and install the Lustre file system definitions.

```
cims1# cd /opt/cray/esms/cray-lustre-control-XX/etc
cims1# lustre_control install scratch.fs_defs
Performing 'install' from cims1 at Mon Apr 22 14:38:42 CDT 2015
Parsing file system definitions file: scratch.fs_defs
Parsed file system definitions file: scratch.fs_defs
The 'scratch' file system definitions were successfully installed!
```

2. Format the Lustre file system (`scratch`). You must enter `y` to enable the `lustre_control` command to reformat the file system.

```
cims1# lustre_control reformat -f scratch
Performing 'reformat' from cims1 at Mon Apr 22 14:42:52 CDT 2015
About to reformat all targets for the following file system(s):
scratch
Continue? (y|n|q) y
```

3. Start the Lustre file system `scratch`.

```
cims1# lustre_control start -p -f scratch
```

4. Check the status of the `scratch` file system.

```
cims1# lustre_control status -f scratch
Performing 'status' from cims1 at Mon Apr 22 14:44:14 CDT 2014

File system: scratch
Device              Host         Mount           OST Active     Recovery Status
MGS                 clfs-mdc001  Mounted         N/A            Unknown
MGS*                clfs-mdc002  Unmounted       N/A            N/A
scratch-MDT0000     clfs-mdc001  Mounted         N/A            INACTIVE
scratch-MDT0000*    clfs-mdc002  Unmounted       N/A            N/A
scratch-OST0000     clfs-oss001    Mounted         Active         INACTIVE
scratch-OST0000*    clfs-oss002    Unmounted       Active         N/A
. . .
```

11.3 Mount a Lustre File System on a CDL Node

About this task

CLFS nodes typically share a Lustre filesystem to CDL nodes and to Cray internal login nodes to create a common development environement. Use this procedure to mount a shared Lustre file system on a CDLnode.

Procedure

1. Log in to the CIMS node as `root`.

2. Use the `chroot` shell to make a mount point in the CDL software image (in this example `ESL-XC-2.2.0-20140116037`).

```
cims1# chroot /cm/images/ESL-XC-2.2.0-20140116037
cims1:/> mkdir -p /lus/scratch
cims1:/> exit
```

3. Start `cmsh`, switch to `device` mode, and update the CDL node (`eslogin001` with the modified `ESL-XC-2.2.0-20140116037` software image.

```
cims1# cmsh
[cims1]% device
[cims1->device]%
[cims1->device]% imageupdate -n eslogin001
Performing dry run (use synclog command to review result, then pass -w to perform real update)...
[cims1->device]%
...
[cims1->device]% imageupdate -w -n eslogin001
```

4. Enter **`quit`** to exit `cmsh`.

5. Use SSH to log in to the CDL node (`eslogin001`).

```
cims1# ssh eslogin001
 Last login: Fri Apr 15 11:28:06 2013 from cims1.cm.cluster
eslogin001#
```

6. Mount the file system on `eslogin001`. Add the following line to `/etc/fstab`, where 10.149.0.2 is the IP address of the `clfs-mdc001` node on `ib-net`. The `scratch.fs_defs` file defines a failover MDS, therefore, inlcude both the primary and failover MDS nodes in the `fstab` entry:

```
eslogin001# vi /etc/fstab
10.149.0.2@o2ib:10.149.0.3@o2ib:/scratch /lus/scratch lustre
rw,flock,lazystatfs 0 0
```

7. Exit the editor and save the file.

8. Enter the following command on `eslogin001` to mount the file system.

```
eslogin001# mount -t lustre -o rw,flock,lazystatfs 10.149.0.2@o2ib:/scratch /
lus/scratch
```

9. Verify the `scratch` file system is mounted.

```
eslogin001# mount | grep lustre
10.149.0.1@o2ib:/scratch on /lus/scratch type lustre (rw,flock,lazystatfs)

eslogin001# df /lus/scratch
Filesystem              1K-blocks   Used   Available  Use% Mounted on
10.149.0.1@o2ib:/scratch 15611116960 922104 14829205576  1% /lus/scratch
```

11.4 Mount a Lustre File System on a CLE System

Prerequisites

A managed CLFS node and a CDL node that share an InfiniBand network with Cray system LNET router nodes.

