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

S–0044–C

RECORD OF REVISION

S–0044–C Published March 2014 Supports the release of Cray SMW High Availability Extension for SLES 11 SP3.

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 Cray SMW High Availability Extension for SLES 11 SP2.

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

Contents

Page

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 SP3 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 SP3 (SLEHA11SP3), and the Cray SMW HA software. Additional procedures in this guide summarize the installation of the required SMW base operating system and the Cray SMW and Cray Linux Environment (CLE) software. See Chapter 3, Installing the Cray SMW HA Release Package on page 17.
- Performing site-specific customization (also called *migration*), for a new preinstalled SMW HA system; see Chapter 4, Customizing a Preinstalled SMW HA System on page 41.
- Updating or upgrading the operating system and Cray software on your SMW HA system. Additional procedures in this guide summarize the upgrade and update of the SMW base operating system and the Cray SMW and CLE software. See Chapter 5, Upgrading or Updating the Cray SMW HA System on page 49.
- Administering the SMW HA system; see Chapter 6, Managing Your Cray SMW HA System on page 77.
- Troubleshooting the SMW HA system; see Chapter 7, Troubleshooting an SMW HA System on page 97.
- Disabling and re-enabling the SMW HA configuration in case of problems; see Appendix A, Disabling and Re-enabling an SMW HA Cluster on page 107.

This guide also includes an overview of the Cray SMW HA system; see Chapter 2, Overview on page 11.

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 procedures for installing the Cray SMW and CLE software, configuring the boot RAID, and updating or upgrading the SMW, CLE, or base operating system 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
- *Upgrading the SMW Base Operating System to SLES 11 SP3* (S–0047)

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 (SLE HA) Extension software, the Pacemaker Cluster Resource Manager (CRM), and related tools.

- *SLE High Availability Extension SP3 High Availability Guide*

  This document is available online at suse.com: https://www.suse.com/documentation/sle_ha/

1.2 Distribution Media

The Cray SMW High Availability Extension for SLES 11 SP3 release distribution media includes the following DVD, which contains the Cray SMW HA software package.

- `Cray SMW HA SLE 11 SP3 (SMW SLEHA11SP3)`

  Note: For an initial installation and upgrade installation, you will also need the release media for the operating system, SMW software, and (optionally) CLE software. For more information, see the *SMW HA Release Notes* and *SMW HA README*, which are provided with your SMW HA release package. Also see the release notes and README files that are provided with the SMW and CLE release packages.

Part I: Installation Guide

Overview [2]

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 SMW*s).

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 `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

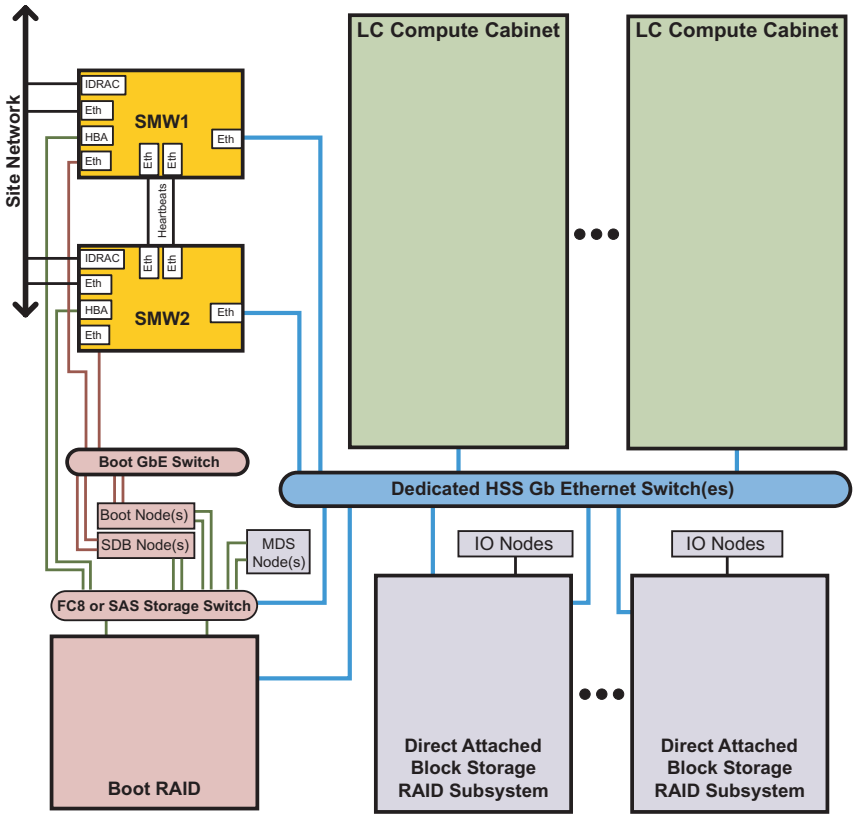
Note: The Pacemaker Cluster Resource Manager uses the term *node* to refer to a host in a CRM cluster. Throughout this document, unless specified otherwise, node refers to an SMW, not to a Cray XC30 compute or service node.

2.1 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 18.

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.

During initial installation, the SMW that is initially configured to be the first active SMW is called the *primary SMW*. The SMW that is initially configured to be passive is called the *secondary SMW*. However, either SMW can be active during normal operation. The cluster configuration does not remember which SMW was initially configured to be active. For more information, see Installing the First SMW With OS and SMW Software on page 21.

2.2 Shared Storage

The SMW HA system uses shared disk devices on the boot RAID for data that must be highly available. The following 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.

- Log disk (`/var/opt/cray/disk/1`), which symbolically links to the following `/var/opt/cray/*` directories:
  - `/var/opt/cray/dump`
  - `/var/opt/cray/install`
  - `/var/opt/cray/log`
- MySQL HSS database (`/var/lib/mysql`). Note that the HSS database server runs on the active SMW only.
- Home directories (`/home`).

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 22.

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.

Note: For systems running the power management feature, the Postgres database runs on both SMWs.

2.3 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 and directories 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 SMW. 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 do 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 89.

2.4 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 97.

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 the Log directory, `/var/opt/cray/disk/1`.

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

`mysqld` Controls and monitors MySQL.

2.5 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 78.
- 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.

Installing the Cray SMW HA Release Package [3]

To completely install a new Cray SMW HA system, you must install the SLES operating system and the SMW, CLE, and SMW HA software. The procedures in this chapter describe how to perform an initial software installation.

Note: Cray ships systems with installation and most of the configuration completed. Unless you need to reinstall the SMW HA system, you can skip the procedures in this chapter. To change the configuration of a preinstalled system, see Chapter 4, Customizing a Preinstalled SMW HA System on page 41.

This chapter uses the following conventions to refer to these SMWs:

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

3.1 Preparing to Install

3.1.1 Before You Start the SMW HA Release Installation

- Read the *SMW HA Release Notes*, the *SMW HA README*, and the *SMW HA Release Errata* to confirm the required versions for the operating system, SMW, and SMW HA software and to determine if there are 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 before you start the installation to ensure the following:
  - Ensure that the prerequisites are satisfied before beginning an initial SMW software installation.
  - Verify that the two SMWs are correctly cabled and that network connections are in place (see Network Connections on page 18).
  - Identify the configuration values for your system (see Configuration Values on page 19).
- Before you install the SMW and CLE software, plan space on the boot RAID for the shared storage for the SMW HA system. For more information, see Shared Storage on page 13 and Configuring the Boot RAID for SMW HA on page 22.
- 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.

3.1.2 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.

3.1.3 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
Virtual host name for SMW HA cluster	`virtual-smw`	
Host name for primary SMW	`smw1`	
Host name for secondary SMW	`smw2`	
DRAC host name on primary SMW	`smw1-drac`	
DRAC host name on secondary SMW	`smw2-drac`	
Virtual IP address for SMW HA cluster	173.31.73.165	
IP address for primary SMW	173.31.73.60 (see Note, below)	
IP address for secondary SMW	173.31.73.61 (see Note, below)	
DRAC IP address on primary SMW	172.31.73.77	
DRAC IP address on secondary SMW	172.31.73.79	

Note: The IP addresses for the virtual SMW HA cluster (`virtual-smw`), primary SMW (`smw1`), and secondary (`smw2`) must be on the same subnet.

3.1.4 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-smw1`)	`initial0`
`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.

3.2 Installing the First SMW With OS and SMW Software

For the first SMW in the SMW HA cluster, the procedures to install the SLES operating system and SMW software are essentially the same as for a single SMW, with a few HA-specific differences. The following procedure summarizes the steps for installing this software.

You can begin the installation on either SMW. The SMW that is installed first (the *primary SMW*) will initially become the active SMW when the SMW HA cluster is fully configured. The examples in this chapter show the host name `smw1` for the primary SMW.

During this procedure, you will need to refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480). Use the procedures for a rack-mount SMW on a Cray XC30 system.

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 the full SMW software. Instead, update the existing SMW to the required SMW release software, then continue to the next section.

Procedure 1. Installing the first SMW with OS and SMW software

1. Log in as `root` to the first SMW.
2. Prepare the SMW for installing the operating system and SMW software, as described in section 2.1 of *Installing Cray System Management Workstation (SMW) Software* (S–2480) and Preparing to Install on page 17 (this guide).

