# Installing, Configuring, and Managing SMW Failover on the Cray XC30™ System

S–0044–B

RECORD OF REVISION

S–0044–B Published July 2013 Corrected initial and upgrade installation procedures. Expanded content to include overview information and administration procedures. Supports the release of the Cray SMW High Availability Extension for SLES 11 SP2.

S–0044–A Published June 2013 Supports the release of the Cray SMW High Availability Extension for SLES 11 SP2 release.

# Contents

*Page*



## Procedures

*Page*



## Tables



## Figures

# Introduction [1]

This guide contains procedures for the Cray System Management Workstation (SMW) failover feature that is available with the Cray SMW High Availability Extension for SLES 11 SP2 release package, also called the *SMW HA package*. This feature is supported only for a Cray XC30 system with two second-generation high-end SMWs (also called *rack-mount SMWs*) that runs the SUSE Linux Enterprise (SLE) High Availability Extension. The following procedures are provided:

- Performing an initial installation of the Cray SMW HA release package. This package consists of the SUSE Linux Enterprise High Availability (HA) Extension 11 SP2 (SLEHA11SP2), and the Cray SMW HA software. See Chapter 2, Installing the Cray SMW HA Release Package on page 11.
- Updating your Cray SMW HA software; see Chapter 3, Updating Your Cray SMW and SMW HA Software on page 31.
- Managing the SMW HA system; see Chapter 5, Managing Your Cray SMW HA System on page 49.
- Troubleshooting the SMW HA system; see Chapter 6, Troubleshooting an SMW HA System on page 67.

This guide also includes an overview of the Cray SMW HA software; see Chapter 4, Overview on page 43.

The intended reader of this guide is a system administrator who is familiar with operating systems derived from UNIX.

**Note:** Some examples in this document are left-justified to fit better on the page. Left-justification has no special significance.

## 1.1 Related Publications

This document does not include the basic procedures for installing the Cray SMW software, configuring the boot RAID, and updating or upgrading the SMW or CLE software. Refer to the following Cray documents for this information.

- *Installing Cray System Management Workstation (SMW) Software* (S–2480)
- *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444), which is provided with your CLE release package

The following documents contain additional information that may be helpful:

- *SMW HA Release Errata* and the *SMW HA README*, which are provided with your SMW HA release package
- *Managing System Software for the Cray Linux Environment* (S–2393)

The following document from Novell, Inc., provides information on the SUSE Linux High Availability (SLEHA) Extension software, the Pacemaker Cluster Resource Manager (CRM), and related tools.

- *SUSE Linux Enterprise High Availability Extension High Availability Guide*

  **Note:** This document is available online at opensuse.org: http://doc.opensuse.org/products/draft/SLE-HA/SLE-ha-guide_sd_draft/

## 1.2 Distribution Media

The Cray SMW High Availability Extension for SLES 11 SP2 release distribution media includes the following DVD, which contains the Cray SMW HA software package.

- `Cray SMW HA SLE 11 SP2 (SMW SLEHA11SP2)`

  **Note:** You will also need the SMW software release media for an initial installation and some SMW upgrades. The Cray SMW HA SLEHA 11 SP2 release requires the following software:

  - SUSE Linux Enterprise Server version 11 Service Pack 2 (SLES 11 SP2) base operating system
  - SMW 7.0.UP03 release software

  For more information, see the *SMW Release Notes* and *SMW README*, which are provided with your SMW release package.

# Part I: Installation Guide

# Installing the Cray SMW HA Release Package [2]

Follow the procedures in this chapter to perform an initial software installation of the SMW HA release package.

**Important:** During these procedures, you will need to refer to the SMW release installation procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

The following conventions are used in this chapter:

- SMW1 or the `smw1` host name specifies the SMW that has been configured to be the active SMW during normal operation. In examples, the prompt `smw1: #` shows a command that runs on this SMW.
- SMW2 or the `smw2` host name specifies the SMW that has been configured to be the passive SMW during normal operation. In examples, the prompt `smw2: #` shows a command that run on this SMW.
- The `cray-smw` host name specifies the currently active SMW (which could be either `smw1` or `smw2`). This virtual host name is defined during installation and configuration.

## 2.1 Preparing to Install

### 2.1.1 Before You Start the SMW HA Release Installation

- Read the *SMW HA Release Errata* and the *SMW HA README* provided with your SMW HA SLEHA 11 SP2 release package for any additional installation-related requirements, corrections to these installation procedures, and other relevant information about the release package.
- Read the Field Notices (FNs) to identify whether there are any changes to this release package or the installation instructions.

- Read this section and ensure that the prerequisites are satisfied before beginning an initial SMW software installation; verify that the two SMWs are correctly cabled and network connections are in place (see Network Connections on page 12) and your configuration values are known (see Configuration Values on page 13).

  **Note:** The Cray SMW HA SLEHA 11 SP2 release requires the following software:

  – SUSE Linux Enterprise Server version 11 Service Pack 2 (SLES 11 SP2) base operating system

  – SMW 7.0.UP03 release software

- For an existing system: Before you install the SMW and SMW HA packages, back up your current SMW software.

  **Note:** Cray recommends that you remove old SMW log files to reduce the amount of time needed to back up your SMW.

### 2.1.2 Network Connections

In addition to the network connections required for single SMW, an SMW HA system uses `eth2` and `eth4` (on the second Ethernet card) for heartbeat connections to the other SMW.

Each SMW must be connected to the customer network through `eth0`, to the HSS network through `eth1`, to the boot node through `eth3`, and to the boot RAID through the Fibre Channel (FC) or SAS card. In addition, `eth2` and `eth4` must directly connect the two SMWs to each other as heartbeat monitoring channels.

Each SMW must have the following private network connections:

`eth0` – To the customer network

`eth1` – To the Hardware Supervisory System (HSS) network

`eth2` – To the other SMW (heartbeat connection)

`eth3` – To the boot node

`eth4` – To the other SMW (redundant heartbeat connection)

Refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480) for more information on the required network connections.

### 2.1.3 Configuration Values

The following IP addresses are set by default and are not site dependent.

**Table 1. Default IP Addresses**

| IP Address | Description |
|---|---|
| 10.1.0.1 | Primary boot RAID controller |
| 10.1.0.2 | Secondary boot RAID controller |
| 10.1.0.15 | Storage RAID controller |
| 10.1.1.1 | SMW, `eth1` – Virtual connection |
| 10.1.1.2 | SMW, `eth1` – Actual `eth1` connection for `smw1` |
| 10.1.1.3 | SMW, `eth1` – Actual `eth1` connection for `smw2` |
| 10.2.1.1 | SMW, `eth2` – Virtual primary heartbeat connection for SMW failover |
| 10.2.1.2 | SMW, `eth2` – Actual `eth2` heartbeat connection for `smw1` |
| 10.2.1.3 | SMW, `eth2` – Actual `eth2` heartbeat connection for `smw2` |
| 10.3.1.1 | SMW, `eth3` – Virtual connection |
| 10.3.1.2 | SMW, `eth3` – Actual heartbeat connection for `smw1` |
| 10.3.1.3 | SMW, `eth3` – Actual heartbeat connection for `smw2` |
| 10.3.1.254 | Boot node |
| 10.4.1.1 | SMW, `eth4` – Virtual redundant heartbeat connection for SMW failover |
| 10.4.1.2 | SMW, `eth4` – Actual `eth4` connection for `smw1` |
| 10.4.1.3 | SMW, `eth4` – Actual `eth4` connection for `smw2` |
| 127.0.0.1 | Localhost (loopback) |
| 225.0.0.1 | Multicast IP address for `eth4` |
| 226.0.0.1 | Multicast IP address for `eth2` |

In addition to the site-specific configuration values for a system with a single SMW (as shown in Table 2 of *Installing Cray System Management Workstation (SMW) Software*, S–2480), the following site-specific SMW and Dell Remote Access Controller (DRAC) values are required for an SMW HA system. You may find it helpful to record the actual values for your site.

**Table 2. Site-dependent Configuration Values for an SMW HA System**

| Description | Example | Actual Value |
|---|---|---|
| SMW virtual host name | `cray-smw` | |
| SMW virtual IP address | 173.31.73.165 | |
| SMW host name for first SMW | `smw1` | |
| SMW host name for second SMW | `smw2` | |
| DRAC host name on first SMW | `smw1-drac` | |
| DRAC IP address on first SMW | 172.31.73.77 | |
| DRAC host name on second SMW | `smw2-drac` | |
| DRAC IP address on second SMW | 172.31.73.79 | |

## 2.1.4 Passwords

The following default account names and passwords are used in the SMW HA software installation process. Cray recommends that sites change these default passwords after completing the installation.

**Table 3. Default Passwords for an SMW HA System**

| ID | Default Password |
|---|---|
| `root` on `smw1` | `initial0` |
| `root` on `smw2` | `initial0` |
| `hacluster` (for logging in to `crm_gui`) | `initial0` |
| `root` on DRAC for `smw1` (for example, `drac-smw2`) | `initial0` |

| ID | Default Password |
|---|---|
| `root` on DRAC for `smw2` (for example, `drac-smw2`) | `initial0` |
| `stonith-1` resource | `initial0` |
| `stonith-2` resource | `initial0` |

**Important:** All of these passwords **must** be the same. When you change the SMW `root` password after installation (as recommended in *Managing System Software for the Cray Linux Environment*, S–2393), you must also change the passwords for `hacluster`, the DRACs, and the `stonith` resources.

## 2.2 Installing the First SMW with Operating System and Cray SMW Software

The following procedure summarizes the steps for installing the SMW base operating system and Cray SMW software release package on the first SMW in an SMW HA cluster.

**Note:** During this procedure, you will need to refer to the SMW release installation procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

**Procedure 1. Installing the first SMW with operating system and Cray SMW software**

**Note:** Either SMW can be installed as the first SMW. In this document, this system is also called the *primary SMW* or `smw1`.

1. Prepare the SMW HA system for installing the SLES 11 SP2 operating system and SMW software, as described in *Installing Cray System Management Workstation (SMW) Software* (S–2480) and Preparing to Install on page 11.
2. Follow the procedures in the *Installing Cray System Management Workstation (SMW) Software* (S–2480) to install the SMW base operating system, then install and configure the SMW 7.0.UP03 software. Use the procedures and steps for a rack-mount SMW on a Cray XC30 system.

**Important:** Note the following differences from a system with a single SMW, as documented in *Installing Cray System Management Workstation (SMW) Software* (S–2480):

- Two Ethernet ports are used for heartbeat connections between the two SMWs: `eth2` (on the first Ethernet card) and `eth4` (on the second quad Ethernet card). Refer to Network Connections on page 12.

  **Note:** These ports are marked as "Reserved for SMW failover" in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

- When installing the SLES 11 SP2 operating system, only the boot disk should be connected to the SMW. All other internal disks should be uncabled. The boot RAID **must** be disconnected to prevent data corruption when installing the operating system.

  **Note:** The boot RAID must be set up for SMW HA before configuring the SMW HA software release package. For more information, see Configuring the Boot RAID for SMW HA on page 19.

- You do not need to back up the first SMW. That is, section 2.10, Backing Up Your Newly-installed SMW Software, is optional at this point. Installing and configuring the SMW HA software will make additional changes that should be backed up after the SMW HA configuration is complete.

**Note:** If you are converting an existing Cray system (with a single SMW) to an SMW HA cluster, you do not need to reinstall the operating system and full SMW software release. Instead, update the existing SMW to the SMW 7.0.UP03 release, shut down the Cray mainframe, then continue to the next section.

## 2.3 Installing the Second SMW with Operating System and Cray SMW Software

The following procedure summarizes the general steps for installing the SMW base operating system and Cray SMW software release package on the second SMW.

**Note:** During these procedures, you will need to refer to the SMW release installation procedures in the *Installing Cray System Management Workstation (SMW) Software* (S–2480).

**Procedure 2. Installing the second SMW with operating system and Cray SMW software**

**Note:** In this document, the second SMW is also called the *secondary SMW* or `smw2`.

1. Prepare the SMW HA system for installing the SLES 11 SP2 operating system and SMW software, as described in *Installing Cray System Management Workstation (SMW) Software* (S–2480) and Preparing to Install on page 11.

   **Important:** Note the following items:

   - The second SMW **must** run the same version of operating system and Cray SMW software.
   - As on the first SMW, when installing the SLES 11 SP2 operating system, only the boot disk should be connected to the SMW. All other internal disks should be uncabled. The boot RAID **must** be disconnected to prevent data corruption when installing the operating system.
   - As on the first SMW, two Ethernet ports are used for heartbeat connections between the two SMWs: `eth2` (on the first Ethernet card) and `eth4` (on the second quad Ethernet card). Refer to Network Connections on page 12.

     **Note:** These ports are marked as "Reserved for SMW failover" in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

2. Install the SMW base operating system, then install and configure the SMW 7.0.UP03 software on the second SMW.

   **Important:** When you install and configure the SMW software, you **must** skip several steps on the second SMW. This step summarizes the differences. For the detailed procedures, refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480).

   a. In section 2.1.1, Before You Start the SMW HA Release Installation on page 11, note that `eth2` and `eth4` are used for the heartbeat between the two SMWs. For SMW HA, refer to Network Connections on page 12.

   b. Skip section 2.1.2, Shutting Down the Cray System. Because this SMW is not yet connected to the Cray system, no system shutdown is required before installing the operating system and SMW software on the second SMW.

   c. Follow all applicable steps in section 2.2, Rack-mount SMW: Installing the SMW Base Operating System.

   d. Follow all applicable steps in section 2.4, Installing and Configuring the SMW Software Packages. There are no differences for the second SMW.

e. Skip section 2.5, Discovering the Cray System Hardware and Powering Up the Full System. Discovery and power-up was done when installing the first SMW.

f. Skip section 2.6, Confirming the SMW is Communicating with the System Hardware. This procedure was already done for the first SMW.

g. Change the default SMW passwords after completing installation, as directed in section 2.7, Changing Default SMW Passwords After Completing Installation.