About this task

CLFS nodes typically share a Lustre file system to CDL nodes and to Cray internal login nodes running Cray Linux Environment (CLE) to create a common development environement. Use this procedure to mount a shared Lustre file system on a CLE system.

Procedure

1. Add a `node_class` entry for LNET router nodes to `CLEinstall.conf`. The `node_class` entry will be added to the `/etc/hosts` file when `CLEinstall` is run. In this example, the LNET routers node IDs (NIDs) are `2` and `30`.

   a. Log in to the SMW as `root`.

   b. Edit the `/home/crayadm/install.xtrel/CLEinstall.conf` file and add an `lnet` class created with nid `2` and `30` as members of that class. After running `CLEinstall`, the new class and its members will be in `/etc/opt/cray/sdb/node_classes` on the `bootroot` and `sharedroot` file systems.

   ```
   smw# vi /home/crayadm/install.xtrel/CLEinstall.conf

   node_class[1]=lnet 2 30
   ```

   c. Save the file and exit.

2. Run `CLEinstall` using the same command line options as if you were performing a software update. `CLEinstall` makes the configuration change to the `node_classes` file. After running `CLEinstall`, boot the CLE system before continuing.

3. Make a mount point in the default view of the `sharedroot`.

```
smw# ssh boot
boot# xtopview
default/:/ # mkdir -p /lus/scratch
```

4. Add the following parameter options to `/etc/modprobe.conf.local` for login nodes and all service nodes. Note that the line containing `options lnet networks=gni` is commented. The setting of `10.149.1.*` is the site IB network address, but the network must be within 10.149.0.0/16.

```
default:/ # vi /etc/modprobe.conf.local
#options lnet networks=gni
options lnet ip2nets="gni0 10.128.*.*; o2ib 10.149.*.*"
options lnet routes="gni0 10.149.1.[3,31]@o2ib; o2ib [2,30]@gni0"
### LNET options
options lnet check_routers_before_use=1
options lnet avoid_asym_router_failure=1
options lnet dead_router_check_interval=60
options lnet live_router_check_interval=60
options lnet router_ping_timeout=50
```

5. Optional: Cray recommends the following parameters be set for CLFS nodes in `modprobe.conf.local`.

```
## o2iblnd parameters
# Increase
options ko2iblnd timeout=100
# Disable peer health
options ko2iblnd peer_timeout=0
# Decrease
options ko2iblnd keepalive=30
# Increase
options ko2iblnd credits=2048
# Increase
options ko2iblnd ntx=2048
# Increase
options ko2iblnd peer_credits=126
# Increase
options ko2iblnd concurrent_sends=63
```

6. Enter **`exit`** to exit `topview`.

7. Create the `lnet` class-specialized files.

```
boot-p1#  -c xtopviewlnet
class/lnet:/ # ls -l /etc/modprobe.conf.local
lrwxrwxrwx 1 root root 45 Feb 1 08:53 /etc/modprobe.conf.local -> /.shared/base/
default/etc/modprobe.conf.local

class/lnet:/ # xtspec -c lnet /etc/modprobe.conf.local
class/lnet:/ # ls -l /etc/modprobe.conf.local
lrwxrwxrwx 1 root root 48 Feb 1 09:06 /etc/modprobe.conf.local -> /.shared/base/
class/lnet/etc/modprobe.conf.local
class/lnet:/ # vi /etc/modprobe.conf.local
```

8. Add the local extensions to the `/etc/modprobe.conf.local` file. The two examples below assume that `ib0` on the LNET router nodes is connected to the DMP IB switch. If `ib1` is used, there is a different format.