   Important: 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), as described in Network Connections on page 18 (this guide). These ports are marked as "Reserved for SMW failover" in *Installing Cray System Management Workstation (SMW) Software* (S–2480).
3. Ensure that the boot RAID is disconnected, then install the SMW base operating system by following the procedure in section 2.2 of *Installing Cray System Management Workstation (SMW) Software* (S–2480).

   Important: When installing the 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.
4. Install and configure the SMW software packages by following the procedure in section 2.4 of *Installing Cray System Management Workstation (SMW) Software* (S–2480).
5. Continue with the other procedures in chapter 2 of *Installing Cray System Management Workstation (SMW) Software* (S–2480) to configure the SMW software on the Cray system.

   Note: You can skip the procedure to back up the newly-installed SMW software (section 2.10 of *Installing Cray System Management Workstation (SMW) Software*, S–2480). The system should be backed up after the SMW HA configuration is complete.

3.3 Configuring the Boot RAID for SMW HA

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

During this procedure, you will need to refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480) and *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Table 4 shows the minimum partition sizes for these additional LUNs. 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 2. 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 and zone your boot RAID, including the required LUNs for the SMW HA cluster as specified in Table 4. Follow the procedures in chapter 3 of *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 installing the CLE software and 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 section 4.4.2 in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

3.4 Installing CLE Software on the First SMW

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 CLE software. Instead, update the existing SMW to the required CLE release software, then continue to the next section.

For the first SMW, the installation procedures for the CLE software are the same as for a single SMW, with a few HA-specific differences. The following procedure summarizes the steps for installing this software.

During this procedure, you will need to refer to *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Procedure 3. Installing CLE on the first SMW

Note: You must be logged in as `root` for this procedure.

1. Prepare the SMW system for installing the CLE software, as described in chapter 2 of *Installing and Configuring Cray Linux Environment (CLE) Software*.
2. Configure the boot RAID as described in chapter 3 of *Installing and Configuring Cray Linux Environment (CLE) Software*. Include the additional three LUNs for shared storage for the SMW HA system, as described in Configuring the Boot RAID for SMW HA on page 22 (this guide).
3. To install the CLE software on the SMW, follow the procedures in chapter 5.1 of *Installing and Configuring Cray Linux Environment (CLE) Software*.
4. To configure the CLE software on the Cray system, continue with the other procedures in Chapter 5 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

3.5 Installing the Second SMW with OS and SMW Software

For the second SMW, you must skip several steps when installing the base operating system and SMW software. The following procedure summarizes the steps for installing this software.

The SMW that is installed second (the *secondary SMW*) will initially become the passive SMW when the SMW HA cluster is fully configured. The examples in this chapter show the host name `smw2` for the secondary SMW.

During this procedure, you will need to refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480). Use the procedures for a rack-mount SMW on a Cray XC30 system.

Important: The second SMW **must** run exactly the same version of operating system and Cray SMW software as the first SMW.

Procedure 4. Installing the second SMW with OS and SMW software

1. Log in as `root` to the second SMW.
2. Prepare the second SMW for installing the operating system and SMW software, as described in section 2.1 of *Installing Cray System Management Workstation (SMW) Software* (S–2480) and Preparing to Install on page 17 (this guide).

   Note: Skip the procedure to shut down the Cray system (section 2.1.2 of *Installing Cray System Management Workstation (SMW) Software*). Because the second SMW is not yet connected to the Cray system, system shutdown is not required at this point.

Important: 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 18 (this guide). These ports are marked as "Reserved for SMW failover" in *Installing Cray System Management Workstation (SMW) Software*.

3. Ensure that the boot RAID is disconnected, then install the SMW base operating system by following the procedure in section 2.2 of *Installing Cray System Management Workstation (SMW) Software*.

   Important: As on the first SMW, when installing the 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.

4. Use the following substeps to install and configure the required portions of the SMW software. You **must** skip several steps on the second SMW.

   a. Install and configure the SMW software packages by following all applicable steps in section 2.4 of *Installing Cray System Management Workstation (SMW) Software*. There are no differences for the second SMW.

   b. Skip the hardware discovery and power-up (section 2.5 of *Installing Cray System Management Workstation (SMW) Software*). Discovery and power-up was done when installing the first SMW.

   c. Skip the procedure to confirm that the SMW is communicating with the system hardware (section 2.6 of *Installing Cray System Management Workstation (SMW) Software*). This procedure was already done for the first SMW.

   d. Change the default SMW passwords after completing installation, as directed in section 2.7 of *Installing Cray System Management Workstation (SMW) Software*.

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

   e. If you set up the SUSE firewall and IP tables on the first SMW, set up the same configuration on the second SMW, as described in section 2.8 of *Installing Cray System Management Workstation (SMW) Software*.

   f. Skip the procedure to collect additional data during a system dump session (section 2.9 of *Installing Cray System Management Workstation (SMW) Software*). 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.

g. As on the first SMW, you can skip the procedure to back up the newly-installed SMW software (section 2.10 of *Installing Cray System Management Workstation (SMW) Software*). The system should be backed up after the SMW HA configuration is complete.

3.6 Installing CLE Software on the Second SMW

On the second SMW, only the actual CLE installation procedure is required. Do not repeat the configuration procedures. The following procedure summarizes the steps for installing this software.

During this procedure, you will need to refer to *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Procedure 5. Installing CLE on the second SMW

1. Prepare the second SMW (`smw2`) for installing the CLE software, as described in section 5.1, Procedure 2, of *Installing and Configuring Cray Linux Environment (CLE) Software*.

2. To install the Cray CLE software on the SMW, use the following steps instead of the full installation procedure.

   Important: Use the following steps rather than Procedure 3 in *Installing and Configuring Cray Linux Environment (CLE) Software*. Installing CLE on the second SMW requires the `-X Aries` option to the `CRAYCLEinstall.sh` command.

   a. As `root`, execute the installation script to install the Cray CLE software on the SMW.

   ```
   smw2:~ # /home/crayadm/install.5.2.14/CRAYCLEinstall.sh \
   -m /home/crayadm/install.5.2.14 -X Aries -v -i -w
   ```

   b. At the prompt '`Do you wish to continue?`', type **`y`** and press `Enter`.

   The output of the installation script displays on the console.

   Note: If this script fails, you can restart it with the same options. However, rerunning this script may generate numerous error messages as it attempts to install already-installed RPMs. You may safely ignore these messages.

3. Skip the other procedures in Chapter 5 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

3.7 Installing the SMW HA Release Package

After the SMW and CLE software has been installed, install the SMW HA release package on both SMWs.

Important: Before beginning this procedure, ensure that both SMWs are running the same base operating system and SMW software. In addition, 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 22.

Procedure 6. Installing the SMW HA release package

Important: You **must** start the SMW HA installation on the first (primary) SMW, which was completely installed and configured with the SMW and CLE software. This SMW will initially be the active SMW when the system is fully configured. The other SMW is the secondary SMW, which will be the passive SMW initially.

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

1. Log on to both SMWs as `root`.
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-SLEHA11SP3`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP3b06.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.

     Important: The ISO must reside on a local disk (such as `/tmp`), not on a shared file system on the boot RAID. For example, do not use a subdirectory in `/home`.

```
smw:~ # mount -o loop,ro /path/SMW-SLEHA11SP3xnn.iso /media/cdrom
```

3. Navigate 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. Navigate out of the `/media/cdrom` directory and unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

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

6. Repeat this procedure on the other SMW.

3.8 Configuring the Cluster

When you configure the SMW HA cluster, the primary SMW (`smw1`) becomes the active SMW. The secondary SMW (`smw2`) becomes the passive SMW.

You will need the following information from Table 1, Table 2, and Table 4 when configuring an SMW HA cluster:

- Host names of two SMWs, such as `smw1` and `smw2`
- Virtual IP address for the SMW cluster
- IP addresses for both SMWs
- 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 the MySQL database, Log, and `/home` directories

Procedure 7. Configuring the SMW HA cluster

1. Log on to both SMWs as `root`.

   Important: Log in as `root` via `ssh`. Do not use `su` from a different account.

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

```
smw1:~ # /etc/init.d/vnc stop
```

   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 the other SMW (`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 primary SMW, and replace *smw2* with the host name of the secondary 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 primary SMW, and replace *smw2* with the host name of the secondary 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 at first (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 at first (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 primary 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 the following commands, replace *smw2* with the host name of the secondary (initially passive) SMW.

    a. On `smw1`, copy the `ssh` host keys to `smw2`.

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

    b. On `smw2`, restart the `ssh` daemon.

    ```
    smw2:~ # /etc/init.d/sshd restart
    ```

    c. On `smw1`, verify that passwordless `ssh` is still functional to the other SMW.

    ```
    smw1:~ # ssh smw2
    ```

    Note: After running `ssh`, you might need to answer the prompt or perform the specified action to complete the connection.

    d. After connecting to `smw2`, exit the connection.

    ```
    smw2:~ # exit
    ```

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

    Important: You **must** be logged in as `root` (via `ssh`). If you used `su` from a different account, you must log out and log back in as `root` to reset your login environment.