   **Important:** Use the same passwords as for the first SMW.

h. For section 2.8, Setting Up the SUSE Firewall and IP Tables: If you performed the procedure to set up the SUSE firewall and IP table, set up the same configuration on the second SMW.

i. Skip section 2.9, Collecting Additional Data During a System Dump Session. After the SMW HA software is installed, the `/home/crayadm` directory will be on a shared disk mounted on the active SMW, so the `/home/crayadm/.xtdumpsys-plugin` file will be available to the active SMW.

j. As on the first SMW, section 2.10, Backing Up Your Newly-installed SMW Software, is optional at this point. Installing and configuring the SMW HA software will make additional changes that should be backed up after the SMW HA configuration is complete.

## 2.4 Installing the SMW HA Release Package

Follow this procedure on both SMWs to install the SMW HA release package.

**Important:** Before beginning this procedure, ensure that both SMWs are running the SMW base operating system and that the SMW software has been correctly installed, as described in the previous two sections.

**Procedure 3. Installing the SMW HA release package**

**Note:** The examples in this procedure show the host names `smw1` for the primary SMW and `smw2` for the secondary SMW.

1. Log on to one SMW as `root`. You can install on either `smw1` or `smw2` first.

2. Mount the Cray SMW HA release media on the SMW.

   **Note:** Because this step is done on both SMWs, the following examples show the generic prompt `smw: #`.

- If you have the release media on DVD, place the Cray SMW HA DVD into the DVD drive and mount it to `/media/cdrom`.

  ```
  smw: # mount /dev/cdrom /media/cdrom
  ```

- If you have an electronic version of the release media, mount the Cray SMW HA ISO to `/media/cdrom`.

  **Note:** The ISO file name depends on the release number, and installer version. The following command shows the generic ISO name `SMW-SLEHA11SP2`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP2b06.iso`). Substitute the actual ISO file name as specified in the SMW HA release information.

  For *path*, substitute the actual path to the ISO on your system (such as `/tmp`).

  ```
  smw: # mount -o loop,ro /path/SMW-SLEHA11SP2xnn.iso /media/cdrom
  ```

3. Go to the `/media/cdrom` directory and execute the `SMWHAinstall` script to install the Cray SMW HA release software on the SMW.

   ```
   smw: # cd /media/cdrom
   smw: # ./SMWHAinstall -v
   ```

4. Examine the initial output and note the process ID (PID) of the `SMWHAinstall` process. `SMWHAinstall` creates a log file in `/var/adm/cray/logs/SMWHAinstall.`*PID*`.log` using this PID.

5. Unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

   ```
   smw: # umount /media/cdrom
   smw: # eject
   ```

6. Repeat this procedure on the second SMW.

## 2.5 Configuring the Boot RAID for SMW HA

**Note:** Cray ships systems with much of this configuration completed. You may not have to perform all of the steps described in this section unless you are making changes to the configuration.

In addition to the LUNs required for a single SMW, a SMW HA system requires three LUNs on the boot RAID for shared storage for the MySQL database, log directory, and home directories (`/home`). This section describes how to move these directories from local disk to the shared boot RAID.

**Important:** You must set up the boot RAID with these LUNs before configuring the SMW HA cluster.

Table 4 shows the minimum partition sizes for these additional LUNs.

**Note:** A large system may require additional space for the shared directories. You should review the requirements of your system in order to determine the appropriate size for these LUNs.

**Table 4. Recommended Boot RAID LUN Sizes for SMW Failover**

| Purpose | Directory on SMW | Size |
|---|---|---|
| Shared MySQL HSS database | `/var/lib/mysql` | 150GB |
| Shared Log directory | `/var/opt/cray/disk/1` | 500GB |
| Shared home directories | `/home` | 500GB |

**Procedure 4. Configuring the boot RAID for SMW HA**

**Warning:** Any existing data on the boot RAID disks will be wiped out during installation. Proceed with care! Make sure you are using the correct disk names.

1. Configure your boot RAID, including the required LUNs for the SMW HA cluster as specified in Table 4. Follow the procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

   **Note:** The recommended boot RAID LUN configuration for a single SMW is included as a table in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

2. Record the device names for the shared directories. These device names are required when configuring the SMW HA cluster.

   **Important:** After initial partitioning, you should always address the storage via its persistent `/dev/disk/by-id/` name. Do not use the short `/dev/sd`*xx* name, which cannot uniquely identify the disk between reboots. For more information, see "About Persistent Boot RAID Device Names" in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

## 2.6 Configuring the Cluster

You will need the following information when configuring an SMW HA cluster:

- Virtual IP address for the SMW cluster
- Host names of two SMWs, such as `smw1` and `smw2`
- DRAC IP addresses for both SMWs

- Multicast IP addresses for the heartbeat connections (`eth2` and `eth4`), both of which use the port 1694
- Disk names on the boot RAID for Log, MySQL database, and `/home`

**Note:** For this information, see Table 1, Table 2, and Table 4.

**Procedure 5. Configuring the SMW HA cluster**

1. Log into both SMWs as `root`.
2. If you are using Virtual Network Computing (VNC) software to enable remote access to the SMW, shut down the VNC server.

   **Note:** In the following command, replace *display* with the necessary display number.

   ```
   smw1: # vncserver -kill :display
   ```

   For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).

3. Update the `ifcfg` files for `eth2` and `eth4` on both SMWs.

   a. On `smw1`, edit the `/etc/sysconfig/network/ifcfg-eth2` file to change the `NAME` value from `'eth2 Reserved'` to `'eth2 SMW HA Heartbeat Network 1'`. The changed file must have the following contents:

   ```
   smw1: # vi /etc/sysconfig/network/ifcfg-eth2
   BOOTPROTO='static'
   IPADDR='10.2.1.1/16'
   NAME='eth2 SMW HA Heartbeat Network 1'
   PREFIXLEN='16'
   STARTMODE='auto'
   USERCONTROL='no'
   ```

   b. Copy the `ifcfg-eth2` file to `ifcfg-eth4`.

```
smw1: # cp /etc/sysconfig/network/ifcfg-eth2 /etc/sysconfig/network/ifcfg-eth4
```

   c. Edit the `/etc/sysconfig/network/ifcfg-eth4` file to change the `IPADDR` value from `'10.2.1.1/16'` to `'10.4.1.1/16'`. Also change the `NAME` value to `'eth4 SMW HA Heartbeat Network 2'`. The changed file must have the following contents:

   ```
   smw1: # vi /etc/sysconfig/network/ifcfg-eth4
   BOOTPROTO='static'
   IPADDR='10.4.1.1/16'
   NAME='eth4 SMW HA Heartbeat Network 2'
   PREFIXLEN='16'
   STARTMODE='auto'
   USERCONTROL='no'
   ```

   d. Repeat step 3.a through step 3.c on `smw2`.

4. Update the cluster IP addresses.

   a. On `smw1`, execute the following command with `0` as the first argument.

      **Note:** In this command, replace *smw1* with the host name of the first (normally active) SMW, and replace *smw2* with the host name of the second (normally passive) SMW.

```
smw1: # /opt/cray/ha-smw/default/hainst/update_addresses 0 smw1 smw2
```

   b. On `smw2`, execute this command with `1` as the first argument:

      **Note:** In this command, replace *smw1* with the host name of the first (normally active) SMW, and replace *smw2* with the host name of the second (normally passive) SMW.

```
smw2: # /opt/cray/ha-smw/default/hainst/update_addresses 1 smw1 smw2
```

5. Initialize `smw1` as the active SMW.

   a. Execute the `sleha-init` command on the SMW that will be the active SMW during normal operation (usually `smw1`).

      **Important:** Run the `sleha-init` command on the active SMW only.

```
smw1: # sleha-init
```

      **Note:** You can safely ignore error messages about file `/etc/corosync/corosync.conf`. This file will be created after `sleha-init` completes.

```
awk: cmd. line:1: fatal: cannot open file `/etc/corosync/corosync.conf' for reading
(No such file or directory)"
```

b. As `sleha-init` runs, it prompts you for required information. Answer the following questions to configure the cluster.

**Note:** For the network address to bind to, specify the IP address of `eth2`. This port is used for the primary heartbeat connection.

```
Network address to bind to (e.g.:192.168.1.0): 10.2.1.0
Multicast address (e.g.:239.x.x.x): 226.0.0.1
Multicast port [5405]: 1694
Configure SBD:
.
.
.
Do you wish to use SBD? [y/N]: N
WARNING: Not configuring SBD - STONITH will be disabled.
.
.
.
Done (log saved to /var/log/sleha-bootstrap.log)
```

**Note:** If SMW HA has been configured before and you wish to rerun `sleha-init`, you will also be prompted to overwrite the existing configuration. In this case, answer the prompt `* - overwrite [y/N]?` with **`y`**.

c. Wait for `sleha-init` to finish (normally, about 1 or 2 minutes).

6. Join `smw2` as the passive SMW.

a. Execute the `sleha-join` command on the SMW that will be the passive SMW during normal operation (usually `smw2`).

**Important:** Execute the `sleha-join` command on the passive SMW only.

```
smw2: # sleha-join
```

b. Answer the following questions to join the passive SMW to the cluster. When asked for password, use the root password for the active SMW (`smw1`).

**Note:** In this command, replace *smw1* with the host name of the first (normally active) SMW.

```
IP address or hostname of existing node (active SMW): smw1

Password: root-password-for-smw1
```

7. Check the cluster status to verify that both `smw1` and `smw2` are online.

```
smw1 # crm_mon -1 | grep Online
Online: [ smw1 smw2 ]
```

8. On both SMWs, configure `eth4` as the redundant heartbeat channel.

   a. On `smw1`, execute `yast2` to open the **YaST2 Control Center**.

   ```
   smw1: # yast2 cluster
   ```

   **Note:** For the GUI version of YaST, either execute this command on the SMW console or connect via an `ssh` connection with X11 port forwarding.

   The cluster wizard starts and opens the cluster configuration window.

   b. In the left panel, select **Communication Channels**.

   c. In the right panel, check the **Redundant Channel** check box, then enter the following information to configure the redundant channel:

   - **Bind Network Address**: Enter **`10.4.1.0`**
   - **Multicast Address**: Enter **`225.0.0.1`**
   - **Multicast Port**: Enter **`1694`**

   d. Click the **Finish** button.

   e. Close the main YaST2 window to exit `yast2`.

   f. Repeat step 8.a through step 8.e on `smw2`.

9. Synchronize the passive SMW.

   ```
   smw1: # csync2 -xv
   ```

10. Synchronize the `ssh` host keys. This step makes both SMWs appear to have the same `ssh` host identity when someone connects to the virtual SMW host name or IP address.

    **Note:** In this command, replace *smw2* with the host name of the second (normally passive) SMW.

    ```
    smw1: # scp -p /etc/ssh/ssh_host_*key* root@smw2:/etc/ssh
    ```

    **Note:** Because this is the first `ssh` access for `root` on `smw2`, a message like the following one appears. Answer `yes` at the prompt.

    ```
    The authenticity of host 'smw2 (172.31.73.76)' can't be established.
    RSA key fingerprint is 5f:87:c3:a9:77:04:56:a7:36:6f:23:60:6c:21:77:b6.
    Are you sure you want to continue connecting (yes/no)? yes
    Warning: Permanently added ''smw2,172.31.73.76' (RSA) to
    the list of known hosts.
    ```

11. Load the `ha-smw` module on both SMWs.

    ```
    smw1: # module load ha-smw
    ```

    ```
    smw2: # module load ha-smw
    ```

    **Note:** After completing the SMW HA configuration, this module is loaded automatically when the root user logs in.

12. On the active SMW, edit the SMW HA configuration file, `/opt/cray/ha-smw/default/hainst/smwha_args` to configure the site-specific IP addresses and boot RAID disk names (defined in Procedure 4 on page 20). Replace the default contents with the values for your site.

```
smw1: # vi /opt/cray/ha-smw/default/hainst/smwha_args
--virtual_ip
172.30.49.160  ← Replace with virtual IP address for the SMW cluster
--log_disk_name
/dev/disk/by-id/scsi-360001ff0200211010bb7ab730c352a00-part5  ← Replace with "by-id" disk name
                                                                for shared log directory
--db_disk_name
/dev/disk/by-id/scsi-360001ff0200211010bb7ab730c352a00-part6  ← Replace with "by-id" disk name
                                                                for shared MySQL database
--home_disk_name
/dev/disk/by-id/scsi-360001ff0200211010bb7ab730c352a00-part7  ← Replace with "by-id" disk name
                                                                for shared /home directories
--drac_ip_active
172.30.49.164  ← Replace with DRAC IP address for active SMW
--drac_ip_passive
172.30.49.165  ← Replace with DRAC IP address for passive SMW
--passive_smw_hostname
smw2  ← Replace with host name of passive SMW
--verbose
```

For more information, see Configuring the Boot RAID for SMW HA on page 19 and the `SMWHAconfig`(8) man page.

13. Configure the SMW HA cluster on the active SMW.

    a. Change to the directory containing the `SMWHAconfig` command.