   For `ib0`:

```
### LNET interfaces
options lnet ip2nets="gni0 10.128.*.*; o2ib 10.149.*.*"
### LNET routes for esFS
options lnet routes="gni0 10.149.1.[3,31]@o2ib; o2ib [2,30]@gni0"
```

For `ib1`:

```
### LNET interfaces
options lnet ip2nets="gni0 10.128.*.*; o2ib(ib1) 10.149.*.*"
### LNET routes for esFS
options lnet routes="gni0 10.149.1.[3,31]@o2ib; o2ib [2,30]@gni0"
```

9. Create node-specialized files for all LNET routers. Repeat these substeps for each LNET router.

a. Run `xtopview` to get the node view of the `sharedroot` for NID 2.

```
boot# xtopview -n 2
```

b. Create and edit the interface file (example shows `ifcfg-ib0`) for NID `2` with IP address 10.149.1.3. This IP address is the IP address of the `clfs-mdc001` node on the `ib-net`.

```
node/2:/ # touch /etc/sysconfig/network/ifcfg-ib0
node/2:/ # xtspec -n 2 /etc/sysconfig/network/ifcfg-ib0
node/2:/ # vi /etc/sysconfig/network/ifcfg-ib0
```

An example `ifcfg-ib0` file follows:

```
BOOTPROTO='static'
IPADDR=10.149.1.3
NETMASK=255.255.0.0
STARTMODE='onboot'
USERCONTROL='no'
MTU=2044
IPOIB_MODE='connected'
```

c. Create and edit the `/etc/sysconfig/infiniband` file.

```
node/2:/ # touch /etc/sysconfig/infiniband
node/2:/ # xtspec -n 2 /etc/sysconfig/infiniband
node/2:/ # vi /etc/sysconfig/infiniband
```

d. Change the following two variables:

```
ONBOOT=no
SRP_LOAD=yes
```

To:

```
ONBOOT=yes
SRP_LOAD=no
```

e. Enter **`exit`** to exit `xtopview`.

10. Update the bootimage so that compute nodes mount the Lustre file system.

a. From the SMW, update the bootimage `/etc/modprobe.conf`. This example uses the `p1` partition of a CLE system.

```
smw# cd /opt/xt-images/templates/default-p1
smw:/opt/xt-images/templates/default-p1 # vi etc/modprobe.conf
```

b. Make the following changes to the `modprobe.conf` file.

```
#options lnet networks=gni
### LNET options
options lnet check_routers_before_use=1
options lnet avoid_asym_router_failure=1
options lnet dead_router_check_interval=60
options lnet live_router_check_interval=60
options lnet router_ping_timeout=50

### LNET interfaces
options lnet ip2nets="gni0 10.128.*.*; o2ib 10.149.*.*"

### LNET routes for esFS
options lnet routes="gni0 10.149.1.[3,31]@o2ib; o2ib [2,30]@gni0"
```

c. Make a mount point for the bootimage.

```
smw:/opt/xt-images/templates/default-p1 # mkdir -p lus/scratch
```

d. Update `/etc/fstab` for bootimage.

```
smw:/opt/xt-images/templates/default-p1 # cd etc
smw:/opt/xt-images/templates/default-p1/etc # cp -p fstab fstab.orig
smw:/opt/xt-images/templates/default-p1/etc # vi fstab
```

e. Add the following line to `/etc/fstab`.

```
10.149.0.2@o2ib:10.149.0.3@o2ib:/scratch /lus/scratch lustre rw,flock 0 0
```

f. Exit and save the `/etc/fstab` file.

g. Rebuild the bootimage. This example uses a script created for the `BLUE` system set.

```
smw# /var/opt/cray/install/shell_bootimage_BLUE.sh -c
```

11. This step requires that the `/etc/modprobe.d/cray.conf` file exist in the CLFS node software image before proceding.

a. Log in to the CIMS node as `root`.

b. Use the `chroot` shell to modify the CLFS node software image (`/cm/images/ESF-XX-2.3.0-201509031326` for example).

```
cims1# cd /cm/images
cims1# chroot ESF-XX-2.3.0-201509031326
```

a. Determine if the `cray.conf` file exists in `/etc/modprobe.d`.

```
[root@cims1 /]# cd /etc/modprobe.d
[root@cims1 /]# ls
blacklist-kvm.conf cray-ib_srp.conf cray-ost.conf ib_ipoib.conf
blacklist.conf cray-lnd.conf dist-alsa.conf ib_sdp.conf
bright-cmdaemon.conf cray-lnet.conf dist-oss.conf mlx4_en.conf
cray-aliases.conf cray-mlx4.conf dist.conf openfwwf.conf

[root@cims1 /]# touch cray.conf
[root@cims1 /]# ls
blacklist-kvm.conf cray-lnd.conf dist-alsa.conf mlx4_en.conf
blacklist.conf cray-lnet.conf dist-oss.conf openfwwf.conf
bright-cmdaemon.conf cray-mlx4.conf dist.conf
cray-aliases.conf cray-ost.conf ib_ipoib.conf
cray-ib_srp.conf cray.conf ib_sdp.conf

[root@cims1 /]# exit
```