    ```
    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 `smw1`, 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 2 on page 23), Replace the default contents with the values for your site.

    Note: Make the following changes in this file:

    - Under `virtual_ip`, replace the *IPaddressA* value with the virtual IP address for the SMW cluster.
    - Under `log_disk_name`, replace the *disknameB* value with the by-id disk name for shared log directory.
    - Under `db_disk_name`, replace the *disknameC* value with the by-id disk name for the shared MySQL database.
    - Under `home_disk_name`, replace the *disknameD* value with the by-id disk name for the shared `/home` directories.
    - Under `drac_ip_active`, replace the *IPaddressE* value with the DRAC IP address for the SMW that is initially configured to be active.
    - Under `drac_ip_passive`, replace the *IPaddressF* value with the DRAC IP address for the SMW that is initially configured to be passive.
    - Under `passive_smw_hostname`, replace the *smw2* value with the host name of the SMW that is initially configured to be passive.

```
smw1:~ # vi /opt/cray/ha-smw/default/hainst/smwha_args
--virtual_ip
IPaddressA
--log_disk_name
/dev/disk/by-id/disknameB
--db_disk_name
/dev/disk/by-id/disknameC
--home_disk_name
/dev/disk/by-id/disknameD
--drac_ip_active
IPaddressE
--drac_ip_passive
IPaddressF
--passive_smw_hostname
smw2
--verbose
```

    For more information, see Configuring the Boot RAID for SMW HA on page 22 and the `SMWHAconfig`(8) man page.

13. Execute the following command to ensure that `/home/crayadm/.gvfs` is **not** mounted.

```
smw1:~ # df -a | grep /home/crayadm/.gvfs && umount -f /home/crayadm/.gvfs
```

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

    Important: Before you continue, ensure that nothing is mounted on `/mnt`. The `SMWHAconfig` script uses `/mnt` to set up the shared storage.

    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. You must include the `@` character before the argument `smwha_args`.

    Important: Execute this command only on the active SMW.

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

    Note: After executing `SMWHAconfig`, you might need to answer a prompt or perform the specified action to complete the `ssh` connection.

    c. The `SMWHAconfig` command prompts for a password so that it can configure the SMW HA cluster and the DRAC; enter the `root` password for the SMW.

    d. `SMWHAconfig` automatically loads the HA cluster configuration settings. If `SMWHAconfig` detects existing data in the shared directories on the boot RAID, it prompts for confirmation to erase the data. For an initial installation, answer `yes` each of the three prompts to erase the space on the shared boot RAID.

    Caution: If you reinstall the SMW HA software on an existing SMW HA cluster, `SMWHAconfig` automatically detects if there is existing data in the shared directories on the boot RAID (shared logs, MySQL database, and `/home`). For each directory with existing data, `SMWHAconfig` prompts you for confirmation to reformat. If you answer `yes`, all data for that shared directory will be erased.

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

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

    ```
    smw1:~ # reboot
    ```

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

    ```
    smw2:~ # reboot
    ```

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

3.9 Configuring Boot Image Synchronization

To ensure that boot images are properly synchronized for the SMW HA system, you must set up the boot image directory. The procedure depends on how boot images are specified in the `/etc/sysset.conf` configuration file.

Note: Cray recommends storing boot images as files in a `/`*bootimagedir* directory. If the boot image directory is a raw device (such as `/raw0`), a change is required on `smw2` to allow the SMW HA system to synchronize the boot images. For more information, see section 5.5 in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Procedure 8. Configuring boot image synchronization

Important: The boot image directory must exist on both SMWs.

1. On the primary SMW only (`smw1`): If boot images are specified as files, configure boot image synchronization.

   Note: If the boot image directory is specified as a raw device, skip this step.

   a. Edit the file `/etc/csync2/csync2_cray.cfg`.

   b. In the `group user_group` section, add an entry for `/`*bootimagedir* using the following format:

      Note: Replace *bootimagedir* with the name of your boot image directory.

      ```
      include /bootimagedir/*;
      ```

   c. Save your changes and exit the editor.

   d. To initialize boot image synchronization, copy the boot images to the other SMW. This manual copy operation speeds up future synchronization. Execute the following command for each boot image in the boot image directory.

      Note: In this command, replace *smw1* with the host name of the primary SMW, and replace *smw2* with the host name of the secondary SMW. Replace *bootimagedir* with the name of your boot image directory.

      ```
      smw1:~ # scp -pr smw1:/bootimagedir smw2:/bootimagedir
      ```

2. On the secondary SMW only (`smw2`): If the boot image directory is specified as a raw device (for example `/raw0`), configure boot image synchronization.

   Note: If boot images are specified as files, skip this step.

   a. Create a symbolic link from the physical device name.

      Note: Replace `/dev/disk/by-id`*xxxx* with the persistent device name for the actual device. Replace `/`*rawdevice* with the raw device name.

```
smw2:~ # ln -s /dev/disk/by-idxxxx /rawdevice
```

3.10 Configuring Failover Notification

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

Tip: For information about configuring email on your SMW, see http://www.postfix.org/BASIC_CONFIGURATION_README.html.

Procedure 9. 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
```

3.11 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 10. Verifying the SMW HA configuration

1. Log on to the active SMW as `root` by using the virtual SMW host name (such as `virtual-smw`). After you have logged in successfully, the prompt displays the host name of the active SMW.

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

2. Verify 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 different resource names, group names, or resource 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: After running `clean_resources`, wait several minutes for cluster activity to settle. You can check cluster status with the `crm_mon -1` command.

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

3.12 Backing Up Your Newly-installed SMW HA Software

After installing and configuring the system, back up your current SMW, CLE, 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).

3.13 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 Table 3.

Procedure 11. Changing SMW, DRAC, and STONITH passwords

1. Log on to the active SMW as `root`, using the virtual SMW host name (such as `virtual-smw`). After you have logged in successfully, the prompt displays the host name of the active SMW.

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

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.

Customizing a Preinstalled SMW HA System [4]

Cray ships SMW HA systems that are completely installed and configured with Cray-specific host names and IP addresses. To complete the configuration on-site, you must reconfigure the system with site-specific IP addresses. Optionally, you can change the default host names that were assigned in the preinstalled system.

The SMW HA system requires five unique IP addresses and three host names. You will need the following information from Table 2 when configuring an SMW HA cluster:

- Virtual IP address for the SMW cluster.
- IP addresses of the two SMWs.
- DRAC IP addresses for both SMWs. These IP addresses are used by the DRAC on each SMW.
- Virtual host name for the SMW cluster. Users access the SMW HA cluster using this host name. This chapter uses the example host name `virtual-smw-default` for the preconfigured host name and `virtual-smw-new` for the site-specific host name.
- Host names of the two SMWs. This chapter uses the example host names `smw1-default` and `smw2-default` for the preconfigured host names, and `smw1-new` and `smw2-new` for the site-specific host names.
- IP addresses for the default gateway and name server.

Important: The IP address for the virtual SMW cluster and the two SMWs must be on the same subnet.

Before beginning the site customization, the network administrator or site administrator must assign the IP addresses to the corresponding host names for the SMW HA cluster.

Procedure 12. Customizing a preinstalled SMW HA system

This procedure updates the IP addresses and host names in the following configuration files:

- `/etc/hosts`
- `/etc/hostname`
- `/etc/csync2/csync2.cfg`
- `/etc/csync2/csync2_cray.cfg`
- `/etc/sysconfig/network/ifcfg-eth0`
- `/etc/sysconfig/network/routes`

Note: In this procedure, host names and command prompts are shown as `smw1-default` and `smw2-default` before customization. After customization, the SMW host names are shown as `smw1-new` and `smw2-new`.

Because this procedure changes host names and IP addresses, you must execute this procedure on the SMW consoles rather than logging in remotely.

1. Shut down both SMWs, if they are not already shut off.

   ```
   smw1-default:~ # shutdown -h now
   smw2-default:~ # shutdown -h now
   ```

2. Connect the Ethernet cables to the network.

3. Power on `smw1-default`.

4. Open a terminal on `smw1-default` and log in as `root`.

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

   ```
   smw1-default:~ # yast2
   ```

6. In the right panel, scroll to the **Network Devices** section and select **Network Settings**.

7. In the **Network Settings** window, select the **Overview** tab.

8. Change the network card setup for `smw1-new`.

   a. Select **eth0 Customer Network Ethernet**, then click the **Edit** button.

   b. Enter the IP address of `smw1-new` in the **IP Address** box.

   c. Enter the host name of `smw1-new` in the **Hostname** box.

   d. Click the **Next** button to return to the **Network Settings** window.

9. Define the name servers for `smw1-new`.

   a. In the **Network Settings** window, select the **Hostname/DNS** tab.

   b. Enter the host name of `smw1-new` in the **Hostname** box.

c. Enter the IP addresses of the name servers into the **Name Server** boxes. You can define up to three name servers.

d. Change the domain names in the **Domain Name** box to the actual names for your system.

e. Change the domain names in the **Domain Search** box to the actual names for your system.

10. Change the route settings.

    a. In the **Network Settings** window, select the **Routing** tab.

    b. Enter the IP address for the site network in the **Default Gateway** box.

11. To finish your changes, click the **OK** button. `yast2` writes the configuration changes.

12. Exit `yast2`.

13. Change the synchronization file `/etc/csync2/csync2.cfg`.

    a. Edit `/etc/csync2/csync2.cfg`.

    b. Locate the following lines in the `ha_group` section:

    `host` *smw1-default*
    `host` *smw2-default*

    c. Change these lines to the actual host names for your system, as in this example:

    `host` *smw1-new*
    `host` *smw2-new*

    d. Save your changes and exit the editor.

14. Change the synchronization file `/etc/csync2/csync2_cray.cfg`.

    a. Edit `/etc/csync2/csync2_cray.cfg`.

    b. Locate the following lines in the `cray_group` section:

    `host` *smw1-default*
    `host` *smw2-default*

    c. Change these lines to the actual host names for your system, as in this example:

    `host` *smw1-new*
    `host` *smw2-new*

    d. Locate the following lines in the `user_group` section:

    `host` *smw1-default*
    `host` *smw2-default*

e. Change these lines to the actual host names for your system, as in this example:

```
host smw1-new
host smw2-new
```

f. Save your changes and exit the editor.

15. Change the CRM cluster configuration file.

a. Edit the cluster configuration file.