    ```
    smw1: # cd /opt/cray/ha-smw/default/hainst
    ```

    b. Execute `SMWHAconfig` with the modified configuration file as an argument.

    **Important:** Execute this command only on the active SMW.

    ```
    smw1: # ./SMWHAconfig @smwha_args
    ```

    **Note:** You must include the `@` character before the argument `smwha_args`.

    c. Because this is the first `ssh` access for `root` on `smw1`, a message like the following one appears. Answer `yes` at the prompt.

```
The authenticity of host 'smw1 (172.31.73.141)' can't be established.
RSA key fingerprint is 5f:87:c3:a9:77:04:56:a7:36:6f:23:60:6c:21:77:b6.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'smw1,172.31.73.141' (RSA) to the list of known hosts.
```

    d. When `SMWHAconfig` runs, it prompts for a password so that it can configure the SMW HA cluster and the DRAC; enter the root password for the SMW.

e. `SMWHAconfig` automatically loads the HA cluster configuration settings, then prompts you to commit (save) these changes. Enter **y** at the following prompt:

```
Do you still want to commit? y
```

**Caution:** If you execute `SMWHAconfig` to reinstall the SMW HA software on an existing SMW HA cluster, `SMWHAconfig` automatically detects if there is existing data on the shared RAID disks and prompts you for confirmation to reformat the RAID disks. If you answer `yes`, all data for the shared logs, MySQL database, and `/home` directory will be erased.

f. If necessary, examine the log file. `SMWHAconfig` creates a log file in `/opt/cray/ha-smw/default/hainst/SMWHAconfig.out`.

14. Reboot `smw1` and wait for the reboot to finish.

```
smw1: # reboot
```

15. Reboot `smw2` and wait for the reboot to finish.

```
smw2: # reboot
```

16. If necessary, restart the VNC server. For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).

## 2.7 Configuring Failover Notification

The SMW HA software includes a `Notification` resource that automatically sends email when a failover occurs.

**Procedure 6. Configuring failover notification**

1. Execute the `crm resource` command.

```
smw1: # crm resource param Notification set email address@yourdomain.com
```

**Note:** Only one email address is allowed. To send notifications to multiple addresses, you can create a group email alias that includes these email addresses.

2. Verify the setting.

```
smw1: # crm resource param Notification show email
address@yourdomain.com
```

If a failover occurs, the `Notification` resource sends several messages that are similar to the following examples.

```
From: root [mailto:root@smw.none]
Sent: Thursday, June 06, 2013 9:20 PM
To: Cray Cluster Administrator
Subject: ***Alert*** A Failover may have occurred. Please investigate! Migrating resource
away at Thu Jun 6 21:20:25 CDT 2013 from smw1

        ***Alert*** A Failover may have occurred. Please investigate! Migrating resource away
                    at Thu Jun  6 21:20:25 CDT 2013 from smw1

        Command line was:
        /usr/lib/ocf/resource.d//heartbeat/MailTo stop
```

```
From: root [mailto:root@smw.none]
Sent: Thursday, June 06, 2013 9:20 PM
To: Cray Cluster Administrator
Subject: ***Alert*** A Failover may have occurred. Please investigate! Takeover in progress
at Thu Jun 6 21:20:25 CDT 2013 on smw2

        ***Alert*** A Failover may have occurred. Please investigate! Takeover in progress
                    at Thu Jun  6 21:20:25 CDT 2013 on smw2

        Command line was:
        /usr/lib/ocf/resource.d//heartbeat/MailTo start
```

## 2.8 Verifying the Configuration

Use this procedure to check that the SMW HA cluster is up and running correctly.

**Note:** After finishing the configuration in the previous section, you must wait for 30 to 60 seconds for the cluster system to come up.

**Procedure 7. Verifying the SMW HA configuration**

1. Log into the virtual SMW as `root`, using the virtual SMW host name (such as `cray-smw`).

   **Note:** The examples in this procedure assume that `smw1` is the active SMW.

2. Find the active SMW by determining where the SMW HA cluster resources are running (such as the `hss-daemons` resource).

   **Note:** All resources except `stonith-2` run on the active SMW.

   ```
   smw1: # crm_mon -1 | grep hss-daemons
    hss-daemons    (lsb:rsms):     Started smw1
   ```

3. Display the cluster status.

```
smw1: # crm_mon -1
============
Last updated: Fri May 31 16:47:25 2013
Last change: Tue May 28 19:02:58 2013 by root via crm_resource on smw1
Stack: openais
Current DC: smw2 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw2 smw1 ]

 stonith-1      (stonith:external/ipmi):        Started smw1
 stonith-2      (stonith:external/ipmi):        Started smw2
 dhcpd  (lsb:dhcpd):    Started smw1
 cray-syslog    (lsb:cray-syslog):      Started smw1
 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP1     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP2     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP3     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP4     (ocf::heartbeat:IPaddr2):       Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 homedir        (ocf::heartbeat:Filesystem):    Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1

Failed actions:
    fsync_monitor_0 (node=smw2, call=11, rc=-2, status=Timed Out):
          unknown exec error
    ml-fs_start_0 (node=smw2, call=31, rc=1, status=complete): unknown error
```

**Note:** `crm_mon` may display the resources in a different order on your system.

4. Examine the `crm_mon` output. Verify that each resource has started by looking for `Started smw1` or `Started smw2`. Also look for any failed actions, as displayed at the end of the output.

5. If not all resources have started or if any failed actions are displayed, execute the `clean_resources` command on either SMW.

```
smw1: # clean_resources
Cleaning resources on node smw1
Cleaning resource on node=smw1 for resource=stonith-1
Cleaning resource on node=smw1 for resource=stonith-2
Cleaning resource on node=smw1 for resource=dhcpd
Cleaning resource on node=smw1 for resource=cray-syslog
Cleaning resource on node=smw1 for resource=ClusterIP
Cleaning resource on node=smw1 for resource=ClusterIP1
Cleaning resource on node=smw1 for resource=ClusterIP2
 ...
Cleaning resources on node smw2
Cleaning resource on node=smw2 for resource=stonith-1
Cleaning resource on node=smw2 for resource=stonith-2
 ...
Cleaning resource on node=smw2 for resource=Notification
```

**Note:** Wait several minutes for the cluster system to come back.

For more information, see the `clean_resources`(8) man page.

## 2.9 Backing Up Your Newly-installed SMW HA Software

After installing and configuring the system, back up your current SMW and SMW HA software. Follow the procedure in Appendix E, Rack-mount SMW: Creating a Bootable Backup Drive, in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

## 2.10 Changing Default SMW Passwords After Completing Installation

**Important:** The `hacluster`, `stonith`, and DRAC passwords **must** be the same as the SMW `root` password.

If you change the SMW `root` password after installation (as recommended in *Managing System Software for the Cray Linux Environment*, S–2393), you must also change the passwords for `hacluster`, both DRACs, and the two `stonith` resources to use the same password.

For more information, see Passwords on page 14.

**Procedure 8. Changing SMW, DRAC, and STONITH passwords**

1. Log on to the active SMW as `root`.
2. To change the SMW `root`, `hacluster`, and `stonith` passwords, execute the following commands:

   **Note:** Replace *drac-smw1* and *drac-smw2* with the host names of the DRAC for `smw1` and `smw2`, respectively.

   ```
   smw1:~# passwd root
   smw1:~# passwd hacluster
   smw1:~# crm resource param stonith-1 set passwd new-passwd
   smw1:~# crm resource param stonith-2 set passwd new-passwd
   ```

   **Important:** The `hacluster`, and `stonith` passwords **must** be the same as the SMW `root` password.

3. To change the DRAC passwords, see *Managing System Software for the Cray Linux Environment* (S–2393).

   **Important:** The DRAC passwords **must** be the same as the SMW `root` password.

# Updating Your Cray SMW and SMW HA Software [3]

Cray provides periodic updates to the SMW or SMW HA software releases. Follow the procedures in this chapter to install an SMW or SMW HA software update (or both). The procedures provided in this chapter do **not** change the base operating system version running on your SMW.

The following conventions are used in this chapter:

- The host name `smw1` specifies the SMW that has been configured to be the active SMW during normal operation. In examples, the prompt `smw1: #` shows a command that runs on this SMW.
- The host name `smw2` specifies the SMW that has been configured to be the passive SMW during normal operation. In examples, the prompt `smw2: #` shows a command that runs on this SMW.
- The host name `cray-smw` host name specifies the currently active SMW (which could be either `smw1` or `smw2`). This virtual host name was defined during initial installation.

## 3.1 Before You Start

**Note:** An SMW HA system requires more time to upgrade, as compared to a system with a single SMW, because you must install the software on both the primary and secondary SMW. Allot one hour of additional time to complete an update or upgrade.

- To determine whether there are any changes to the procedures in this chapter, read the *SMW SLEHA README* and *SMW SLEHA Errata* that are included in the SMW SLEHA update directory. Also read any Field Notices (FNs) related to kernel security fixes.
- Determine if you have made local changes to the file `/etc/csync2/csync2_cray.cfg`. The installation procedure saves local changes in a temporary file; you will restore those changes in a post-installation step.

- Before you install the update package, back up your current SMW and SMW HA software. For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480), Appendix E, Rack-mount SMW: Creating a Bootable Backup Drive.

## 3.2 Updating SMW Software

To update the SMW software, you will update the active SMW first, then update the passive SMW.

**Procedure 9. Updating SMW software on an SMW HA system**

**Note:** During this procedure, you will need to refer to the SMW release installation procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

1. Log into both SMWs as `root`.
2. If you are using Virtual Network Computing (VNC) software to enable remote access to the SMW, shut down the VNC server.

   **Note:** In the following command, replace *display* with the necessary display number.

   ```
   smw1: # vncserver -kill :display
   ```

   For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).
3. Find the active SMW by determining where the SMW HA cluster resources are running (such as the `hss-daemons` resource).

   ```
   smw1: # crm_mon -1 | grep hss-daemons
    hss-daemons    (lsb:rsms):     Started smw1
   ```

   **Note:** The examples in this procedure assume that `smw1` is currently the active SMW.
4. Record the DRAC IP address of both SMWs in case you need to power-cycle either SMW; see Table 2.

**Tip:** Usually, the DRAC host name follows the naming convention *hostname*`-drac`. For example, if the host names are `smw1` and `smw2`, the DRAC host names would be `smw1-drac` and `smw2-drac`, In this case, the following `ping` commands would display the DRAC IP addresses.

**Note:** In these commands, replace *smw1-drac* with the host name of the DRAC on the active SMW. Replace *smw2-drac* with the host name of the DRAC on the passive SMW.

```
smw1: # ping smw1-drac
PING smw1-drac.us.cray.com (172.31.73.77) 56(84) bytes of data.
64 bytes from smw1-drac.us.cray.com (172.31.73.77): icmp_seq=1 ttl=64 time=1.85 ms
64 bytes from smw1-drac.us.cray.com (172.31.73.77): icmp_seq=2 ttl=64 time=0.398 ms
64 bytes from smw1-drac.us.cray.com (172.31.73.77): icmp_seq=3 ttl=64 time=0.408 ms
...

smw1: # ping smw2-drac
PING smw2-drac.us.cray.com (172.31.73.79) 56(84) bytes of data.
64 bytes from smw2-drac.us.cray.com (172.31.73.79): icmp_seq=1 ttl=64 time=1.85 ms
64 bytes from smw2-drac.us.cray.com (172.31.73.79): icmp_seq=2 ttl=64 time=0.398 ms
64 bytes from smw2-drac.us.cray.com (172.31.73.79): icmp_seq=3 ttl=64 time=0.408 ms
...
```

5. On the active SMW, shut down the system.

```
smw1: # xtbootsys -s last -a auto.xtshutdown
```

6. Stop file synchronizing.

```
smw1: # crm resource stop fsync
```

7. Turn on maintenance mode.

```
smw1: # crm configure property maintenance-mode=true 2> /dev/null
```

**Note:** You **must** execute this command from the active SMW.

8. Turn on `mysql`.

```
smw1: # chkconfig mysql on
```

9. Install the SMW software update on the active SMW.

**Note:** The SMW HA system requires SMW 7.0.UP03 or later release software.

Follow all the upgrade install steps in the *Installing Cray System Management Workstation (SMW) Software* (S–2480), including the step to flash the mainframe.

**Important:** The flash step must be done on the active SMW only, not the passive SMW.

10. On the active SMW, turn off maintenance mode.

```
smw1: # crm configure property maintenance-mode=false 2> /dev/null
```

11. On the active SMW, force a failover to the passive SMW, then wait 30 seconds for the failover operation to complete.

```
smw1: # crm node standby
smw1: # sleep 30
```

12. On the other SMW (which is now the active one), clear the resource failcounts.

```
smw2: # clear_failcounts
```

13. Bring the first SMW online.

    **Note:** Replace *smw1* with the host name of the SMW that is normally active.

```
smw2: # crm node online smw1
```

14. Turn on MySQL.

```
smw2: # chkconfig mysql on
```

15. Turn on maintenance mode on the second SMW.

```
smw2: # crm configure property maintenance-mode=true 2> /dev/null
```

16. Install the SMW software on the second SMW.

    **Important:** When you install the SMW software update, you **must** skip several steps on the second SMW. This step summarizes the differences. For the detailed procedures, refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480).

    a. Follow the procedure in section 4.2, Backing Up Your Current Software, to back up the second SMW's software.

    b. In section 4.3, Installing an SMW Update Package, skip Procedure 20 (Shutting down the Cray system). No system shutdown is required before installing the SMW software on the second SMW.

    c. Follow all applicable steps in Procedure 21. Updating the SMW software and configuration (also in section 4.3).

    d. Skip section 4.4, For Cray XC30 Systems Only: Updating the BC and CC Firmware.