12. Update the CLFS category finalize script for `esfs-odd-`*`filesystem`*, `esfs-even-`*`filesystem`*, and `esfs-failed-`*`filesystem`* categories.

a. Edit the `site.esf_finalize.sh` script.

```
cims1# cd /opt/cray/esms/cray-es-finalize-scripts-XX/etc
cims1# vi site.esf_finalize.sh
```

b. Add the entry before the `exit 0` line. This example uses the IP address 10.149.1.3 for NID 2 and 10.149.1.31 for NID 30.

```
echo "options lnet routes=\"gni0 10.149.1.[3,31]@o2ib\" >> /localdisk/etc/
modprobe.d/cray.conf
```

c. Exit and save the `site.esf_finalize.sh` script.

d. Update each of the CLFS node categories in Bright with the new finalize script (`esfs-odd-`*`filesystem`*, `esfs-even-`*`filesystem`*, and esfs-failed-*`filesystem`*.

```
cims1# cmsh
[cims1]% category
[cims1->category]% use esfs-even-filesystem
[cims1->category[esfs-even-filesystem]]% set finalizescript /opt/cray/esms/cray-es-finalize-
scripts-XX/etc/site.esf_finalize.sh
[cims1->category*[esfs-even-filesystem*]]% commit
[cims1->category[esfs-even-filesystem]]%
```

e. Verify the finalize script.

```
[cims1->category[esfs-even-filesystem]]% get finalizescript
```

f. Enter **`quit`** to exit `cmsh`.

13. From the `eslogin001` root login window, unmount the Lustre clients and stop the Lustre file system.

```
eslogin001# umount /lus/scratch
```

14. From the CIMS node root login window, stop the Lustre file system (`scratch`).

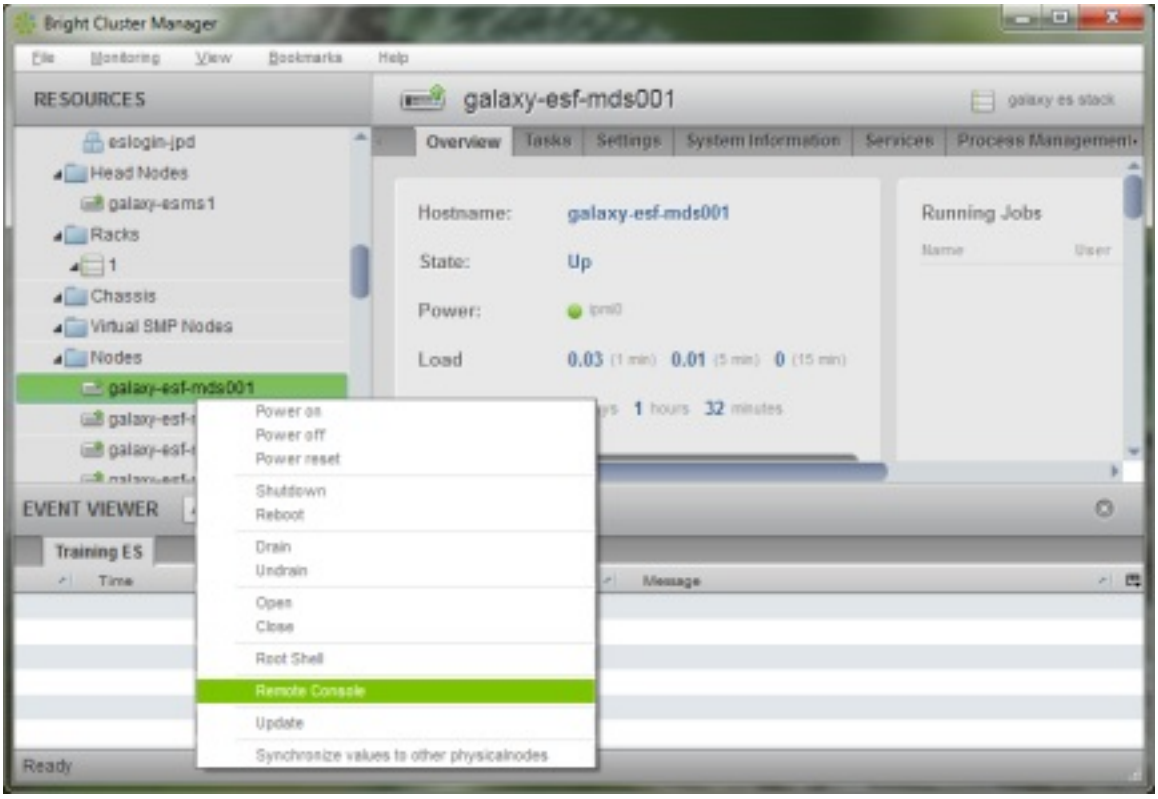
```
cims1# lustre_control status -f scratch
cims1# lustre_control stop -f scratch
```