```
smw1-default:~ # crm configure edit
```

The configuration file opens in the `vi` editor.

b. Locate the following lines.

```
node smw1-default \
node smw2-default \
params ip="virtual-smw-default-ip"
params hostname="smw1-default" ipaddr="drac-smw1-ip-default" userid="root"
params hostname="smw2-default" passwd="initial0" ipaddr="drac-smw2-ip-default"
location stonith-1-loc stonith-1 -inf: smw1-default
location stonith-2-loc stonith-2 -inf: smw2-default
```

c. Change the host names and IP addresses in these lines to the actual values for your system.

```
node smw1-new \
node smw2-new \
params ip="virtual-smw-new-ip"
params hostname="smw1-new" ipaddr="drac-smw1-ip-new" userid="root"
params hostname="smw2-new" passwd="initial0" ipaddr="drac-smw2-ip-new"
location stonith-1-loc stonith-1 -inf: smw1-new
location stonith-2-loc stonith-2 -inf: smw2-new
```

d. Save your changes and exit the editor.

16. Shut down `smw1-default`, and wait for the system to finish shutting down.

17. Power on `smw2-default`.

18. Open a terminal on `smw2-default` and log in as `root`.

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

```
smw2-default:~ # yast2
```

20. In the right panel, scroll to the **Network Devices** section and select **Network Settings**.

21. In the **Network Settings** window, click the **Overview** tab.

22. Change the network card setup for `smw2-new`.

a. Select **eth0 Customer Network Ethernet**, then click the **Edit** button.

b. Enter the IP address of `smw2-new` in the **IP Address** box.

c. Enter the host name of `smw2-new` in the **Hostname** box.

d. Click the **Next** button to return to the **Network Settings** window.

23. Define the name servers for `smw2-new`.

    a. In the **Network Settings** window, select the **Hostname/DNS** tab.

    b. Enter the host name of `smw2-new` in the **Hostname** box.

    c. Enter the IP addresses of the name servers into the **Name Server** boxes. You can define up to three name servers.

    d. Change the domain names in the **Domain Name** box to the actual names for your system.

    e. Change the domain names in the **Domain Search** box to the actual names for your system.

24. Change the route settings.

    a. In the **Network Settings** window, select the **Routing** tab.

    b. Enter the IP address for the site network in the **Default Gateway** box.

25. To finish your changes, click the **OK** button. `yast2` writes the configuration changes.

26. Exit `yast2`.

27. Shut down `smw2-default`, and wait for the system to finish shutting down.

28. Power on `smw1-default` and wait for it to come up.

    Note: After the system powers on, the prompt displays the new host name (for example, `smw1-new`).

29. On `smw1-new`, remove the default nodes from the CRM configuration.

    Note: In the following commands, replace *smw1-default* with the default (pre-configured) host name of the first SMW. Replace *smw2-default* with the default host name of the second SMW.

```
smw1-new:~ # crm node delete smw1-default
INFO: node smw1-default not found by crm_node
INFO: node smw1-default deleted                 =====> deleted

smw1-new:~ # crm node delete smw2-default
INFO: node smw2-default not found by crm_node
INFO: node smw2-default deleted                 =====> deleted
```

    For each command, the second message confirms that the node has been deleted. You may safely ignore the first message that the node is not found.

30. Restart the OpenAIS service on `smw1`.

```
smw1-new:~ # rcopenais stop
smw1-new:~ # rcopenais start
```

31. Power on `smw2-default` and wait for it to come up.

    Note: After the system powers on, the prompt displays the new host name (for example, `smw2-new`).

32. Copy the synchronization files `/etc/csync2/csync2.cfg` and `/etc/csync2/csync2_cray.cfg` from `smw1-new` to `smw2-new`.

    Note: Replace *smw2-new* with the actual host name of the second SMW.

```
smw1-new:~ # scp /etc/csync2/csync2.cfg smw2-new:/etc/csync2/
smw1-new:~ # scp /etc/csync2/csync2_cray.cfg smw2-new:/etc/csync2/
```

33. Synchronize the `csync` files between `smw1-new` to `smw2-new`.

```
smw1-new:~ # csync2 –xv
```

    If all files are synchronized successfully, `csync2` will finish with no errors.

34. Display the cluster status.

```
smw1-new:~ # crm_mon -1
============
Last updated: Tue Aug 20 17:36:34 2013
Last change: Tue Aug 20 14:15:34 2013 by root via cibadmin on smw1-new
Stack: openais
Current DC: smw1-new - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Online: [ smw1-new smw2-new ]

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

Note: `crm_mon` may display different resource names, group names, or resource order on your system.

35. Verify that all resources have started. If necessary, see Procedure 10 on page 37 for additional steps to examine cluster status and fix problems with stopped resources or failed actions.

36. 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).

37. Change the default SMW passwords as described in Changing Default SMW Passwords After Completing Installation on page 38.

Note: To further customize your SMW HA cluster, see Customizing the SMW HA Cluster on page 88. This section describes how to change the email address for failover notification, add site-specific files and directories to the synchronization list, and change the migration threshold for SMW HA cluster resources.

Upgrading or Updating the Cray SMW HA System [5]

Cray provides periodic upgrades or updates to the SMW, CLE, and SMW HA software releases, as well as infrequent updates to the base operating system version running on the SMW. The procedures in this chapter describe how to install these upgrades or updates.

Important: Cray recommends a concurrent upgrade for the following items:

1. Operating system software
2. SMW software
3. SMW HA software

You can upgrade the CLE software at any point in the process.

For each upgrade release package, upgrade the active SMW first, then upgrade the passive SMW.

This chapter uses the following conventions to refer to the SMWs:

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

5.1 Before You Start

- Read the *SMW HA Release Notes* and the *SMW HA README* provided with your SMW HA release package to confirm the required versions for the operating system, SMW, and SMW HA software release, as well as the supported upgrade paths.
- To determine whether there are any changes to the procedures in this chapter, read the *SMW HA README* and *SMW HA 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 list of synchronized files and directories in `/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 upgrade or update packages, 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.

Note: An SMW HA system requires more time to upgrade or update, as compared to a system with a single SMW, because you must install the software on both the primary and secondary SMW. Allow at least two hours of additional time to complete an upgrade or update.

5.2 Upgrading Operating System Software

For a system running SLEHA 11 SP2, you must upgrade to the SLES 11 SP3 operating system before upgrading the SMW and SMW HA software.

Important: Cray recommends a concurrent upgrade for the following items:

1. Operating system software
2. SMW software
3. SMW HA software

You can upgrade the CLE software at any point in the process.

Procedure 13. Upgrading the operating system software

To upgrade the operating system software on an SMW HA system, upgrade the active SMW first, then upgrade the passive SMW. You **must** complete the upgrade on the active SMW before starting to upgrade the passive SMW.

Note: During this procedure, you will need to refer to the operating system upgrade procedures in *Upgrading the SMW Base Operating System to SLES 11 SP3* (S–0047).