       **Important:** Do **not** flash the mainframe from the second SMW.

    e. Skip section 4.6, Confirming the SMW is Communicating with the System Hardware. This procedure was already done for the first SMW.

    f. In section 4.9, Updating SMW Software on the Boot Root and Shared Root, execute all applicable steps.

17. Turn off maintenance mode on the second SMW.

```
smw2: # crm configure property maintenance-mode=false 2> /dev/null
```

18. Start file synchronizing on the first SMW.

```
smw1: # crm resource start fsync
```

19. Turn off `rsms`, `cray-mzwatcher`, and `mysql` on both SMWs.

```
smw1: # chkconfig rsms off
smw1: # chkconfig cray-mzwatcher off
smw1: # chkconfig dbMonitor off
smw1: # chkconfig mysql off

smw2: # chkconfig rsms off
smw2: # chkconfig cray-mzwatcher off
smw2: # chkconfig dbMonitor off
smw2: # chkconfig mysql off
```

20. Reboot the first SMW and wait for the reboot to finish.

```
smw1: # reboot
```

21. Reboot the second SMW and wait for the reboot to finish.

```
smw2: # reboot
```

22. From either SMW, execute the `clear_failcounts` command to clean up any SMW HA resource errors.

```
smw1: # clear_failcounts
```

23. Display the cluster status and verify that each resource has been started.

```
smw1: # crm_mon -1
============
Last updated: Fri May 31 16:47:25 2013
Last change: Tue May 28 19:02:58 2013 by root via crm_resource on smw1
Stack: openais
Current DC: smw2 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw2 smw1 ]

 stonith-1      (stonith:external/ipmi):        Started smw1
 stonith-2      (stonith:external/ipmi):        Started smw2
 dhcpd  (lsb:dhcpd):    Started smw1
 cray-syslog    (lsb:cray-syslog):      Started smw1
 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP1     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP2     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP3     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP4     (ocf::heartbeat:IPaddr2):       Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 homedir        (ocf::heartbeat:Filesystem):    Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw11
```

**Note:** `crm_mon` may display the resources in a different order on your system.

24. If any resource status is `Stopped`, execute the `clean_resources` command.

```
smw1: # clean_resources
```

25. Display the resource failcount data. All failcounts should be zero.

```
smw1:~# show_failcounts
node=smw1 scope=status  name=fail-count-stonith-1 value=0
node=smw1 scope=status  name=fail-count-stonith-2 value=0
node=smw1 scope=status  name=fail-count-dhcpd value=0
node=smw1 scope=status  name=fail-count-cray-syslog value=0
...
```

26. If you are using Virtual Network Computing (VNC) software to enable remote access to the SMW, restart the VNC server. For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).

## 3.3 Updating SMW HA Software

To update the SMW software, you will update the active SMW first, then fail over to the other SMW to update that system.

**Procedure 10. Updating SMW HA software on an SMW HA system**

**Note:** The examples in this procedure assume that `smw1` is currently the active SMW.

1. Log into both SMWs as `root`.
2. Find the active SMW by determining where the SMW HA cluster resources are running (such as the `hss-daemons` resource).

```
smw1: # crm_mon -1 | grep hss-daemons
 hss-daemons    (lsb:rsms):     Started smw1
```

3. Install the SMW HA software update on both SMWs.

   a. Mount the Cray SMW HA release media on the SMW.

      - If you have the release media on DVD, place the Cray SMW HA DVD into the DVD drive and mount it to `/media/cdrom`.

```
smw1: # mount /dev/cdrom /media/cdrom
```

      - If you have an electronic version of the release media, mount the Cray SMW HA ISO to `/media/cdrom`.

        **Note:** The ISO file name depends on the release number, and installer version. The following command shows the generic ISO name `SMW-SLEHA11SP2`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP2b06.iso`). Substitute the actual ISO file name as specified in the SMW HA release information.

        For *path*, substitute the actual path to the ISO on your system (such as `/tmp`).

```
smw: # mount -o loop,ro /path/SMW-SLEHA11SP2xnn.iso
/media/cdrom
```

   b. Put the active SMW into standby mode. This command causes the system to fail over to the other SMW (`smw2`).

```
smw1: # crm node standby
```

   c. Go to the `/media/cdrom` directory and execute the `SMWHAinstall` script to install the Cray SMW HA release software on the SMW.

```
smw1: # cd /media/cdrom
smw1: # ./SMWHAinstall -v
```

   d. Examine the initial output and note the process ID (PID) of the `SMWHAinstall` process. `SMWHAinstall` creates a log file in `/var/adm/cray/logs/SMWHAinstall.`*PID*`.log` using this PID.

e. Unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

```
smw1: # umount /media/cdrom
smw1: # eject
```

f. Repeat step 3 on the other SMW (which became the active SMW in step 3.b).

**Note:** When repeating the commands in this step, note that the system prompt will show the host name of the second SMW (for example, `smw2: #` instead of `smw1: #`).

4. If you have made local changes to the file `/etc/csync2/csync2_cray.cfg`, restore your local changes to the updated file.

   The installation procedure saves local changes in the file `/etc/csync2/csync2_cray.cfg.sav`. You must copy these changes into `csync2_cray.cfg`.

   a. Change to the `/etc/csync2` directory.

   b. In `csync2_cray.cfg.sav`, locate the `group user_group` section, and copy the include and exclude lines.

   c. In the distribution file `csync2_cray.cfg`, add your local include and exclude lines to the `group user_group` section.

5. Update the SMW HA cluster configuration.

   a. If you are using Virtual Network Computing (VNC) software to enable remote access to the SMW, shut down the VNC server.

   **Note:** In the following command, replace *display* with the necessary display number.

   ```
   smw1: # vncserver -kill :display
   ```

   For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).

   b. On the active SMW, change to the directory containing the `SMWHAconfig` command.

   ```
   smw1: # cd /opt/cray/ha-smw/default/hainst
   ```

   c. Load the `ha-smw` module.

   ```
   smw1: # module load ha-smw
   ```

   d. Execute the `SMWHAconfig` command with the `--update` option.

   **Important:** Execute this command only on the active SMW.

   ```
   smw1: # ./SMWHAconfig --update
   ```

e. When `SMWHAconfig` runs, it prompts for a password so that it can configure the SMW HA cluster and the DRAC; enter the root password for the SMW.

f. If necessary, examine the log file. `SMWHAconfig` creates a log file in `/opt/cray/ha-smw/default/hainst/SMWHAconfig.out`.

g. Reboot `smw1` and wait for the reboot to finish.

```
smw1: # reboot
```

h. Reboot `smw2` and wait for the reboot to finish.

```
smw2: # reboot
```

i. If necessary, restart the VNC server. For more information, see *Installing Cray System Management Workstation (SMW) Software* (S–2480).

6. From either SMW, execute the `clear_failcounts` command to clean up any SMW HA resource errors.

```
smw1: # clear_failcounts
```

7. Display the cluster status and verify that each resource has been started.

```
smw1: # crm_mon -1
============
Last updated: Fri May 31 16:47:25 2013
Last change: Tue May 28 19:02:58 2013 by root via crm_resource on smw1
Stack: openais
Current DC: smw2 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw2 smw1 ]

 stonith-1      (stonith:external/ipmi):        Started smw1
 stonith-2      (stonith:external/ipmi):        Started smw2
 dhcpd  (lsb:dhcpd):    Started smw1
 cray-syslog    (lsb:cray-syslog):      Started smw1
 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP1     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP2     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP3     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP4     (ocf::heartbeat:IPaddr2):       Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 homedir        (ocf::heartbeat:Filesystem):    Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw11
```

**Note:** `crm_mon` may display the resources in a different order on your system.

8. If any resource status is `Stopped`, execute the `clean_resources` command.

```
smw1: # clean_resources
```

9. Display the resource failcount data. All failcounts should be zero.

```
smw1:~# show_failcounts
node=smw1 scope=status  name=fail-count-stonith-1 value=0
node=smw1 scope=status  name=fail-count-stonith-2 value=0
node=smw1 scope=status  name=fail-count-dhcpd value=0
node=smw1 scope=status  name=fail-count-cray-syslog value=0
...
```

**Note:** The `SMWHAconfig` command backs up the cluster configuration before beginning the update.

The previous configuration is stored in the directory `/opt/cray/ha-smw/default/hainst`, in a file named `_CLUSTER_CONFIG_BACKUP_`*YYYY-MM-DD-hh:mm* (for example, `_CLUSTER_CONFIG_BACKUP_2013-06-11-10:11`).

If there is a problem with the update, you can use the following commands to restore the previous configuration (in the file name, specify the actual timestamp instead of *YYYY-MM-DD-hh:mm*):

```
smw1: # crm node standby
smw2: # crm node standby
smw1: # crm configure load replace \
/opt/cray/ha-smw/default/hainst/_CLUSTER_CONFIG_BACKUP_YYYY-MM-DD-hh:mm 2> /dev/null
.
.
.
Do you still want to commit? y
```

# Part II: Administration Guide

# Overview [4]

This SMW High Availability (HA) release implements the SMW failover feature. This feature is supported only for Cray XC30 systems with two high-end second-generation SMWs (also called *rack-mount SMWs*). You must be running the SUSE Linux Enterprise Server version 11 Service Pack 2 (SLES 11 SP2) SMW base operating system and the SMW 7.0.UP03 release.

The SMW failover feature provides improved reliability, availability, and serviceability (RAS) of the SMW, allowing the mainframe to operate correctly and at full speed. This feature adds SMW failover, fencing, health monitoring, and failover notification.

The administrators of the SMW can be notified of SMW software or hardware problems in real time and be able to react by manually shutting down nodes, or allowing the software to manage the problems. In the event of a hardware failure or SMW-based HSS (rsms) daemon failure, the software will failover to the passive SMW node, which becomes the active node. The failed node, once repaired, can be returned to the configuration as the passive node.

Administration and monitoring of the SMW HA system is available via the SUSE Pacemaker Cluster Resource Manager (CRM), which has a command line interface (`crm`) and a GUI (`crm_gui`). With this interface and associated commands, the SMW administrator can do the following:

- Display the current state of the SMW HA cluster
- Monitor the HSS daemons, which are configured as cluster resources
- Customize the SMW failover thresholds for each resource
- Configure automatic email notification of failover

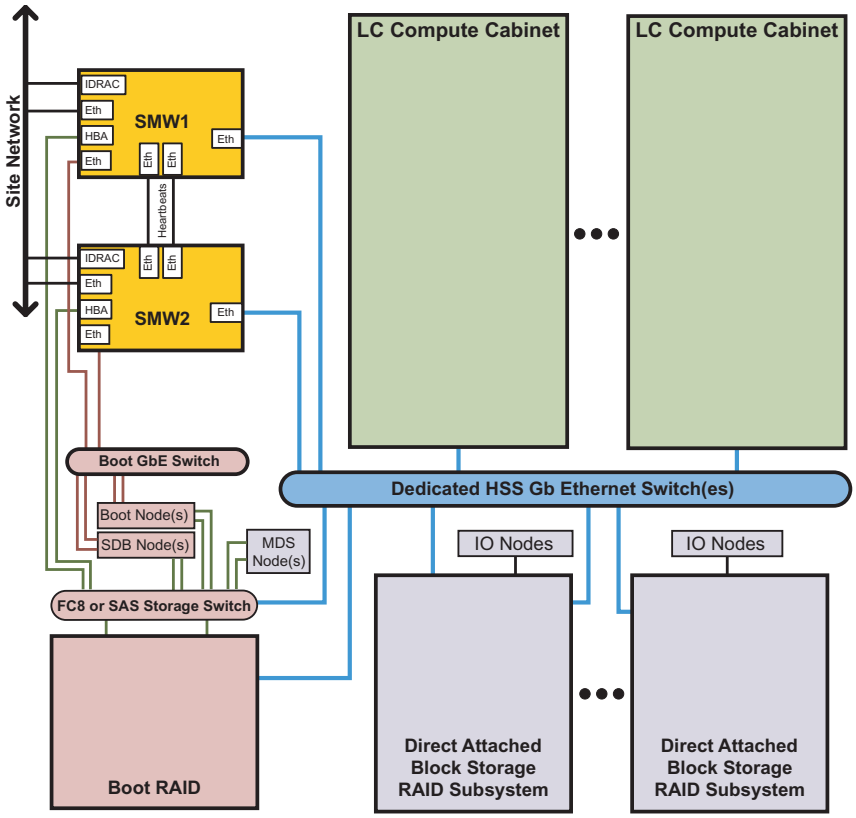
**Note:** The Pacemaker Cluster Resource Manager uses the term *node* to refer to a host in a CRM cluster. Throughout this document, unless specified otherwise, node refers to an SMW, not to a Cray XC30 compute or service node.

## 4.1 SMW Cluster Configuration

The two SMWs are connected with heartbeat cables between the `eth2` and `eth4` ports on each SMW. The heartbeat connection monitors the health of the cluster. In addition, each SMW is connected to the boot RAID (through FC or SAS cards), to the site network through `eth0`, to the HSS network through `eth1`, and to the boot node through `eth3`. For more information, see Network Connections on page 12.

Figure 1 shows the major connections between components in an SMW HA system.