15. Start `cmsh`, and reboot the MDS and OSS node(s).

```
cims1# cmsh
cims1# device
cims1# foreach clfs-mds001,clfs-mds002,clfs-oss001,clfs-oss002,clfs-oss003,clfs-oss004 (reboot)
clfs-mds001: Reboot in progress ...
clfs-mds002: Reboot in progress ...
```

16. Wait for each node to reboot before continuing. Use Bright `cmgui` to remote consoles for each node to monitor to reboot process.

Figure 14. Bright GUI Slave Node Remote Console



17. Enter **`quit`** to exit `cmsh`.

18. From the CIMS node, start the Lustre file system (`scratch`).

```
[cims1# lustre_control start -f scratch
Performing 'start' from cims1 at Tue Jul 29 10:30:14 CDT 2014
```

19. Either configure LDAP on MDS nodes or execute the following command on the `clfs-mdc001` node after starting Lustre file system to disable upcall and make the MGS/MDS use the UID/GID supplied by the client.

```
cims1# ssh clfs-mdc001
Last login: Tue Jul 29 10:30:14 2014 from cims1.cm.cluster
clfs-mdc001# lctl conf_param extlus-MDT0000.mdt.identity_upcall=NONE
```

20. Log out of the `clfs-mdc001` node and return to the CIMS node.

```
clfs-mdc001# exit
cims1#
```

21. Use `ssh` to log in to `eslogin001` and mount the Lustre `scratch` file system.

```
cims1# ssh eslogin001
eslogin001# mount -t lustre -o rw,flock,lazystatfs 10.149.0.2@o2ib:
10.149.0.3@o2ib:/scratch
/lus/scratch
```

22. Reboot the CLE system. The compute nodes will mount the Lustre file system as specified in the bootimage `/etc/fstab` file.

12 Configure LDAP on MDS nodes

About this task

LDAP is configured in the category finalize script for the MDS node. The `named`, `nslcd`, `nscd`, and `ldap` services are turned off by default. To enable these services, uncomment the corresponding lines in the finalize script.

> **IMPORTANT:** Files created or modified by a finalize script must be listed in the `excludelistupdate` exclude list for the category. Software updates will overwrite customized files if the files are not specified in an exclude list for the category. Customized files must also be specified in the `excludelistgrab`, and `excludelistgrabnew` exclude lists to prevent customized files from being copied to the CIMS node.

Procedure

1. Log in to the CIMS node as .
2. Edit the finalize script for the MDS node category (`site.esf_finalize.sh`).

   ```
   cims1# cd /opt/cray/esms/cray-es-finalize-scripts-XX/etc
   cims1# vi site.esf_finalize.sh
   ```

3. Enable the name service caching daemon (`nscd`) and the LDAP name service daemon `nslcd`. `named`, `nslcd`, `nscd`, and `ldap` are turned off in the default ESF software image.