1. Log on to 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
```

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

3. Upgrade the active SMW (`smw1`) by following the applicable procedures in *Upgrading the SMW Base Operating System to SLES 11 SP3*.

   a. Back up your current software, as described in section 1.1 of *Upgrading the SMW Base Operating System to SLES 11 SP3*.

   b. Shut down the Cray system, as described in section 1.2 of *Upgrading the SMW Base Operating System to SLES 11 SP3*.

      Important: Ensure that the boot RAID is powered off or disconnected before continuing.

   c. Upgrade the SMW base operating system to SLES 11 SP3, as described in section 1.3 of *Upgrading the SMW Base Operating System to SLES 11 SP3*. Follow the procedure for a rack-mount SMW.

      The installation process automatically reboots the SMW to finish setting up SP3. After the reboot completes, reconnect the boot RAID to the SMW, then reboot the SMW again to ensure that the boot RAID connection is recognized correctly.

      Important: You **must** complete the operating system upgrade on `smw1` before you start the upgrade on `smw2`.

   Tip: After the operating system has been upgraded on `smw1`, Cray recommends that you upgrade the operating system on `smw2` before upgrading the SMW and SMW HA software. However, if you need to complete all installation tasks on `smw1` before working on `smw2`, you can go to Procedure 24 on page 68, step 1 through step 3 **only**, then return to this procedure to continue with the next step. Do **not** upgrade the SMW HA cluster configuration until the SMW HA update release has been installed on both SMWs.

4. Upgrade the other SMW (`smw2`) by following the applicable procedures in *Upgrading the SMW Base Operating System to SLES 11 SP3*.

   a. **Skip** the step to back up your current software. This step was already done on `smw1`.

   b. **Skip** the step to shut down the Cray system. This step was already done on `smw1`.

   c. Upgrade the operating system to SLES 11 SP3, as described in section 1.3 of *Upgrading the SMW Base Operating System to SLES 11 SP3*. Follow the procedure for a rack-mount SMW.

      The installation process automatically reboots the SMW to finish setting up SP3. After the reboot completes, you must reconnect the boot RAID to the SMW, then reboot the SMW again to ensure that the boot RAID connection is recognized correctly.

5. Create a single bootable backup drive, as described in appendix A of *Upgrading the SMW Base Operating System to SLES 11 SP3*.

After upgrading the operating system, continue to the following section to upgrade the SMW software.

5.3 Upgrading the SMW HA System

For a system running SMW 7.1.UP01 (or earlier) software, the procedures in this section are required for the first upgrade to the SMW 7.2.UP00 release package.

If the system is already running the SMW 7.2.UP00 (or later) software, use the procedures in Updating the SMW HA System on page 64.

Important: Cray recommends a concurrent upgrade for the following items:

1. Operating system software
2. SMW software
3. SMW HA software

You can upgrade the CLE software at any point in the process.

For each upgrade release package, upgrade the active SMW first, then upgrade the passive SMW.

Note: During these procedures, you will need to refer to the SMW release installation procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480) and the CLE release installation procedures in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

5.3.1 Preparing for an Upgrade

Procedure 14. Preparing for an upgrade

1. Log on to 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.

   ```
   smw1:~ # /etc/init.d/vnc stop
   ```

   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.

   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 (`smw1`), shut down the system.

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

6. Stop file synchronizing.

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

7. On the active SMW, turn on maintenance mode.

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

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

Continue to the next procedure to upgrade the SMW software on the active SMW.

5.3.2 Upgrading SMW Software on the Active SMW

Procedure 15. Upgrading SMW software on the active SMW

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the upgrade.

1. Mount the `/home` directory from the boot RAID. Use the persistent (by-id) device name for the mount command.

   Note: In the following command, replace `scsi-`*xxxxxxxxx* with the persistent (by-id) device name for the `/home` directory on the boot RAID.

   ```
   smw1:~ # mount /dev/disk/by-id/scsi-xxxxxxxxx /home
   ```

   Tip: Execute this command as `root` to display the configured device names.

```
smw1:~ # crm configure show | grep device | awk '{print $2 " " $3}' | sed 's/"//g'

device=/dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01 directory=/home
device=/dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf directory=/var/lib/mysql
device=/dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9 directory=/var/opt/cray/disk/1
```

2. Install the SMW software update on the active SMW (`smw1`).

   Refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480). Follow the applicable steps in Procedure 20 and Procedure 21, including the step to reboot the SMW.

   Important: Before you continue, wait until the SMW has finished rebooting.

3. Update the SMW HA `mysqld` resource to use the new path for MySQL 5.5.

   a. Display the current paths for the `mysqld` resource.

   ```
   smw1:~ # crm resource param mysqld show binary
   /opt/MySQL/default/sbin/mysqld
   smw1:~ # crm resource param mysqld show client_binary
   /opt/MySQL/default/bin/mysql
   ```

   b. Change to the new paths for the server and client.

```
smw1:~ # crm resource param mysqld set binary /usr/sbin/mysqld
smw1:~ # crm resource param mysqld set client_binary /usr/bin/mysql
```

   c. Verify the changes.

   ```
   smw1:~ # crm resource param mysqld show binary
   /usr/sbin/mysqld
   smw1:~ # crm resource param mysqld show client_binary
   /usr/bin/mysql
   ```

4. Put the passive SMW into standby mode.

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

   ```
   smw2:~ # crm node standby smw2
   ```

5. On the active SMW (`smw1`), turn off maintenance mode.

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

6. Continue with the remaining update procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

7. Execute the `clean_resources` command on `smw1`.

```
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
```

After running `clean_resources`, wait several minutes for cluster activity to settle.

Tip: You can check cluster status with the `crm_mon -1` command.

Continue to the next procedure to upgrade the SMW software on the passive SMW.

5.3.3 Upgrading SMW Software on the Passive SMW

Procedure 16. Upgrading SMW software on the passive SMW

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the upgrade.

1. Bring the passive SMW online (take `smw2` out of standby mode).

```
smw2:~ # crm node online smw2
```

2. 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
```

Note: Ignore the failover errors. The failover operation will not complete successfully, because the second SMW has not been upgraded yet.

3. On the second SMW (`smw2`), turn on maintenance mode.

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

4. Mount the `/home` directory from the boot RAID. Use the persistent (by-id) device name for the mount command.

   Note: In the following command, replace `scsi-`*xxxxxxxxx* with the persistent (by-id) device name for the `/home` directory on the boot RAID.

   ```
   smw2:~ # mount /dev/disk/by-id/scsi-xxxxxxxxx /home
   ```

   Tip: Execute this command as root to display the configured device names.

```
smw2:~ # crm configure show | grep device | awk '{print $2 " " $3}' | sed 's/"//g'

device=/dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01 directory=/home
device=/dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf directory=/var/lib/mysql
device=/dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9 directory=/var/opt/cray/disk/1
```

5. Install the SMW software on the second SMW (`smw2`).

   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 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.

   d. Skip section 4.4, For Cray XC30 Systems Only: Updating the BC and CC Firmware. This procedure was done when installing the SMW software on the first 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.

   g. At the end of the SMW installation procedure, be sure that the second SMW reboots successfully.

6. On `smw2`, turn off maintenance mode.

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

7. Verify that the SMW HA services have started and that the second SMW is working properly.

```
smw2:~ # crm_mon -1
============
Last updated: Fri Mar  7 16:10:36 2014
Last change: Fri Mar  7 16:09:21 2014 by root via cibadmin on smw2
Stack: openais
Current DC: smw1 - partition with quorum
Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
2 Nodes configured, 2 expected votes
16 Resources configured.
============

Node smw1: standby
Online: [ smw2 ]

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

Note: `crm_mon` may display different resource names, group names, or resource order on your system.

Continue to the next procedure to finish the SMW software upgrade.

5.3.4 Finishing the SMW Upgrade

Procedure 17. Finishing the SMW upgrade

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the upgrade and is currently the passive SMW.

1. Reboot the currently passive SMW (`smw1`), if you did not reboot it during the previous step. Wait for the reboot to complete.

2. From the currently active SMW (`smw2`), bring the first SMW (`smw1`) online.

   Note: Replace *smw1* with the host name of the SMW that was active at the start of the upgrade.

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

3. On `smw1`, start file synchronizing.

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

4. Reboot the first SMW (`smw1`) and wait for the reboot to finish.

   ```
   smw1:~ # reboot
   ```

5. Reboot the second SMW (`smw2`) and wait for the reboot to finish.

   ```
   smw2:~ # reboot
   ```

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

   ```
   smw1:~ # clear_failcounts
   ```

7. 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).

Continue to the next procedure to upgrade the SMW HA software.

5.3.5 Upgrading SMW HA Software

To upgrade the SMW HA software, upgrade the active SMW first, then fail over to the other SMW to upgrade that system.

Important: Before you start, ensure that both SMWs are running the required operating system and SMW software. For more information, see Before You Start on page 49.

Procedure 18. Upgrading SMW HA software

1. Log on to 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
   ```

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

3. Install the SMW HA software update on the active SMW (`smw1`).

   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-SLEHA11SP3`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP3b06.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.

  Important: The ISO must reside on a local disk (such as `/tmp`), not on a shared file system on the boot RAID. For example, do not use a subdirectory in `/home`.

```
smw1:~ # mount -o loop,ro /path/SMW-SLEHA11SP3xnn.iso /media/cdrom
```

b. Put the active SMW (`smw1`) into standby mode. This command causes the system to fail over to the other SMW.

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

c. On `smw1`, navigate 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. Navigate out of the `/media/cdrom` directory and unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

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

f. Restart the OpenAIS service on `smw1`.

   ```
   smw1:~ # rcopenais stop
   smw1:~ # rcopenais start
   ```

g. Put `smw1` back online (take it out of standby mode).

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

4. Install the SMW HA software upgrade on the other SMW (`smw2`). Note that this system became the active SMW in step 3.b.

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`.

```
smw2:~ # 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-SLEHA11SP3`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP3b06.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.

  Important: The ISO must reside on a local disk (such as `/tmp`), not on a shared file system on the boot RAID. For example, do not use a subdirectory in `/home`.

```
smw2:~ # mount -o loop,ro /path/SMW-SLEHA11SP3xnn.iso /media/cdrom
```

b. Put the SMW (`smw2`) into standby mode. This command causes the system to fail back to the first SMW.