**Figure 1. SMW HA Hardware Components**




In a Cray SMW HA cluster, the two SMWs are configured in an active/passive configuration. This configuration lets the passive node take over the SMW functions if a software or hardware fault occurs on the active node. All HSS daemons run on the active SMW. (An additional STONITH daemon, which monitors SMW health, runs on both SMWs.) At failover, all daemons move to the passive SMW, which then becomes the active one.

## 4.2 Shared Storage

The SMW HA system uses shared disk devices on the boot RAID for data that must be highly available. The following files and directories on the boot RAID are shared by both SMWs.

- Log disk (`/var/opt/cray/disk/1`), which symbolically links to the `/var/opt/cray/*` directories:
  - `/var/opt/cray/dump`
  - `/var/opt/cray/install`
  - `/var/opt/cray/log`
- MySQL HSS database (`/var/lib/mysql`)

  **Note:** The HSS database server runs on active SMW only.
- Home (`/home`)

These directories are mounted only on the active SMW. When a failover occurs, access to these directories is automatically transferred to the other SMW as part of the failover process.

**Note:** For systems running the power management feature, the Postgres database runs on both SMWs, because there is insufficient space on shared boot RAID.

The boot RAID is configured with these shared directories during initial installation. For more information, see Configuring the Boot RAID for SMW HA on page 19.

**Important:** Because several file systems are shared between the two SMWs, an SMW HA system has a slightly increased risk for double-mount problems. Do not mount `/bootroot`, `/sharedroot`, or any other CLE RAID file systems on both SMWs at the same time.

## 4.3 Synchronized Files

For files not located on the shared storage device, the SLEHA Extension software includes the `csync2` utility to synchronize (*sync*) important files between the two SMWs. When a file changes on the active SMW, it is automatically synchronized to the passive SMW.

File synchronization is automatically configured during initial installation. The file `/etc/csync2/csync2_cray.cfg` specifies the Cray-specific files and directories that must be synchronized, as well as small files that are convenient to keep in sync. The synchronized files include the following:

- `/opt/cray/hss/*/etc/*`
- `/opt/cray/hss-images` (with some exclusions)
- `/etc/hosts`
- `/etc/passwd`
- `/etc/group`
- `/root/authorized_keys*`
- `/etc/motd`

**Note:** Very large files are explicitly excluded from synchronization (such as `/opt/cray/hss-images/master`).

File synchronization happens in one direction only: from the active SMW to the passive. If you change a synchronized file on the passive SMW, the change will not be propagated to the active SMW in the course of normal operations and could be overwritten on the passive SMW later if there is a subsequent change to the corresponding file on the active SMW.

However, if a failover occurs, the previously passive SMW becomes the active SMW. If the change is still in place, the changed file becomes a candidate for propagation to the other SMW (subject to the rules of file conflict resolution).

The `fsync` resource controls file synchronized operations. Every 100 seconds, `fsync` checks for files that need to be synchronized.

**Important:** If a failover occurs before a file synchronization operation has completed, it could result in the loss of the latest updates.

Although it is possible to customize file synchronization, Cray recommends that you not change the list of synchronized files, or add only small files to `/etc/csync2/csync2_cray.cfg`. If `csync2` must monitor many directories or synchronize a large amount of data, it can become overloaded and failures may not be readily apparent. For more information, see Managing Synchronized Files on page 60.

## 4.4 Cluster Resources

A resource is any type of service or application that is managed by the Pacemaker Cluster Resource Manager, such as a daemon or file system. In an SMW HA system, the HSS (rsms) daemons are configured as resources.

Each time a resource fails, it is automatically restarted and its failcount is raised. If the failcount exceeds the defined migration threshold for the resource, a failover occurs and management of all cluster resources migrates to the other SMW, making it the active SMW. The original SMW will no longer be allowed to run the failed resource, so no failback can occur until the resource's failcount is reset for that SMW.

> **Tip:** Failcounts can be reset using the `clean_resources` or `clear_failcounts` command. For more information, see If Resources Are Not Running on page 67.

An SMW HA system includes the following resources:

`stonith-1` and `stonith-2`

Monitors the health of the other SMW. Each SMW monitors its peer and has the capability to power off that peer at failover time, using the STONITH (Shoot The Other Node In The Head) mechanism. STONITH failovers are used when the state of the failing SMW cannot be determined. A STONITH failover powers off the failing SMW to guarantee that the newly active SMW has exclusive access to all cluster managed resources.

`dhcpd` Controls and monitors `dhcpd` as used by the SMW HA feature.

`cray-syslog`

Controls and monitors Lightweight Log Management (LLM).

`ClusterIP`, `ClusterIP1`, `ClusterIP2`, `ClusterIP3`, and `ClusterIP4`

Controls and monitors the Ethernet connections (`eth0`, `eth1`, `eth2`, `eth3`, and `eth4`, respectively).

`fsync` Provides file synchronization using `csync2`.

`homedir` Mounts and unmounts `/home`.

`hss-daemons`

Controls and monitors HSS daemons; corresponds to the `/etc/init.d/rsms` startup script.

`Notification`

Provides automatic notification email when a failover occurs.

`ml-fs` Mounts, unmounts, and monitors `/var/opt/cray/disk/1`.

`md-fs` Mounts, unmounts, and monitors `/var/lib/mysql`.

`mysqld` Controls and monitors MySQL.

## 4.5 Limitations of SMW Failover

The SMW HA failover feature has the following limitations:

- There is no support for seamless failover (also called *double failure*) if errors occur while the system is doing error handling for another system component. If an HSS daemon or other SMW process were doing some type of error handling that got interrupted by an (unrelated) failover, when that daemon restarts on the new SMW it may not be able to resume operation where it left off and complete the recovery from the first error. In this case, even though a failover occurs, manual intervention might still be required to return the system to an operational state.
- There is no support for seamless failover during operational commands. An interrupted operation such as `xtbootsys`, shutdown, dump, warm-swap, or flash will need to be reissued after failover has completed and the other SMW becomes active.
- Partial migration of managed resources is not supported. For example, the SMW HA system does not support migration of individual HSS daemons or resources to the other SMW. A particular SMW is either *active*, with complete responsibility for all HSS daemons, or *passive* with no HSS daemons running.
- If both SMWs are started (powered on) at the same time, a race condition can develop that could result in one SMW being powered off via the STONITH mechanism. Before starting the second SMW, wait until the first SMW has completed startup and initialized all cluster resources. For more information, see Booting an SMW HA Cluster on page 49.
- SMW and CLE upgrades in an HA environment require some duplication of effort, with portions of the procedure done individually to each SMW. System down-time requirements for operating system upgrades are somewhat longer as a result.
- Both SMWs must run the same versions of SLES and SMW/HSS software.
- Auto-throttling of applications is likely while an actual SMW failover is taking place. Blades begin to auto-throttle if essential HSS daemons (`erd`, `state-manager`, or `xtnlrd`) are unavailable and lasts until those daemons resume operation on the other SMW. On a single-cabinet system, the throttled period was fairly consistent, lasting 37 seconds. The throttled period may increase for larger systems.
- For systems running the power management feature: Power management data collected before an SMW failover will be lost or not easily accessible after failover.
- System administration of an SMW HA environment is more complex than administration of a system with a single SMW.

# Managing Your Cray SMW HA System [5]

**Important:** When using the procedures in this chapter, be careful not to change the cluster configuration, to migrate a single resource (all resources must migrate as a group), or to change the system list of synchronized files. Exceptions:

- You can set the migration threshold (maximum failcount value) for each resource by using the `set_migration_threshold` command; see Cray Cluster Commands on page 54.
- You can define which local (site-specific) files are synchronized or excluded from synchronization; see Managing Synchronized Files on page 60.
- You can change the email address for the `Notification` resource; see Configuring Failover Notifications on page 59.

The following conventions are used in this chapter:

- SMW1 or the `smw1` host name specifies the SMW that has been configured to be the active SMW during normal operation. In examples, the prompt `smw1: #` shows a command that runs on this SMW.
- SMW2 or the `smw2` host name specifies the SMW that has been configured to be the passive SMW during normal operation. In examples, the prompt `smw2: #` shows a command that run on this SMW.
- The `cray-smw` host name specifies the currently active SMW (which could be either `smw1` or `smw2`). This virtual host name is defined during installation and configuration.

## 5.1 Booting an SMW HA Cluster

**Important:** When SMW HA is enabled, do **not** start both SMWs at the same time. Doing so can cause a race condition that could result in one SMW being powered off via the STONITH mechanism. Before starting the second SMW, wait until the first SMW has completed startup and initialized all cluster resources.

**Procedure 11. Booting an SMW HA cluster**

Follow these steps to boot or reboot both SMWs.

1. Boot `smw1` (or the SMW that you want to be active).
2. Wait until the SMW is completely booted and SMW HA services have started on that SMW.

   **Tip:** You can check the status of the SMW HA services with the `crm_mon -1` command. For more information, see Displaying Cluster Status on page 56.

3. Boot `smw2` (or the SMW that you want to be passive).
4. Verify that both SMWs are online and that all cluster resources have started. For more information, see Displaying Cluster Status on page 56.

## 5.2 Logging In

Cray recommends that you always connect to the SMW cluster using the virtual host name. Avoid connecting to an SMW by specifying the actual host names, except for host-specific maintenance. In the event of a failover, all connections made using the virtual host name will be terminated. A connection to the active SMW via the actual host name could be confusing after a failover occurs, because the login session would remain open, but there is no indication that the SMW is now passive.

**Procedure 12. Logging into the SMW**

- To log into the active SMW, specify the virtual SMW host name.

  **Note:** This example shows the virtual host name *cray-smw*. Specify the virtual host name of your SMW HA cluster.

  ```
  remote-system% ssh root@cray-smw
  smw1:~ #
  ```

  After you log in, the prompt displays the host name of the active SMW (in this example, `smw1`).

To log into a specific SMW, use the actual host name of the SMW (such as `smw1` or `smw2`).

## 5.3 Identifying the Active SMW

**Procedure 13. Identifying the active SMW**

1. The easiest way to find the active SMW is to log in using the virtual SMW host name and look at the system prompt, as described in Procedure 12 on page 50.

2. Another way to find the active SMW is to determine where the SMW HA cluster resources are running (such as the `hss-daemons` resource).

   **Note:** One `stonith` resource runs on each SMW to monitor the other SMW; all other resources run only on the active SMW.

   As `root` on either SMW, execute the following command.

```
smw1: # crm_mon -1 | grep hss-daemons
 hss-daemons    (lsb:rsms):     Started smw1
```

## 5.4 Using SMW HA Commands

Only the `root` user can execute the Cray SMW HA commands. These commands are included in the `ha-smw` module, which is automatically loaded when the `root` user logs in. If necessary, use the following command to load the `ha-smw` module:

```
smw1: # module load ha-smw
```

### 5.4.1 `crm`

The `crm` command provides a command-line interface to the SUSE Pacemaker Cluster Resource Manager (CRM). This command can be used either as an interactive shell or as a single command entered on the command line.

As a simple example, you can use this single command to display the status of the `fsync` resource.

```
smw1:~ # crm resource status fsync
resource fsync is running on: smw1
```

Or you can use the interactive method to display the same information:

```
smw1:~ # crm
crm(live)# resource
crm(live)resource# status fsync
resource fsync is running on: smw1
crm(live)resource# end
crm(live)# quit
smw1:~ #
```

**Tip:** The `crm` command has multiple levels. You can use the `help` keyword to display the commands at each level and the valid options and arguments for each command; for example, `crm help`, `crm resource help`, or `crm resource failcount help`.

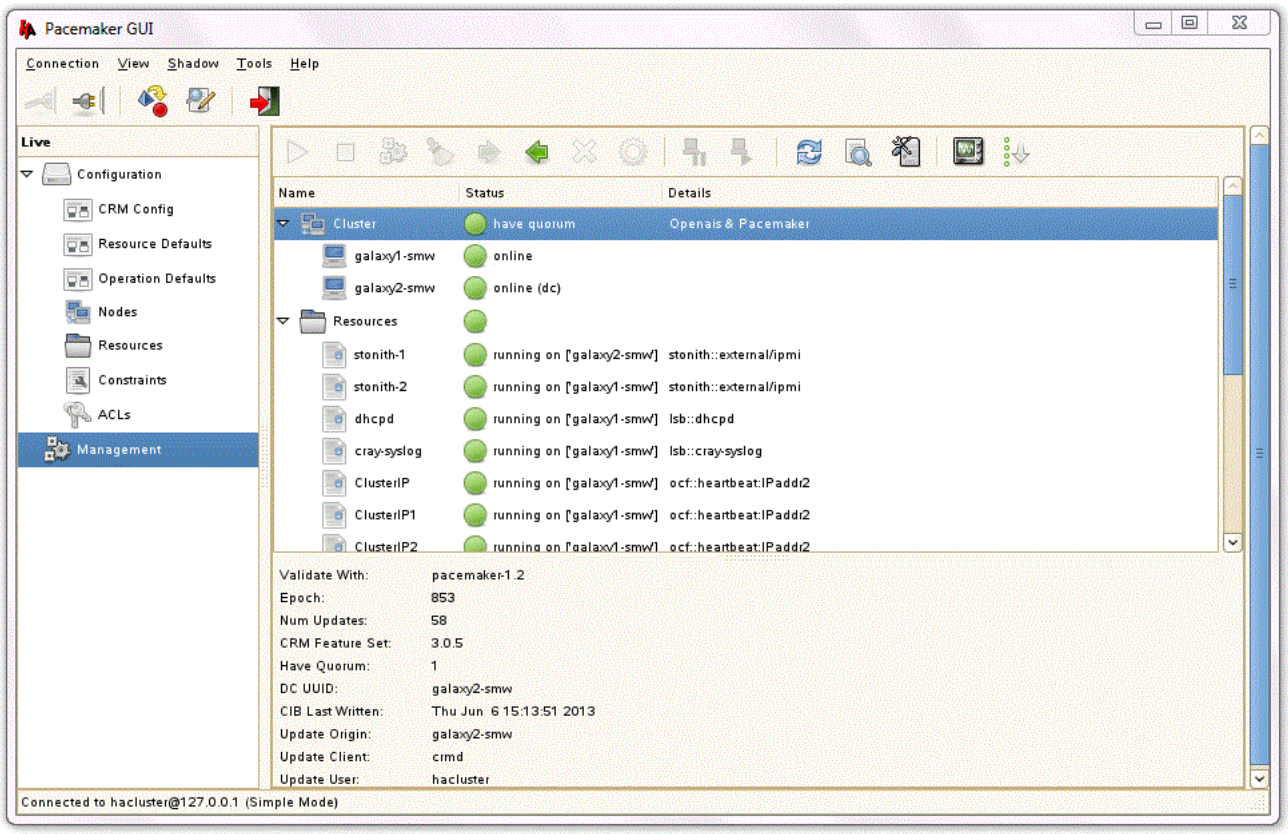
For more information, see the `crm`(8) man page and the *SUSE Linux Enterprise High Availability Extension High Availability Guide*.