   Add or uncomment the following commands in the `site.esf_finalize.sh` script:

   ```
   chkconfig nscd on
   service nscd start
   chkconfig nslcd on
   service nslcd start
   ```

4. Save `site.esf_finalize.sh` file and exit the editor.
5. Use the `chroot` shell to edit the ESF software image:

   ```
   cims1# cd /cm/images
   cims1# chroot imagename
   [root@cims1 /]#
   ```

6. Verify the LDAP service (`ldap`), is turned off in the ESF software image. The MDS should not act as an LDAP server. Use the following command to disable the LDAP server instance. The LDAP client configuration settings are typically in `/etc/passwd` and `/etc/nsswitch.conf`, and in the pluggable authentication modules (PAM).

```
[root@cims1 /]# chkconfig ldap off
```

7. Edit `/etc/openldap/ldap.conf` file, and uncomment (enable) the `TIMELIMIT 15` line:

```
[root@cims1 /]# cd /etc/openldap
[root@cims1 /]# cp ldap.conf ldap.conf.orig
[root@cims1 /]# vi ldap.conf
TIMELIMIT     15
```

8. Add/edit the following lines and enter the specific IP and base value settings for your site's information structure:

```
   URI ldap://aaa.bbb.ccc.ddd/
   BASE dc=somedomain,dc=somedomain,dc=com
   base ou=people,dc=somedomain,dc=somedomain,dc=com
```

9. Edit `/etc/nslcd.conf` file:

```
[root@cims1 /]# cd /etc
[root@cims1 /]# cp nslcd.conf nslcd.conf.orig
[root@cims1 /]# vi nslcd.conf
```

10. Add/edit the following lines and enter the specific IP and base value settings for your site's information structure:

```
   uri ldap://aaa.bbb.ccc.ddd/
   base dc=somedomain,dc=somedomain,dc=com
   base ou=people,dc=somedomain,dc=somedomain,dc=com
```

11. Exit the `chroot` shell.

12. Reboot all service nodes using the `ESF-XX-2.3.0-201509031326` software image.

13 Update Software on a CLFS Node

NOTE: The ESF XX-2.3.0 initial release supports initial software installations on CLFS servers for the following platforms: Cray XC30, Cray XE6, Cray XE6m and Cray XK6 systems. An upgrade path for this software release is currently not available, an initial install is required.

To upgrade previous versions of ESF software, refer to *Installing Lustre File System by Cray (CLFS) Software*, S-2521-C available from *CrayPort*.

13.1 Install ESF XX-2.3.0 Software Updates

Prerequisites

The ISO files or DVD release media required to install ESF XX-2.3.0 are named (ESF-XX-2.3.0-201509031326a06.iso), CentOS base operating system (CentOS-6.6-x86_64-bin-DVD1.iso), and Bright Cluster Manager updates for CentOS, (bright6.1-centos6u6.iso). This procedure requires that each ISO file is copied to the `/root/isos` directory on the CIMS node.

The ESF XX-2.3.0 software image will be named similar to ESF-XX-2.3.0-201509031326, and is a fresh install. It does not automatically copy site customizations made to the previously released ESF-XX-2.2.0 software image. Site customizations to the existing software image must be migrated the newly installed ESF-XX-2.3.0-201509031326 software image. Copy the relevant files and site customizations from the exising ESF-XX-2.2.0 image into the new software image that is created in this procedure (ESF-XX-2.3.0-201509031326, for example). Capture the following configuration settings and copy the following files from the existing ESF software image:

Multipath	`chkconfig multipath --list` `/etc/multipath.conf`
Kdump	`/etc/kdump.conf` `/boot/pxelinux.cfg/default`

Ldap	`chkconfig nscd --list`
	`chkconfig nslcd --list`
	`/etc/openldap/ldap.conf`
	`/etc/nslcd.conf`

Modprobe	`/etc/modprobe.d/`

Procedure

1. Log in to the CIMS node as `root`.

```
remote% ssh root@esms1
```

2. If necessary, create a /media/cdrom directory on the CIMS node.

```
cims1# mkdir -p /media/cdrom
```

3. Mount the release media on the CIMS node and copy each ISO file to /root/isos on the CIMS node.

   - If you have the release media on DVD, insert the DVD into the DVD drive and mount it to /media/cdrom. cims1

     ```
     cims1# mount /dev/cdrom /media/cdrom
     cims1# dd if=/dev/cdrom of=/root/isos/ISO_image
     ```

   - If you have an ISO file, copy the ISO file to /root/isos on the CIMS node.