```
smw2:~ # crm node standby
```

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

```
smw2:~ # cd /media/cdrom
smw2:~ # ./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. Navigate out of the `/media/cdrom` directory and unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

```
smw2:~ # cd
smw2:~ # umount /media/cdrom
smw2:~ # eject
```

f. Restart the OpenAIS service on `smw2`.

```
smw2:~ # rcopenais stop
smw2:~ # rcopenais start
```

g. Put `smw2` back online (take it out of standby mode).

```
smw2:~ # crm node online
```

5. If you have made local changes to the list of synchronized files and directories in `/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. Navigate to the `/etc/csync2` directory.

   b. Edit the file `csync2_cray.cfg.sav`.

   c. Locate the `group user_group` section in `csync2_cray.cfg.sav`, and copy the include and exclude lines into `csync2_cray.cfg`.

      Note: You can ignore the generic host entries near the top of the file. The `SMWHAconfig` command will restore site-specific host entries later in this procedure.

6. 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.

      ```
      smw1:~ # /etc/init.d/vnc stop
      ```

      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. On the active SMW, 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).

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

```
smw1:~ # clear_failcounts
```

8. 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 different resource names, group names, or resource order on your system.

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

```
smw1:~ # clean_resources
```

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

10. 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
...
```

If there is a problem with the update, see Restoring a Previous SMW HA Configuration After Update Problems on page 105 for information on how to restore the previous configuration.

Continue to the next procedure to upgrade the CLE software.

5.3.6 Upgrading CLE Software

When upgrading the SMW HA system, you can upgrade the CLE software at any point in the process.

Note: During these procedures, you will need to refer to the CLE release installation procedures in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Procedure 19. Upgrading CLE software

Note: This procedure assumes that `smw1` was the active SMW at the start of the upgrade.

1. Upgrade the CLE release software on the active SMW by following all applicable procedures in chapters 7 and 8 of *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

2. If boot images are stored as files, log on to `smw1` as `root` and copy the boot image to the other SMW. This manual copy operation speeds up future synchronization.

   Note: In this command, replace *smw1* with the host name of the primary SMW, and replace *smw2* with the host name of the secondary SMW. Replace *bootimagedir* with the name of your boot image directory, and replace *file* with the name of the boot image.

   ```
   smw1:~ # scp -p smw1:/bootimagedir/file smw2:/bootimagedir/file
   ```

   Important: The *bootimagedir* directory must already exist on the passive SMW.

3. Upgrade the CLE release software on the passive SMW by following the procedures in section 8.2 of *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

   Important: Skip the other procedures in chapter 8 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

5.4 Updating the SMW HA System

For a system running the SMW 7.2.UP00 (or later) software, the procedures in this section are required for updates to the SMW 7.2 release package.

If the system is running the SMW 7.1.UP01 (or earlier) software, use the procedures in Upgrading the SMW HA System on page 52.

To update the SMW software on an SMW HA system, update the active SMW first, then update the passive SMW.

Note: During these procedures, you will need to refer to the SMW release installation procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480) and the CLE release installation procedures in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

5.4.1 Preparing for an Update

Procedure 20. Preparing for an update

1. Log on to 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.

   ```
   smw1:~ # /etc/init.d/vnc stop
   ```

   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.

   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. On the active SMW (`smw1`), turn on maintenance mode.

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

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

Continue to the next procedure to update the SMW software on the active SMW.

5.4.2 Updating SMW Software on the Active SMW

Procedure 21. Updating SMW software on the active SMW

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the update.

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

   Refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480). Follow the applicable steps in Procedure 20 and Procedure 21, including the step to reboot the SMW.

2. After the SMW has finished rebooting, disable then enable maintenance mode to start the cluster resources.

```
smw1:~ # crm configure property maintenance-mode=false 2> /dev/null
smw1:~ # sleep 10
smw1:~ # crm configure property maintenance-mode=true 2> /dev/null
```

3. Continue with the remaining update procedures in *Installing Cray System Management Workstation (SMW) Software* (S–2480).

Continue to the next procedure to update the SMW software on the passive SMW.

5.4.3 Updating SMW Software on the Passive SMW

Procedure 22. Updating SMW software on the passive SMW

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the update.

1. On the active SMW (`smw1`), turn off maintenance mode.

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

2. On the active SMW, force a failover to the passive SMW (`smw2`), then wait 30 seconds for the failover operation to complete.

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

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

```
smw2:~ # clear_failcounts
```

4. Bring the first SMW online.

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

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

5. On the second SMW (`smw2`), turn on maintenance mode.

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

6. Install the SMW software on the second SMW (`smw2`).

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 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.

d. Skip section 4.4, For Cray XC30 Systems Only: Updating the BC and CC Firmware. This procedure was done when installing the SMW software on the first 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.

Continue to the next procedure to finish the SMW update.

5.4.4 Finishing the SMW Update

Procedure 23. Finishing the SMW update

Note: You must be logged into both SMWs as `root`.

The examples in this procedure assume that `smw1` was the active SMW at the start of the update.

1. On the second SMW (`smw2`), turn off maintenance mode.

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

2. On `smw1`, start file synchronizing.

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

3. Check that the `rsms`, `dbMonitor`, and `mysql` services are disabled on both SMWs. These services **must** be off when the SMWs reboot.

```
smw1:~ # chkconfig -list rsms dbMonitor mysql
rsms                0:off  1:off  2:off  3:off  4:off  5:off  6:off
dbMonitor           0:off  1:off  2:off  3:off  4:off  5:off  6:off
mysql               0:off  1:off  2:off  3:off  4:off  5:off  6:off

smw2:~ # chkconfig -list rsms dbMonitor mysql
rsms                0:off  1:off  2:off  3:off  4:off  5:off  6:off
dbMonitor           0:off  1:off  2:off  3:off  4:off  5:off  6:off
mysql               0:off  1:off  2:off  3:off  4:off  5:off  6:off
```

If any of these services are on, use the following commands to turn them off.

```
smw1:~ # chkconfig rsms off
smw1:~ # chkconfig dbMonitor off
smw1:~ # chkconfig mysql off

smw2:~ # chkconfig rsms off
smw2:~ # chkconfig dbMonitor off
smw2:~ # chkconfig mysql off
```

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

```
smw1:~ # reboot
```

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

```
smw2:~ # reboot
```

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

```
smw1:~ # clear_failcounts
```

7. 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).

Continue to the next procedure to update the SMW HA software.

5.4.5 Updating SMW HA Software

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

Important: Before you start, ensure that both SMWs are running the required operating system and SMW software. For more information, see Before You Start on page 49.

Procedure 24. Updating SMW HA software

1. Log on to 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
```

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

3. Install the SMW HA software update on the active SMW (`smw1`).

   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-SLEHA11SP3`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP3b06.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.

  Important: The ISO must reside on a local disk (such as `/tmp`), not on a shared file system on the boot RAID. For example, do not use a subdirectory in `/home`.

```
smw1:~ # mount -o loop,ro /path/SMW-SLEHA11SP3xnn.iso /media/cdrom
```

b. Put the active SMW (`smw1`) into standby mode. This command causes the system to fail over to the other SMW.

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

c. On `smw1`, navigate 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. Navigate out of the `/media/cdrom` directory and unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

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

f. Put `smw1` back online (take it out of standby mode).

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

4. Install the SMW HA software update on the other SMW (`smw2`). Note that this system became the active SMW in step 3.b.

   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`.

  ```
  smw2:~ # 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-SLEHA11SP3`*xnn*`.iso`, where *xnn* is the installer version (for example, `SMW-SLEHA11SP3b06.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.

  Important: The ISO must reside on a local disk (such as `/tmp`), not on a shared file system on the boot RAID. For example, do not use a subdirectory in `/home`.

```
smw2:~ # mount -o loop,ro /path/SMW-SLEHA11SP3xnn.iso /media/cdrom
```

b. Put the SMW into standby mode. This command causes the system to fail back to the first SMW.

   ```
   smw2:~ # crm node standby
   ```

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

   ```
   smw2:~ # cd /media/cdrom
   smw2:~ # ./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. Navigate out of the `/media/cdrom` directory and unmount the SMW HA release media. If you are using a physical DVD, also eject the DVD.

   ```
   smw2:~ # cd
   smw2:~ # umount /media/cdrom
   smw2:~ # eject
   ```

f. Put `smw2` back online (take it out of standby mode).

   ```
   smw2:~ # crm node online
   ```

5. If you have made local changes to the list of synchronized files and directories in `/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. Navigate to the `/etc/csync2` directory.

b. Edit the file `csync2_cray.cfg.sav`.

c. Locate the `group user_group` section in `csync2_cray.cfg.sav`, and copy the include and exclude lines into `csync2_cray.cfg`.

> **Note:** You can ignore the generic host entries near the top of the file. The `SMWHAconfig` command will restore site-specific host entries later in this procedure.

6. 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.

```
smw1:~ # /etc/init.d/vnc stop
```

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. On the active SMW, 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).

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

```
smw1:~ # clear_failcounts
```

8. 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 different resource names, group names, or resource order on your system.