## 5.4.2 `crm_gui`

The `crm_gui` command provides a graphical interface to the SUSE Pacemaker Cluster Resource Manager (CRM).

When the `crm_gui` window opens, it is blank. Connect to the cluster with **Connection→Login**, then log in as the `hacluster` user. For the password, see Passwords on page 14.

**Figure 2. Pacemaker GUI (`crm_gui`) Management Window**



Click on **Management** in the left panel to display node and resource status, as shown in Figure 2. A green circle marks an node or resource that is running without errors; a red circle marks an item with problems, such as an offline node or stopped resource. Click on a node or resource to display status details (including errors) in the bottom panel of the window.

> **Note:** The management display marks one of the nodes with `(dc)`, which stands for *designated coordinator*. This is a Pacemaker CRM concept that is not related to the SMW's current active or passive role. The active SMW is not necessarily the CRM designated coordinator.

For information on using `crm_gui`, see the *SUSE Linux Enterprise High Availability Extension High Availability Guide*.

### 5.4.3 `crm_mon`

The SUSE `crm_mon` command allows you to monitor cluster status and configuration. The output includes the number of nodes, host names, SMW status, the resources configured in your cluster, and the current status of each resource.

By default (if no options are specified), `crm_mon` runs continuously, updating the cluster status every 15 seconds. To display a single snapshot of cluster status, use the `-1` option:

```
smw1: # crm_mon -1
============
Last updated: Fri May 31 16:47:25 2013
Last change: Tue May 28 19:02:58 2013 by root via crm_resource on smw1
Stack: openais
Current DC: smw2 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw2 smw1 ]

 stonith-1      (stonith:external/ipmi):        Started smw1
 stonith-2      (stonith:external/ipmi):        Started smw2
 dhcpd  (lsb:dhcpd):    Started smw1
 cray-syslog    (lsb:cray-syslog):      Started smw1
 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 .
 .
 .
```

> **Note:** The `crm_mon` output marks one of the nodes as the `Current DC`, which stands for *designated coordinator*. This is a Pacemaker CRM concept that is not related to the SMW's current active or passive role. The active SMW is not necessarily the CRM designated coordinator.

For more information, see the `crm_mon`(8) man page and the *SUSE Linux Enterprise High Availability Extension High Availability Guide*.

### 5.4.4 crm_resource

The SUSE crm_resource command displays resource information. To list the cluster resources, execute the following command as root on either SMW.

```
smw1:~ # crm_resource -l
stonith-1
stonith-2
dhcpd
cray-syslog
ClusterIP
ClusterIP1
ClusterIP2
ClusterIP3
ClusterIP4
fsync
homedir
hss-daemons
Notification
ml-fs
md-fs
mysqld
```

For more information, see the crm_resource(8) man page and the *SUSE Linux Enterprise High Availability Extension High Availability Guide*.

### 5.4.5 Cray Cluster Commands

The Cray SMW HA software provides several commands to monitor the cluster status, clean up resource problems, and configure migration thresholds.

**Note:** You must be root to execute these commands. Except as noted below, all commands can be run on either the active or passive SMW.

- show_failcounts: Displays the failcounts of all SMW failover resources on both SMWs. This command shows the failcounts (number of failures) for all resources on both SMWs.

  **Note:** The show_failcounts command provides a quick way to access the failcount data for all resources in an SMW HA cluster, rather than running multiple crm or crm_failcount commands.

- show_failcount: Displays the failcount of a specific SMW failover resource. This command shows the failcount (number of failures) of the specified resource. This command provides a simple way to display the failcount data of a resource, rather than running the crm or crm_failcount command.

- clear_failcounts: Resets the failcounts for all SMW failover resources. This command resets the resource failcounts (number of failures) in an SMW HA cluster. This command resets all resources on both the active and passive SMW.

  **Note:** clear_failcounts provides a quick way to clear all failcount data, rather than running multiple crm or crm_failcount commands.

- `clean_resources`: Cleans up all SMW failover resources on both SMWs. This command sets each resource's status to the default clean state and sets the failcount (number of failures) to 0. If some resources did not start after system boot or are marked as unclean after failover, you can use this command to quickly clean up all resources on both SMWs.

  **Note:** The command `crm resource cleanup` also cleans up resources, but requires you to enter each resource name separately.

- `set_migration_threshold`: Sets the migration threshold for an SMW failover resource. A migration threshold is defined as the maximum number of failures (the failcount) allowed for the resource. If the failcount exceeds this threshold, a failover occurs and management of all cluster resources migrates to the other SMW, making it the active SMW. By default, the migration threshold is `1000000`.

- `show_migration_threshold`: Displays the migration threshold for an SMW failover resource. A migration threshold is defined as the maximum number of failures (the failcount) allowed for a resource (any type of service or application that is managed by the Pacemaker Cluster Resource Manager, such as a daemon or file system). If the failcount exceeds this threshold, a failover occurs and management of all cluster resources migrates to the other SMW, making it the active SMW. The original SMW will no longer be allowed to run the failed resource until the resource's failcount is reset for that SMW.

  **Note:** Before executing `show_migration_threshold`, you must explicitly set the migration threshold with the `set_migration_threshold` command. If the migration threshold has not been set (that is, if it has the default value), `show_migration_threshold` displays an error message.

- `SMWHAconfig`: Configures SMW failover on both SMWs in an SMW HA cluster. After installing or updating the Cray SMW HA software, execute this command on the active SMW to configure both SMWs through `ssh`.

  **Note:** The `SMWHAconfig` command must be run on the active SMW.

For more information, see the man pages for these commands.

# 5.5 Monitoring the SMW HA Cluster

## 5.5.1 Displaying Cluster Status

**Procedure 14. Displaying cluster and resource status**

You can use some or all of the following steps to check the health of the SMW HA cluster.

**Note:** You must execute the CRM and Cray SMW HA commands as `root`. Unless otherwise noted, you can execute these commands on either SMW.

1. Verify that both SMWs are online.

```
smw1: # crm_mon -1 | grep Online
Online: [ smw1 smw2 ]
```

2. Display the cluster status with `crm_mon`.

```
smw1: # crm_mon -1
============
Last updated: Fri May 31 16:47:25 2013
Last change: Tue May 28 19:02:58 2013 by root via crm_resource on smw1
Stack: openais
Current DC: smw2 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw2 smw1 ]

 stonith-1      (stonith:external/ipmi):        Started smw1
 stonith-2      (stonith:external/ipmi):        Started smw2
 dhcpd  (lsb:dhcpd):    Started smw1
 cray-syslog    (lsb:cray-syslog):      Started smw1
 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP1     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP2     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP3     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP4     (ocf::heartbeat:IPaddr2):       Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 homedir        (ocf::heartbeat:Filesystem):    Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw11
```

**Note:** `crm_mon` may display the resources in a different order on your system.

Note that all resources run only on the active SMW (except for one `stonith` resource, which is a special case). In the previous example, `smw1` is the active SMW.

3. Display the status of the cluster resources.

```
smw1: # crm resource status
 stonith-1      (stonith:external/ipmi) Started
 stonith-2      (stonith:external/ipmi) Started
 dhcpd  (lsb:dhcpd) Started
 cray-syslog    (lsb:cray-syslog) Started
 ClusterIP      (ocf::heartbeat:IPaddr2) Started
 ClusterIP1     (ocf::heartbeat:IPaddr2) Started
 ClusterIP2     (ocf::heartbeat:IPaddr2) Started
 ClusterIP3     (ocf::heartbeat:IPaddr2) Started
 ClusterIP4     (ocf::heartbeat:IPaddr2) Started
 fsync  (ocf::smw:fsync) Started
 homedir        (ocf::heartbeat:Filesystem) Started
 hss-daemons    (lsb:rsms) Started
 Notification   (ocf::heartbeat:MailTo) Started
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem) Started
     md-fs      (ocf::heartbeat:Filesystem) Started
     mysqld     (ocf::heartbeat:mysql) Started
```

For information on restarting a stopped resource, see If Resources Are Not Running on page 67.

4. Display failcount data for all resources.

```
smw1:~# show_failcounts
node=smw1 scope=status  name=fail-count-stonith-1 value=0
node=smw1 scope=status  name=fail-count-stonith-2 value=0
node=smw1 scope=status  name=fail-count-dhcpd value=0
node=smw1 scope=status  name=fail-count-cray-syslog value=0
node=smw1 scope=status  name=fail-count-ClusterIP value=0
   .
   .
   .
node=smw2 scope=status  name=fail-count-hss-daemons value=0
node=smw2 scope=status  name=fail-count-Notification value=0
node=smw2 scope=status  name=fail-count-ml-fs value=0
node=smw2 scope=status  name=fail-count-md-fs value=0
node=smw2 scope=status  name=fail-count-mysqld value=0
```

**Note:** You can display the failcount data for a single resource on one SMW. This example shows the failcount data for the `fsync` resource. (Replace *smwX* with the actual SMW host name.)

```
smw1: # show_failcount smwX fsync
scope=status  name=fail-count-fsync value=0
```

For information on clearing the failcount values, see If Resources Are Not Running on page 67.

5. Test file synchronization by creating a temporary file in a synchronized directory on the active SMW, then check for it on the passive SMW.

   **Note:** This example assumes that `smw1` is the active SMW.

```
smw1: # cp /etc/motd /opt/cray/hss/default/etc/my_test_file
smw1: # ls -l /opt/cray/hss/default/etc/my_test_file
smw1: # md5sum /opt/cray/hss/default/etc/my_test_file
```

   *... (wait about 2 minutes for the next file synchronization operation to complete) ...*

```
smw1: # ssh smw2
...
smw2: # ls -l /opt/cray/hss/default/etc/my_test_file
smw2: # md5sum /opt/cray/hss/default/etc/my_test_file
```

   Finally, return to the active SMW to delete the test file. Within several minutes, the file will be automatically removed from the passive SMW.

## 5.5.2 Displaying SMW Power Status

**Procedure 15. Displaying SMW power status**

If you are not near the SMWs to check the LEDs, you can use one of the following methods to display the power status for the SMWs:

1. As `root` on either SMW, use the `crm_mon` command to check the SMW status.

```
smw1: # crm_mon -1
============
Last updated: Mon Jul 15 15:32:58 2013
Last change: Wed Jun 26 11:35:09 2013 by root via crm_attribute on smw1
Stack: openais
Current DC: smw1 - partition WITHOUT quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw1 ]
OFFLINE: [ smw2 ]

 stonith-2      (stonith:external/ipmi):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw11
    .
    .
    .
```

2. As `root` on either SMW, use the `ipmitool` command to check the power status of a specific SMW.

   **Note:** Replace *smw-DRAC-IP-addr* with the SMW's DRAC IP address.

```
smw1: # /usr/bin/ipmitool -I lan -U root -H smw-DRAC-IP-addr -a chassis power status
Password:
Chassis Power is on
```

   **Note:** At the password prompt, enter the root password for the DRAC.

For the procedure to restore power and join the SMW to the cluster, see If an SMW is Powered Off on page 69.

## 5.6 Customizing Your SMW HA Cluster

### 5.6.1 Configuring Failover Notifications

The SMW HA software includes a `Notification` resource that automatically sends email when a failover occurs. Only one recipient (a single user ID or email group) is allowed. This resource is configured during initial installation.

If a failover occurs, the `Notification` resource sends several messages that are similar to the following examples.

```
From: root [mailto:root@smw.none]
Sent: Thursday, June 06, 2013 9:20 PM
To: Cray Cluster Administrator
Subject: ***Alert*** A Failover may have occurred. Please investigate! Migrating resource
away at Thu Jun 6 21:20:25 CDT 2013 from smw1

        ***Alert*** A Failover may have occurred. Please investigate! Migrating resource away
                    at Thu Jun  6 21:20:25 CDT 2013 from smw1

        Command line was:
        /usr/lib/ocf/resource.d//heartbeat/MailTo stop
```

```
From: root [mailto:root@smw.none]
Sent: Thursday, June 06, 2013 9:20 PM
To: Cray Cluster Administrator
Subject: ***Alert*** A Failover may have occurred. Please investigate! Takeover in progress
at Thu Jun 6 21:20:25 CDT 2013 on smw2

        ***Alert*** A Failover may have occurred. Please investigate! Takeover in progress
                    at Thu Jun  6 21:20:25 CDT 2013 on smw2

        Command line was:
        /usr/lib/ocf/resource.d//heartbeat/MailTo start
```

**Procedure 16. Changing failover notification**

**Note:** Only one email address is allowed. To send notifications to multiple addresses, you can create a group email alias that includes all necessary email addresses.