     ```
     cims1# cp ISO_image /root/isos/
     ```

   The ISO files for ESF XX-2.3.0 (ESF-XX-2.3.0-201509031326a06.iso), CentOS base operating system (CentOS-6.6-x86_64-bin-DVD1.iso), and Bright Cluster Manager updates for CentOS, (bright6.1-centos6u6.iso) must reside in /root/isos directory on the CIMS node before continuing.

4. Mount the ESF-XX-2.3.0-201509031326a06.iso file to /media/cdrom.

```
cims1# mount /root/isos/ESF-XX-2.3.0-201509031326a06.iso /media/cdrom/
```

5. Create a release directory on the CIMS node in /root/release.

```
cims1# mkdir /root/release/
cims1# cp -pr /media/cdrom/* /root/release/ESF-XX-2.3.0-201509031326
```

6. Copy the new esfsmon 2.1.0 and lustre_control service node RPMs to the correct directory and remove the old versions. The old RPMs must be moved from /root/install.ESM/* and its sub directories in order for the new RPMs to be installed in the image.

```
cims1# mkdir /root/esfsmon.old
cims1# mv /root/install.ESM/cray-lustre-control-es* /root/esfsmon.old/
cims1# mv /root/install.ESM/cray-esfsmon-slave-XX* /root/esfsmon.old/
cims1# cd /root/release/ESF-XX-2.3.0-201509031326/
cims1# cp esm.updates/cray-lustre-control-es* /root/install.ESM/
cims1# cp esm.updates/cray-esfsmon-slave-XX-*  /root/install.ESM/
```

7. Move the CentOS 6.6 manifest to the esinstaller directory.

```
cims1# cp esm.updates/centos6.6.manifest /opt/cray/esms/cray-esinstaller-XX/default/sysconfig/
```

8. Run EFSinstall from /root/release/ESF-XX-2.3.0-201509031326/ to install the new CLFS software image. The software is installed to /cm/images/*imagename* on the CIMS node. By default, ESFinstall uses the release version string as the image name.

```
cims1# ./ESFinstall -v
./ESFinstall: exec to temp script: /tmp/ESFinstall.PID
Installation output will be captured in /var/adm/cray/logs/ESFinstall.PID.log.ESF-
XX-2.3.0-201509031326
Tue July 28 15:56:48 CST 2015: ESFinstall.PID Starting, Version XX-2.3.0-201507151419-a06.
```

   Examine the initial output and note the process ID (PID) of the ESFinstall process while the installation runs. ESFinstall creates a log file in /var/adm/cray/logs/ESFinstall.*PID*.log.*imagename* using this PID.

9. Make the necessary site customization to the software image as needed for multipath, kdump, and modprobe settings in the ESF-XX-2.3.0-201509031326 software image.

The ESF XX-2.3.0 software image will be named similar to ESF-XX-2.3.0-201509031326, and is a fresh install. It does not automatically copy site customizations made to the previously released ESF-XX-2.2.0 software image. Site customizations to the existing software image must be migrated the newly installed ESF-XX-2.3.0-201509031326 software image. Copy the relevant files and site customizations from the exising ESF-XX-2.2.0 image into the new software image that is created in this procedure (ESF-XX-2.3.0-201509031326, for example). Capture the following configuration settings and copy the following files from the existing ESF software image:

Multipath	`chkconfig multipath --list` `/etc/multipath.conf`
Kdump	`/etc/kdump.conf` `/boot/pxelinux.cfg/default`

Ldap	`chkconfig nscd --list`
	`chkconfig nslcd --list`
	`/etc/openldap/ldap.conf`
	`/etc/nslcd.conf`

Modprobe	`/etc/modprobe.d/`

10. Turn off the SRP daemon process in the ESF-XX-2.3.0-201509031326 software image.

```
cims1# chkconfig off srpd
cims1# exit
```