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

```
smw1:~ # clean_resources
```

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

10. 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
...
```

    If there is a problem with the update, see Restoring a Previous SMW HA Configuration After Update Problems on page 105 for information on how to restore the previous configuration.

Continue to the next procedure to update the CLE software.

5.4.6 Updating CLE Software

Note: During these procedures, you will need to refer to the CLE release installation procedures in *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Procedure 25. Updating CLE software

The examples in this procedure assume that `smw1` was the active SMW at the start of the update.

1. Update the CLE release software on the active SMW (`smw1`) by following all applicable procedures in chapters 7 and 8 of *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

2. If boot images are stored as files, log on to `smw1` as `root` and copy the boot image to the other SMW. This manual copy operation speeds up future synchronization.

    Note: In this command, replace *smw1* with the host name of the primary SMW, and replace *smw2* with the host name of the secondary SMW. Replace *bootimagedir* with the name of your boot image directory, and replace *file* with the name of the boot image.

```
smw1:~ # scp -p smw1:/bootimagedir/file smw2:/bootimagedir/file
```

    Important: The *bootimagedir* directory must already exist on the passive SMW.

3. Update the CLE release software on the other SMW (`smw2`) by following the procedures in section 8.2 of *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

    Important: Skip the other procedures in chapter 8 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

Part II: Administration Guide

Managing Your Cray SMW HA System [6]

The administration tasks for an SMW HA system are generally the same as those for a system with a single SMW. This chapter summarizes the operational differences for an SMW HA system and includes the following HA-specific procedures:

- Booting an SMW HA Cluster on page 78
- Logging onto the active SMW; see Logging In on page 79
- Identifying the Active SMW on page 79
- Monitoring the SMW HA Cluster on page 85
- Customizing the SMW HA Cluster on page 88
- Handling Failover on page 91

This chapter also describes the SMW HA commands; see Using SMW HA Commands on page 80.

This chapter uses the following conventions to refer to the SMWs:

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

6.1 Operational Differences on an SMW HA System

The SMW HA system includes the following operational differences:

- On an SMW HA system, you must control the `rsms` daemon as `root` rather than as `crayadm`. In addition, restarting `rsms` behaves differently than on a system with a single SMW. Running `/etc/init.d/rsms restart` does not display the expected output, because the HA cluster returns immediately rather than waiting for the HSS daemons to start.

  Tip: To display the daemon status, run `/etc/init.d/rsms status`.

- Users may notice differences in the behavior of the `find` command for the shared file systems on the boot RAID. By default, `find` does not follow symbolic links (for example, in the log file system). To follow symbolic links, use `find -L`.

- 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.

- 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 system on both SMWs at the same time.

For additional differences, see Limitations of SMW Failover on page 16.

6.2 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 26. 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 has 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 85.

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 85.

6.3 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 27. Logging into the SMW

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

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

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

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

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

6.4 Identifying the Active SMW

Procedure 28. 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 27 on page 79.
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
   ```

6.5 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
```

6.5.1 `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.

For example, execute the following command to display a list of all cluster resources on the system.

```
smw1:~ # crm resource show
 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
 hss-daemons (lsb:rsms): Started
 Notification (ocf::heartbeat:MailTo): Started
 Resource Group: HSSGroup
     homedir (ocf::heartbeat:Filesystem): Started
     ml-fs (ocf::heartbeat:Filesystem): Started
     md-fs (ocf::heartbeat:Filesystem): Started
     mysqld (ocf::heartbeat:mysql): Started
```

To display the status of a single resource, such as `fsync`, execute the following command:

```
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, the following commands display different levels of help:

- `crm help`
- `crm resource help`
- `crm resource failcount help`

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

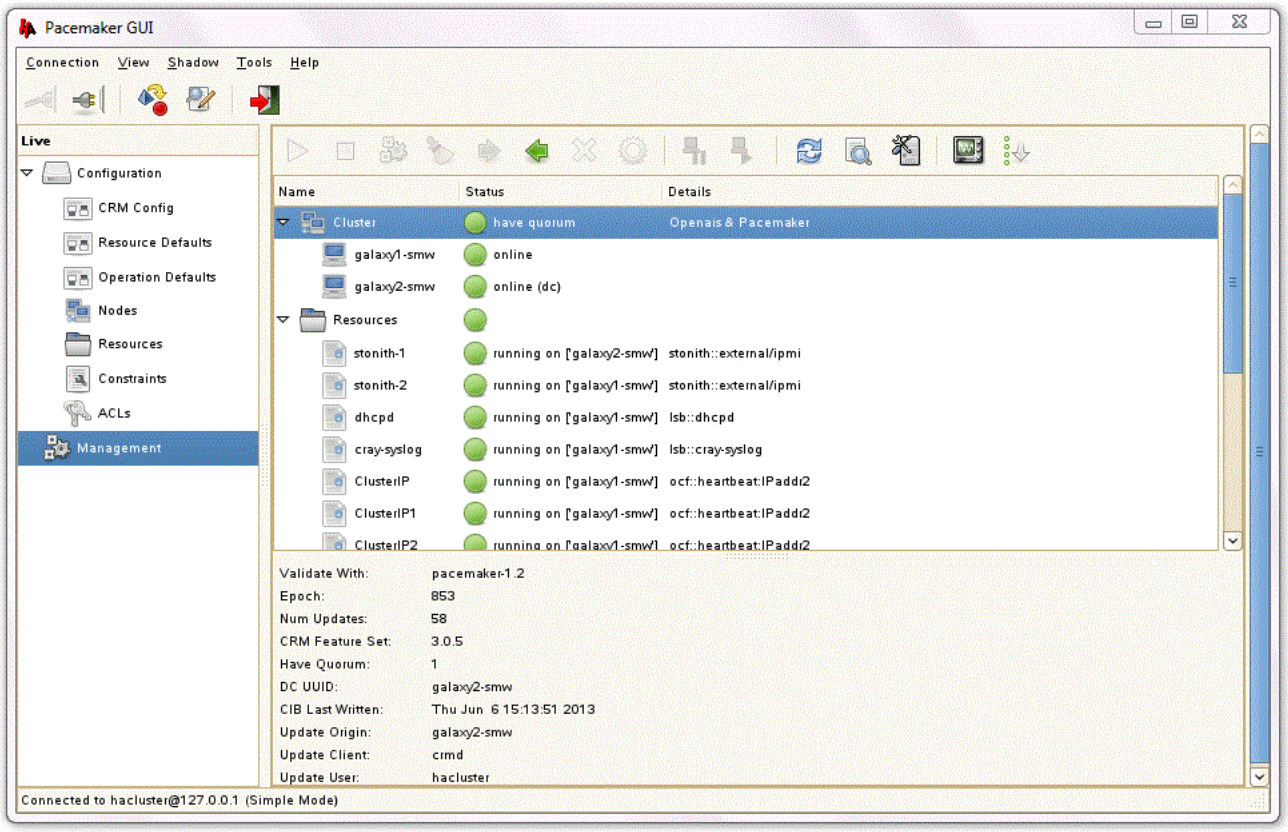
6.5.2 `crm_gui`

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 20.

To display node and resource status, click on **Management** in the left pane.

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



In the management display, a green circle marks a 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 also 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*.

6.5.3 `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: `crm_mon` may display different resource names, group names, or resource order on your system.

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*.

6.5.4 `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*.

6.5.5 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; it 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 the status of each resource 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. The command `crm resource cleanup` also cleans up resources, but requires you to enter each resource name separately.

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

- `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: Execute the `SMWHAconfig` command **only** on the active SMW.

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

6.6 Monitoring the SMW HA Cluster

6.6.1 Displaying Cluster Status

Procedure 29. 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 different resource names, group names, or resource 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 97.

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 97.

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.

6.6.2 Displaying SMW Power Status

Procedure 30. 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
   .
   .
   .
```

   Note: `crm_mon` resource names, group names, or resource order on your system.

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
```

   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 99.

6.7 Customizing the SMW HA Cluster

The SMW HA system is configured during installation. You can customize the system by changing the failover notification address, resource migration threshold, and list of synchronized files.

When customizing the SMW HA system, follow these basic rules:

- Do not change the cluster configuration, except for the migration threshold (maximum failcount value). You can set the migration threshold for each resource by using the `set_migration_threshold` command. For more information, see Cray Cluster Commands on page 83.
- Do not attempt to migrate a single resource. All resources must migrate as a group. For more information, see Cluster Resources on page 14.
- Do not change the system list of synchronized files. You can define which local (site-specific) files are synchronized or excluded from synchronization, but do not add large files or directories to the local list of synchronized files. For more information, see Managing Synchronized Files on page 89.