1. As `root` on either SMW, execute the following `crm resource` command.

```
smw1: # crm resource param Notification set email address@yourdomain.com
```

2. Verify the setting.

```
smw1: # crm resource param Notification show email
address@yourdomain.com
```

## 5.6.2 Managing Synchronized Files

The `fsync` resource controls file synchronization operations. If `fsync` stops, no file synchronization occurs.

**Procedure 17. Monitoring the `fsync` resource**

- To check the status of the `fsync` resource, execute the following command as `root` on either SMW:

```
smw1: # crm_mon -1 | grep fsync
 fsync  (ocf::smw:fsync):       Started smw1
```

For information on restarting a stopped `fsync` resource, see If Resources Are Not Running on page 67.

**Procedure 18. Adding site-specific synchronized files**

The file `/etc/csync2/csync2_cray.cfg` specifies the Cray-specific files and directories that must be synchronized, as well as small files that are convenient to keep in sync.

**Warning:** Cray recommends that you not change the list of synchronized files, or add only small files to `/etc/csync2/csync2_cray.cfg`. If `csync2` must synchronize a large amount of data, it can become overloaded and failures may not be readily apparent.

For example, do not synchronize the following files or directories:

- `/home`
- `/home/crayadm/.ssh/authorized_keys`
- `/opt/xt-images` (Cray boot images are very large)
- Very large files
- Directories that change frequently

**Tip:** You can use `scp` to copy an unsynchronized file (such as new boot image) or directory to the other SMW, as in this example:

```
smw1: # scp -pr smw1:/path/file smw2:/path/file
```

- To add a site-specific file or directory, add the full path (one entry per line) to `/etc/csync2/csync2_cray.cfg`. Comments in this file explain how to make changes.

  **Important:** The parent directories for all entries in this file must already exist on the passive SMW. In some cases, you must either manually create directories on the passive SMW or copy the directory structure from the active SMW. With either method, be sure that owner, group, and permissions are maintained, because `csync2` can be sensitive to mismatches.

  **Note:** For a symbolic link, only the link itself is synchronized, not the content (destination) of the symbolic link.

### 5.6.3 Setting the Migration Threshold for Resources

The `set_migration_threshold` command sets the migration threshold for a resource in an SMW HA cluster. A migration threshold is defined as the maximum number of failures (the failcount) allowed for the resource. If the failcount exceeds this threshold, a failover occurs and management of all cluster resources migrates to the other SMW, making it the active SMW. By default, the migration threshold is `1,000,000`.

**Important:** Cray recommends that you either leave migration thresholds at the default values or set them to a very high value until you have experience with SMW HA operation. Migration threshold settings that are too low could cause the resource to be ineligible to run if the failcount exceeds that value on both SMWs. If lower settings are used, Cray recommends that you monitor failcounts regularly for trends and clear the failcount values as appropriate. Otherwise, transient errors over time could push failcount values beyond the migration threshold, which could lead to one of the following scenarios:

- Failovers could be triggered by a transient error condition that might otherwise have been handled by a less disruptive mechanism.
- Failovers might not be possible because both SMWs have exceeded the migration threshold.

**Procedure 19. Setting the migration threshold for a resource**

**Note:** Execute these commands as `root` on either SMW.

1. Determine the resource name. To display a list of resource names, execute the `crm_resource` command.

   ```
   smw1: # crm_resource -l
   ```

2. Use the `set_migration_threshold` command to change the migration threshold for a resource.

   **Note:** For *resource*, specify a resource name. For *value*, specify an integer in the range of 0 - 1000000.

   ```
   smw1: # set_migration_threshold resource value
   ```

3. Verify the change.

   ```
   smw1: # show_migration_threshold resource
   ```

For more information, see the `set_migration_threshold`(8) man page.

## 5.7 Handling Failover

The following critical events cause a failover from the active SMW to the passive SMW:

- Hardware fault on the active SMW.
- Lost heartbeat between the two SMWs.
- Kernel fault (panic) on the active SMW.
- Failed resource (HSS daemon or cluster service). If a resource stops, the cluster manager automatically restarts it and increments the failcount by 1. When the failcount exceeds the migration threshold (by default, `1,000,000`), a failover occurs.

The failover type (STONITH or non-STONITH) depends upon whether the newly active SMW can determine the health of the failing SMW. A STONITH failover occurs only if there is no other way for the new SMW to ensure the integrity of the cluster.

- In the case of STONITH failover, the original SMW is powered off (via the STONITH mechanism) if it is not already off. This guarantees that file synchronization is stopped and the failed SMW no longer holds any cluster-managed resources so that the new SMW will have exclusive access to those resources.
- In the case of non-STONITH failover, the original SMW is still powered up. In addition:
  - HSS daemons are stopped on the original SMW.
  - Lightweight Log Manager (LLM) logging to shared disk is stopped.
  - File synchronization (`csync2`) between SMWs is stopped.
  - The shared storage versions of `/home`, `/var/opt/cray/disk/1`, and `/var/lib/mysql` are unmounted on the original SMW.
  - Network connections using the `eth0`, `eth1`, `eth2`, `eth3`, and `eth4` virtual IP addresses are dropped and those interfaces begin accepting connections to their actual IP addresses only.

For both types of failover, the following actions then occur on the new SMW:

- The `eth0`, `eth1`, `eth2`, `eth3`, and `eth4` interfaces begin accepting connections using the virtual IP addresses in addition to their actual IP addresses.
- The shared storage versions of `/home`, `/var/opt/cray/disk/1`, and `/var/lib/mysql` are mounted on the new SMW.
- File synchronization (`csync2`) between SMWs usually resumes (depending on the reason for failover).
- LLM logging to the shared disk resumes.
- The HSS database (MySQL) is started on the original SMW.
- HSS daemons are started on the new SMW (including, if necessary, any `xtbootsys`-initiated daemons).

**Important:** When failover occurs:

- All user commands that were started from the active SMW are terminated. These commands must be restarted on the new active SMW. The restarted commands might not start with the same internal states, if those commands do not provide persistent capabilities.
- During failover, if there is no communication between the SMW and the Cray mainframe for about 30 seconds, workload throttling can occur. For more information, see Limitations of SMW Failover on page 48.

## 5.7.1 Restoring Normal Operations After Failover

While a failover is automatic, adding the failed SMW back into the cluster requires manual intervention to identify the reason for failover, take corrective action if needed, and return the failed SMW to an online state. Another failover (that is, a "failback" to the originally active SMW) is not possible until the failed SMW returns to online status and its failcounts are cleared so that it is eligible to run all cluster resources.

**Procedure 20. Restoring normal operations after failover**

1. Identify and fix the problems that caused the failover (such as a hardware fault, kernel panic, or HSS daemon issues).

   **Tip:** Use the following methods to help diagnose problems:

   - Execute the `show_failcounts` command and note any resources with non-zero failcounts.
   - From the active SMW, examine `/var/opt/cray/log/smwmessages-`*yyyymmdd* for relevant messages.
   - Examine the failing SMW for additional clues.
     - For a non-STONITH failover: In most cases, the failing SMW will still be running; additional clues may be available in `dmesg` or via other commands.
     - For a STONITH failover: The failing SMW will be powered off. Before powering it back on, place it into standby mode so that it does not automatically try to rejoin the cluster at startup before we are assured that it is in fact healthy. For more information, see Procedure 22 on page 67.

2. Log into the failing SMW (either from the console or remotely by using the actual host name). Identify the reason for the failure and take corrective action as needed. This might include administrative actions such as freeing space on a file system that has filled up or hardware actions such as replacing a failing component.

3. After the SMW is ready to rejoin the cluster, run the `clean_resources` command as described in Procedure 22 on page 67. This command also resets all failcounts to zero.

   **Note:** After running the `clean_resources`, wait several minutes for cluster activity to settle. You can check cluster status with the `crm_mon -1` command.

4. Return the SMW to online status as the passive SMW.

   **Note:** Replace *smw2* with the host name of the failed SMW.

   ```
   smw1: # crm node online smw2
   ```

## 5.7.2 Manually Causing Failover

From an operational perspective, both SMWs are equal peers. Because the recommended access is by virtual host name, it usually does not matter which SMW is active and which is passive. However, the following procedure can be used to switch their roles.

**Procedure 21. Performing a manual failover**

1. As `root` on the active SMW, put the active SMW into standby mode. This command forces a failover, which stops all resources on the active SMW and moves them to the passive SMW.

   ```
   smw1: # crm node standby smw1
   ```

   At this point, the other SMW (`smw2`) is now the active SMW.

2. Bring the previously active SMW (`smw1`) online as the passive SMW.

   ```
   smw1: # crm node online smw1
   ```

3. Check the cluster status.

   ```
   smw1: # crm_mon -1
   ```

# Troubleshooting an SMW HA System [6]

This chapter describes how to troubleshoot issues on an SMW HA system.

## 6.1 If Resources Are Not Running

A successful failover should restart all cluster-managed resources on the other SMW. Unless the cluster has been placed into maintenance mode, stopped resources do not occur during normal operation. A stopped resource usually indicates that the resource's failcount on the active SMW has exceeded the migration threshold for that resource and that no failover is possible because the passive SMW is offline, unclean, or is also not eligible to run that resource (typically, because its failcount has also exceeded the migration threshold).

**Procedure 22. Restarting stopped resources**

You can use this procedure on either the active or passive SMW. Execute the commands in this procedure as `root`.

1. Use the following commands to check the status of cluster resources:

   - Execute the `crm_gui` command, then check the 'Management' display and verify that all resources are marked with green circles (see Figure 2).

   - Execute the following command as `root` on either SMW.

     ```
     smw1: # crm resource status
      stonith-1      (stonith:external/ipmi) Stopped
      stonith-2      (stonith:external/ipmi) Started
      dhcpd  (lsb:dhcpd) Started
      cray-syslog    (lsb:cray-syslog) Started
      ClusterIP      (ocf::heartbeat:IPaddr2) Started
      ClusterIP1     (ocf::heartbeat:IPaddr2) Started
      ClusterIP2     (ocf::heartbeat:IPaddr2) Started
      ClusterIP3     (ocf::heartbeat:IPaddr2) Started
      ClusterIP4     (ocf::heartbeat:IPaddr2) Started
      fsync  (ocf::smw:fsync) Started
      homedir        (ocf::heartbeat:Filesystem) Started
      hss-daemons    (lsb:rsms) Started
      Notification   (ocf::heartbeat:MailTo) Stopped
      Resource Group: HSSGroup
          ml-fs      (ocf::heartbeat:Filesystem) Started
          md-fs      (ocf::heartbeat:Filesystem) Started
          mysqld     (ocf::heartbeat:mysql) Started
     ```

   The status `Stopped` is usually caused by exceeding the failcount for a resource.

2. Display the failcount data for all resources.

```
smw1:~# show_failcounts
node=smw1 scope=status  name=fail-count-stonith-1 value=0
node=smw1 scope=status  name=fail-count-stonith-2 value=0
node=smw1 scope=status  name=fail-count-dhcpd value=0
node=smw1 scope=status  name=fail-count-cray-syslog value=0
...
```

**Note:** You can also use the `show_failcount` command to display the failcount data for a single resource on the specified SMW.

**Note:** Replace *smw* with the SMW host name.

```
smw1: # show_failcount smw fsync
```

3. Clear the failcounts and return all values to zero.

```
smw1: # clear_failcounts

Clearing failcount on node smw1
Clearing failcount on node=smw1 for resource=stonith-1
Clearing failcount on node=smw1 for resource=stonith-2
Clearing failcount on node=smw1 for resource=dhcpd
Clearing failcount on node=smw1 for resource=cray-syslog
Clearing failcount on node=smw1 for resource=ClusterIP
 .
 .
 .
Clearing failcount on node=smw2 for resource=hss-daemons
Clearing failcount on node=smw2 for resource=Notification
Clearing failcount on node=smw2 for resource=ml-fs
Clearing failcount on node=smw2 for resource=md-fs
Clearing failcount on node=smw2 for resource=mysqld
```

4. After all failcounts have been cleared, the resource should be up and running. Check the cluster status again to verify that the resource has been restarted.

```
smw1: # crm resource status
 stonith-1      (stonith:external/ipmi) Started
 stonith-2      (stonith:external/ipmi) Started
 dhcpd  (lsb:dhcpd) Started
 cray-syslog    (lsb:cray-syslog) Started
 ClusterIP      (ocf::heartbeat:IPaddr2) Started
 ClusterIP1     (ocf::heartbeat:IPaddr2) Started
 ClusterIP2     (ocf::heartbeat:IPaddr2) Started
 ClusterIP3     (ocf::heartbeat:IPaddr2) Started
 ClusterIP4     (ocf::heartbeat:IPaddr2) Started
 fsync  (ocf::smw:fsync) Started
 homedir        (ocf::heartbeat:Filesystem) Started
 hss-daemons    (lsb:rsms) Started
 Notification   (ocf::heartbeat:MailTo) Started
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem) Started
     md-fs      (ocf::heartbeat:Filesystem) Started
     mysqld     (ocf::heartbeat:mysql) Started
```

5. If not all resources have started, execute the `clean_resources` command.

```
smw1: # clean_resources
Cleaning resources on node smw1
Cleaning resource on node=smw1 for resource=stonith-1
Cleaning resource on node=smw1 for resource=stonith-2
Cleaning resource on node=smw1 for resource=dhcpd
Cleaning resource on node=smw1 for resource=cray-syslog
Cleaning resource on node=smw1 for resource=ClusterIP
Cleaning resource on node=smw1 for resource=ClusterIP1
Cleaning resource on node=smw1 for resource=ClusterIP2
 ...
Cleaning resources on node smw2
Cleaning resource on node=smw2 for resource=stonith-1
Cleaning resource on node=smw2 for resource=stonith-2
 ...
Cleaning resource on node=smw2 for resource=Notification
```

## 6.2 If an SMW is Powered Off

If a critical event happens, one of the SMWs may be powered off. After resolving the problem, use the following procedure to power on the SMW and join it to the cluster.