11. Configure the Bright categories for esfsmon 2.1.0 to use the new ESF-XX-2.3.0-201509031326 software image (`esfs-odd-`*`filesystem`*, `esfs-even-`*`filesystem`*, and `esfs-failed-`*`filesystem`*).

```
cims1# cmsh
[cims1]% category
[cims1->category]% use esfs-odd-filesystem
[cims1->category[esfs-odd-filesystem]]% set softwareimage ESF-XX-2.3.0-201509031326


[cims1->category]% use esfs-even-filesystem
[cims1->category[esfs-even-filesystem]]% set softwareimage ESF-XX-2.3.0-201509031326


[cims1->category]% use esfs-failed-filesystem
[cims1->category[esfs-failed-filesystem]]% set softwareimage ESF-XX-2.3.0-201509031326


[cims1->category*[clfs_catagory*]]% commit
```

12. Verify and modify the finalize script for each CLFS node category. The InfiniBand configuration file has changed in ESF XX-2.3.0. The following two lines should replace `sed … > /localdisk/etc/infiniband/openib.conf` lines in the finalize scripts.

```
sed -i -e "s/SRP_LOAD=no/SRP_LOAD=yes/" /localdisk/etc/rdma/rdma.conf
sed -i -e "s/NFSoRDMA_LOAD=yes/NFSoRDMA_LOAD=no/" /localdisk/etc/rdma/rdma.conf
```

Change the device `-i` and port `-p` options as needed for each InfiniBand port connected to storage to start the SRP daemon on that port. Add these after the entry to start the subnet manager on the corresponding port.

```
echo "/usr/sbin/srp_daemon -e -c -n -i mlx4_1 -p 1 -R 60 &" >> /localdisk/etc/rc.d/rc.local
```

```
[cims1->category]% use esfs-odd-filesystem
[cims1->category*[esfs-odd-filesystem]]% set finalizescript

[cims1->category]% use esfs-even-filesystem
[cims1->category*[esfs-even-filesystem]]% set finalizescript

[cims1->category]% use esfs-failed-filesystem
[cims1->category*[esfs-failed-filesystem]]% set finalizescript

[cims1->category*]]% commit
[cims1->category*]]% quit
```

13. Stop `esfsmon` monitoring.

```
cims1 # touch /var/esfsmon/esfsmon_suspend_filesystem
```

14. Stop the Lustre® file system.

```
cims1# lustre_control stop -f filesystem
```

15. Install updated esfsmon 2.1.0 and `lustre_control` RPMs on the CIMS node.

```
cims1# mkdir -p /root/updates/CRAY

cims1# cp  /root/release/ESF-XX-2.3.0-201509031326/esm.updates/cray-lustre-control-XX-*
 /root/updates/CRAY/
cims1# cp /root/release/ESF-XX-2.3.0-201509031326/esm.updates/cray-esfsmon-XX-* /root/updates/CRAY/
cims1# rpm -Uvh /root/updates/CRAY/cray-lustre-control-XX-2.3.3-1.000000.1256.0.x86_64.rpm
cims1# rpm -Uvh /root/updates/CRAY/cray-esfsmon-XX-2.0.0-1.040000.1251.0.x86_64.rpm
```

The ESF XX-2.3.0 version requires `cray-lustre-control` version `.1256`. If you cannot access the `lustre_control` command after the update, log out and back into the CIMS node to repair links and path name changes that occurred during the update.

> **NOTE:** If you have an HA CIMS configuration be sure to follow the Cray recommended procedure to install the RPMs on both CIMS. If `esms1` is active and `esms2` is passive, perform the following sequence.

1. Install RPMs on the current passive CIMS (`esms2`).
2. Run `cmha mackeactive` on the passive CIMS (`esms2`).
3. `esms2` is now the active CIMS node.
4. Test `esms2` to verify it is functioning as expected.
5. Install RPMs on the current passive CIMS node (`esms1`).
6. Run `cmha makeactive` on current passive CIMS node (`esms1`).
7. Test `esms1` to verify it functions as expected.

16. The ESF software image changed so the old SSH keys on the CIMS node no longer work. Delete old SSH keys for the CLFS nodes.

17. Reboot the MDS and OSS CLFS nodes to verify the new software image is functioning properly.

18. After each MDS and OSS has successfully rebooted, re-install the Lustre `fs_defs`.

```
cims1#  lustre_control install fs_defs_file
```

19. Start the Lustre file system.

```
cims1#  lustre_control start -f filesystem
```

20. Start `esfsmon` monitoring.

```
cims1 # rm /var/esfsmon/esfsmon_suspend_filesystem
```