**Procedure 23. Returning an SMW to the HA cluster after it has been powered off**

1. As `root` on either SMW, check the SMW status with the `crm_mon` command.

```
smw1: # crm_mon -1
============
Last updated: Mon Jul 15 15:32:58 2013
Last change: Wed Jun 26 11:35:09 2013 by root via crm_attribute on smw1
Stack: openais
Current DC: smw1 - partition WITHOUT quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw1 ]
OFFLINE: [ smw2 ]

 stonith-2      (stonith:external/ipmi):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw11
...
```

2. Determine the cause of the problem and resolve it before continuing with this procedure.

3. On the active SMW, put the passive SMW into standby mode.

   **Note:** Replace *smw2* with the host name of the passive SMW.

```
smw1: # crm node standby smw2
```

4. Check the power status of the passive SMW.

   **Note:** Replace *smw2-DRAC-IP-addr* with the passive SMW's DRAC IP address.

```
smw1: # /usr/bin/ipmitool -I lan -U root -H smw2-DRAC-IP-addr -a chassis power status
Password:
Chassis Power is off
```

   **Note:** At the `Password:` prompt, enter the root password for the DRAC.

5. If the power status is `off`, use the following command to turn power on.

   **Note:** Replace *smw2-DRAC-IP-addr* with the passive SMW's DRAC IP address.

```
smw1: # /usr/bin/ipmitool -I lan -U root -H smw2-DRAC-IP-addr -a chassis power on
```

6. Verify the changed power status.

   **Note:** Replace *smw2-DRAC-IP-addr* with the passive SMW's DRAC IP address.

```
smw1: # /usr/bin/ipmitool -I lan -U root -H smw2-DRAC-IP-addr -a chassis power status
Password:
Chassis Power is on
```

   **Note:** At the `Password:` prompt, enter the root password for the DRAC.

7. Wait for the SMW to reboot.

8. Join the passive SMW to the cluster.

   **Note:** Replace *smw2* with the host name of the passive SMW.

```
smw1: # crm node online smw2
```

## 6.3 If STONITH Keeps Killing an SMW

If the cluster manager keeps killing one or both SMWs, it usually means that the cluster has lost the heartbeat because of a communication issue. In this situation, check that the `eth2` and `eth4` cables are connected correctly on each SMW. For more information, see Network Connections on page 12.

## 6.4 If Failover Occurs During the `xtbootsys` Boot Process

The SMW failover feature does not support failover during boot and certain other HSS operations. If a failover should occur during an `xtbootsys` session, it is possible that an HSS database lock could remain in place, causing subsequent boot attempts to fail with a `"Locking components failed"` message that is similar to this example:

```
...
Network topology: class 2

              Nodeid|       Flags: Result
-------------------------------------------------------------------
                  p0|    noflags|: Locking components failed
-------------------------------------------------------------------
Sat Feb 02 11:38:22 CST 2013
It took 0 seconds for 'xtcli' to complete.
'xtcli -s boot SNL0 -o bootnode p0' completed with status 254
WARNING: 'xtcli' could not find any nodes to boot
'crms_boot_loadfile SNL0 bootnode p0 linux' apparently failed
...
```

If this problem occurs, use the following procedure to clear the HSS lock.

**Procedure 24. Clearing an HSS lock after failover occurs during mainframe boot**

1. As `crayadm` on the active SMW, determine the lock ID.

   ```
   crayadm@smw1:~> xtcli lock show
   Network topology: class 2
   ======== SM Session Info ========
   :3:s0:  mtoken=0
   session id:     1
   time       :    Sat Feb  2 11:22:16 2013
   target type:    rt_node
   members:        c0-0
   --------------------------------
   ```

   In this example, the line "`:3:s0:  mtoken=0`" indicates that service number 3 (boot manager) holds a lock. The lock ID is shown in the line "`session id: 1`", indicating a lock ID of `1`.

2. On the active SMW, manually clear the lock.

   **Note:** Replace *id-number* with the actual lock ID.

   ```
   crayadm@smw1:~> xtcli lock -u id-number
   Network topology: class 2
   ```

3. Verify that the lock has been cleared.

   ```
   crayadm@smw1:~> xtcli lock show
   Network topology: class 2
   ======== SM Session Info ========
   No session found in the SM.
   ```

4. If the lock remains in place, log into the active SMW as `root` and restart HSS.

   **Note:** Replace *smw1* with the host name of the active SMW.

   ```
   crayadm@smw1:~> ssh root@smw1
   Password:
   ...
   smw1: # /etc/init.d/rsms restart
   ```

5. Ensure that CLE is not running (that is, the boot node is not partially or fully booted) before running `xtbootsys` again.

   ```
   crayadm@smw:~> ping boot
   ```

## 6.5 If System Settings Are Lost After a Failover During Discovery

If a failover occurs during the system discovery process, you must rerun the `xtdiscover` command on the other SMW once failover completes. However, if a failover occurs early in the `xtdiscover` process, manual intervention is required to recover the system settings.

After rerunning `xtdiscover`, check whether the prompts for the boot node, SDB node, and cpio archive have reverted to default settings (`c0-0c0s0n1`, `c0-0c1s0n1`, and `/raw0,` respectively) instead of the system's previous settings. Information about disabled nodes may also be missing.

`xtdiscover` automatically backs up the HSS database before starting, so previous settings can be recovered if this problem occurs.

**Note:** The HSS database is saved in `/home/crayadm/hss_db_backup/db_backup.*.sql`.

**Procedure 25. Recovering system settings after failover during discovery**

1. Restore the previously saved HSS database, as described in the NOTES section of the `xtdiscover`(8) man page. The recovery procedure is the same as that for a system with a single SMW.

2. Rerun `xtdiscover`.

## 6.6 Correcting Problems with File Synchronization

If file synchronization does not seem to be working, use the following procedure to check for basic problems.

**Procedure 26. Checking for file synchronization problems**

1. Check the `/var/opt/cray/log/smwmessages` file for the following `file is marked dirty` error or any other `fsync` errors.

   ```
   While syncing file /etc/corosync/corosync.conf:
   ERROR from peer hex-14: File is also marked dirty here!
   Finished with 1 errors.
   ```

2. Check that exactly one `corosync` process is running on each SMW. Execute the following `ps` command on **both** SMWs.

   ```
   smw1: # ps h -C corosync
   10840 ?        Ssl    3:45 /usr/sbin/corosync
   smw1: # ssh smw2
   ...
   smw2: # ps h -C corosync
    7621 ?        Ssl    2:44 /usr/sbin/corosync
   ```

   If you see multiple `corosync` processes, continue to the following procedure.

**Procedure 27. Stopping extra `corosync` processes**

If file synchronization (`csync2`) does not seem to be working and no explanation can be found in the `/var/opt/cray/log/smwmessages-`*yyyymmdd* file, use the following procedure to ensure that both SMWs are running one (and only one) instance of the `corosync` daemon.

1. Check the status of the `corosync` process on both SMWs.

2. Stop the OpenAIS service on both SMWs.

**Important:** Stopping OpenAIS is likely to trigger a failover.

```
smw1: # /etc/init.d/openais stop
Stopping OpenAIS/corosync daemon (corosync): 1
.2
.3
.4
.5
.6
.7
.8
.9
.10
.11
.done OK

smw2: # /etc/init.d/openais stop
Stopping OpenAIS/corosync daemon (corosync): 1
.2
.3
.4
.5
.6
.7
.8
.9
.10
.11
.done OK
```

3. Verify that `corosync` is no longer running on either SMW.

```
smw1: # ps h -C corosync

smw2: # ps h -C corosync
```

4. If `corosync` is still running on either SMW, attempt to kill it manually.

```
smw1: # killall -9 corosync

smw2: # killall -9 corosync
```

5. Once no `corosync` processes are running on either SMW, restart OpenAIS on both SMWs.

```
smw1: # /etc/init.d/openais start
Starting OpenAIS/Corosync daemon (corosync): starting... OK

smw2: # /etc/init.d/openais start
Starting OpenAIS/Corosync daemon (corosync): starting... OK
```

# Disabling an SMW HA Cluster [A]

If problems occur during system configuration and testing, it may be helpful to convert the SMW HA cluster back to two unclustered SMWs. The following procedure describes how to disable the cluster configuration without uninstalling cluster software, so that the system can be converted back to a cluster in the future. To avoid the interference between the two SMWs, the second SMW will be powered off.

The following information is required for this procedure:

- DRAC IP address of the both SMWs.
- Virtual host name and IP address for the cluster.
- Device names of the shared directories on the boot RAID (`/home`, `/var/lib/mysql`, and `/var/opt/cray/disk/1`). These device names were defined the `smwha_args` file during cluster configuration; see step 12 in Procedure 5 on page 21.

**Procedure 28. Disabling an SMW cluster**

**Note:** The examples in this procedure show the host names `cray-smw` (virtual host name for the cluster), `smw1`, and `smw2`. Substitute the actual host names for your system.

1. Log in as `root` to the first SMW.
2. Put both SMWs in standby mode.

   **Note:** Replace *smw1* with the host name of the first SMW. Replace *smw2* with the host name of the second SMW.

   ```
   smw1: # crm node standby smw1
   smw1: # crm node standby smw2
   ```

3. On each SMW, edit the `/etc/sysconfig/network/ifcfg-eth*` files to restore the original IP addresses for the Ethernet ports (`eth1`, `eth2`, `eth3`, and `eth4`).

   **Note:** This step assumes that your site uses the standard fixed IP addresses for these ports. If your site uses different IP addresses, ensure that the final digit in the dotted quad is `1`, not `2` or `3` (that is, use the virtual IP address).

   a. In `/etc/sysconfig/network/ifcfg-eth1`, change the `IPADDR` value to `'10.1.1.1/16'`.

b. In `/etc/sysconfig/network/ifcfg-eth2`, change the `IPADDR` value to `'10.2.1.1/16'`.

c. In `/etc/sysconfig/network/ifcfg-eth3`, change the `IPADDR` value to `'10.3.1.1/16'`.

d. In `/etc/sysconfig/network/ifcfg-eth4`, change the `IPADDR` value to `'10.4.1.1/16'`.

4. Set up the service configuration.

   a. Execute the following commands on `smw1`:

```
smw1: # chkconfig openais off
smw1: # chkconfig cray-mzwatcher on
smw1: # chkconfig xinetd off
smw1: # chkconfig dbMonitor on
smw1: # chkconfig mysql on
smw1: # chkconfig rsms on
```

   b. On `smw2`, execute the following `chkconfig` command:

```
smw2: # chkconfig openais off
```

5. From the first SMW, power off the second SMW (`smw2`). For *drac-ip-address*, specify the DRAC IP address of `smw2`.

```
smw1: # /usr/bin/ipmitool -I lan -U root -H drac-ip-address -a chassis power off
```

**Note:** Enter the root password when prompted.

6. Change the actual host name and IP address of the first SMW to the virtual host name and IP address of the cluster, so that users can use the same name to access the system. (For example, change `smw1` to `cray-smw`.)

   a. Execute `yast2` to open the **YaST2 Control Center**

```
smw1: # yast2
```

   **Note:** For the GUI version of YaST, either execute this command on the SMW console or connect via an `ssh` connection with X11 port forwarding.

   b. In the left menu panel, select the **Network Services** group.

   c. In the right panel, select **Hostnames**. The **Host Configuration** window opens.

   d. Select `smw1` from the **Current Hosts** list, then click **Edit**. A pop-up window displays the IP address, host name, and host aliases for the selected host.

e. Change to the cluster's virtual host name and IP address.

- **IP Address**: Enter the cluster's virtual IP address.
- **Hostname**: Enter the cluster's virtual host name (for example, `cray-smw`).
- **Host Aliases**: Enter the cluster's virtual host name (for example, `cray-smw`).

f. Click **OK** to confirm your changes.

g. Click **Finish** to save your changes and close the **Host Configuration** window.

h. Close the main YaST2 window to exit `yast2`.

7. Edit `/etc/fstab` to restore to the mount points for the local disks. Add mount points to the local disks for `/var/lib/mysql` (MySQL database) and `/var/opt/cray/disk/1` (log directory).

   After this change, the SMW will use the local disks for the MySQL database and log directory.

   **Note:** The original data is left on the boot RAID. If you need to recover the shared data, you can mount the boot RAID disks and copy the data to the local disks. The RAID disk names are recorded in the `smwha_args` file.

8. Reboot the first SMW.

```
smw1: # reboot
```

The first SMW now functions as a standard, unclustered SMW with the host name `cray-smw`.