6.7.1 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 31. 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
```

6.7.2 Managing Synchronized Files

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

Procedure 32. 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 97.

Procedure 33. 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)
- `/tmp/SEDC_FILES`
- 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.

6.7.3 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 34. 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.

6.8 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 16.
- Before using a command that interacts with the HSS daemons, wait for 30 – 60 seconds after failover to ensure that all cluster resources have started. In the first 30 seconds after failover, resources may appear to be started, then change to another state. Although you might be able to log in via the virtual IP address before this period is over, the cluster is not ready for use until all resources are fully started.

  Tip: Use `crm_mon` to verify that all cluster resources have started after failover. For more information, see `crm_mon` on page 82.

6.8.1 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 35. 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 ensuring that the node is healthy. For more information, see Procedure 37 on page 97.

2. Log on to 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 37 on page 97. This command also resets all failcounts to zero.

   Note: After running `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. If the boot node mounts any SMW directories, you must unmount the SMW from the boot node, then remount it.

6.8.2 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 36. 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
   ============
   Last updated: Tue Aug 20 17:36:34 2013
   Last change: Tue Aug 20 14:15:34 2013 by root via cibadmin on smw1
   Stack: openais
   Current DC: smw1-new - partition with quorum
   Version: 1.1.6-b988976485d15cb702c9307df55512d323831a5e
   2 Nodes configured, 2 expected votes
   16 Resources configured.
   ============

   Online: [ smw smw2 ]

    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
    Notification (ocf::heartbeat:MailTo): 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
   ```

   Note: `crm_mon` may display different resource names, group names, or resource order on your system.

Troubleshooting an SMW HA System [7]

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

7.1 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 37. 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 (click on **Management** in the left pane) to verify that all resources are marked with green circles. For more information, see `crm_gui` on page 81.

   - 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
```

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

7.2 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 38. 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
...
```

Note: `crm_mon` may display different resource names, group names, or resource order on your system.

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
   ```

7.3 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 18.

7.4 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 39. 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 on to the active SMW as `root` and restart the RSMS service.

   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
   ```

7.5 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. For more information, see *Managing System Software for the Cray Linux Environment* (S–2393).

Tip: The HSS database is saved in the file `/home/crayadm/hss_db_backup/db_backup.`*MM-DD-YYYY.hhmm*`.sql`.

Procedure 40. 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`.

7.6 Correcting Problems With File Synchronization

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

Procedure 41. 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 42. 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
```

7.7 Restoring a Previous SMW HA Configuration After Update Problems

The `SMWHAconfig` command backs up the cluster configuration before beginning the update. If there are serious problems with the update, you can restore the previous cluster configuration.

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
.
.
.
```

Caution: If you reinstall the SMW HA software on an existing SMW HA cluster, `SMWHAconfig` automatically detects if there is existing data in the shared directories on the boot RAID (shared logs, MySQL database, and `/home`). For each directory with existing data, `SMWHAconfig` prompts you for confirmation to reformat. If you answer `yes`, all data for that shared directory will be erased.

Disabling and Re-enabling an SMW HA Cluster [A]

If problems occur during system configuration and testing, it may be helpful to temporarily disable the SMW HA cluster without uninstalling HA cluster software, then re-enable the cluster after fixing the problems.

A.1 Disabling an SMW HA Cluster

You can disable an SMW HA cluster by converting it to two unclustered SMWs. The active SMW is converted to a stand-alone SMW. The passive SMW is powered off to prevent interference between the two SMWs.

The following information is required for this procedure:

- DRAC IP address of the passive SMW (see Table 2).
- Virtual host name and virtual IP address of the cluster (see Table 2). The examples in this procedure use the virtual host name `virtual-smw`; substitute the actual host name for your system.
- Host names of the active and passive SMWs (see Table 2). The examples in this procedure use the host names `smw1` and `smw2`; substitute the actual host names for your system.
- IP addresses of the original (virtual) Ethernet ports (see Table 1).
- Device names of the shared directories on the boot RAID. You will need the persistent (by-id) device names for the following directories:
  - `/var/lib/mysql` (MySQL database)
  - `/var/opt/cray/disk/1` (Log directory)
  - `/home` (home directories)

  For more information, see Configuring the Boot RAID for SMW HA on page 22.

  Tip: Execute this command as root to display the configured device names.

```
smw1:~ # crm configure show | grep device | awk '{print $2 " " $3}' | sed 's/"//g'

device=/dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01 directory=/home
device=/dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf directory=/var/lib/mysql
device=/dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9 directory=/var/opt/cray/disk/1
```

- Device names of the local MySQL database, Log directory, and home directories. You will need the by-path device names for the following directories on the local disk:
  - `/var/lib/mysql`
  - `/var/opt/cray/disk/1`
  - `/home`

  For more information, see Configuring the Boot RAID for SMW HA on page 22.

Procedure 43. Disabling an SMW HA cluster

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

1. Log on to the active SMW (`smw1`) as `root`.
2. Put both SMWs in standby mode.

   Note: Replace *smw1* with the host name of the active SMW. Replace *smw2* with the host name of the passive SMW.

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

3. On `smw1`, edit the `/etc/sysconfig/network/ifcfg-eth*` files to specify the original IP addresses for the Ethernet ports (`eth0`, `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` (the virtual IP address), not `2` or `3`.

   a. In `/etc/sysconfig/network/ifcfg-eth0`, change the `IPADDR` value to the virtual IP address of the SMW (for example, the IP address for `virtual-smw.cray.com`).

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

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

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

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

4. Disable the SMW HA service configuration.

   Important: The command order is important. Do not change the order of these commands.

```
smw1:~ # chkconfig openais off
smw1:~ # chkconfig mysql on
smw1:~ # chkconfig xinetd off
smw1:~ # chkconfig dbMonitor on
smw1:~ # chkconfig rsms on
smw1:~ # chkconfig dhcpd on
```

5. Power off the passive 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. Edit `/etc/hosts` to replace the active SMW's host name and IP address with the virtual values, so that users can use the same name to access the system. (For example, change `smw1` to `virtual-smw`.)

   a. Locate the line that specifies the active SMW's IP address and host name, as in this example:

```
172.30.49.161   smw1 virtual-smw1
```

   b. Change this line to the virtual IP address and host name of the cluster, as in this example:

```
172.30.49.160   smw virtual-smw
```

7. Execute the following commands to update `/etc/HOSTNAME` with the virtual host name for the cluster.

   Note: This example shows the host names `smw1.us.cray.com` and `virtual-smw.us.cray.com`. Substitute your full host name for the cluster.

```
virtual-smw:~# cat /etc/HOSTNAME
smw1.us.cray.com
virtual-smw:~# echo virtual-smw.us.cray.com > /etc/HOSTNAME
```

8. Define the mount points in `/etc/fstab` for the directories `/var/lib/mysql` (MySQL database), `/var/opt/cray/disk/1` (log directory), and `/home` (home directories). If the boot RAID is available, you can mount the shared directories; otherwise, you must use the local disks.

   a. To use the RAID disks:

      1) Check whether the shared RAID disks are mountable.

         Note: In the following commands, replace `scsi-`*deviceX* with the persistent (by-id) device name for each shared directory on the boot RAID.

```
smw1:~ # mkdir -p /mnt/test
smw1:~ # mount /dev/disk/by-id/scsi-deviceA /mnt/test
smw1:~ # echo $?
0
smw1:~ # mount /dev/disk/by-id/scsi-deviceB /mnt/test
smw1:~ # echo $?
0
smw1:~ # mount /dev/disk/by-id/scsi-deviceC /mnt/test
smw1:~ # echo $?
0
smw1:~ # umount /mnt/test
```

         If the `echo` command displays the value `1` (as the error status for each `mount` command), the shared RAID disks are not mountable. Skip to step 8.b.

      2) Edit `/etc/fstab` to add the RAID disk names.

      3) Change the permissions of directory `/var/lib/mysql/hssds` to `mysql`.

```
smw1:~ # chgrp -R mysql /var/lib/mysql/hssds
smw1:~ # chown -R mysql /var/lib/mysql/hssds
```

      4) Continue to step 9.

   b. If the RAID disks are not available, use the local disks:

      1) Edit `/etc/fstab` and locate the lines containing the by-path device names for `/var/lib/mysql` and `/var/opt/cray/disk/1`. These lines are commented out in a cluster system.

         For example, locate the following lines:

```
# /dev/disk/by-path/pci-0000:05:00.0-sas-phy4-0x4433221104000000-lun-0-part1   /var/opt/cray/disk/1 ...
# /dev/disk/by-path/pci-0000:05:00.0-sas-phy5-0x4433221105000000-lun-0-part1   /var/lib/mysql ...
```

      2) Remove the comment character from these lines, as in this example:

```
/dev/disk/by-path/pci-0000:05:00.0-sas-phy4-0x4433221104000000-lun-0-part1   /var/opt/cray/disk/1 ...
/dev/disk/by-path/pci-0000:05:00.0-sas-phy5-0x4433221105000000-lun-0-part1   /var/lib/mysql ...
```

9. Reboot the SMW.

```
smw1:~ # reboot
```

The formerly active SMW now functions as a standard, unclustered SMW with the cluster's virtual host name (for example, `virtual-smw`). The other (formerly passive) SMW must remain powered off because it is still configured for the SMW HA cluster.

A.2 Re-enabling an SMW HA Cluster

To re-enable a disabled SMW HA cluster, you must undo the changes to the active SMW (`smw1`) that were made in Procedure 43 on page 108. The two SMWs will be returned to the active/passive configuration for the SMW HA cluster.

The following information is required for this procedure:

- Virtual host name and virtual IP address of the cluster (see Table 2).
- Host names of the active and passive SMWs (see Table 2). The examples in this procedure use the host names `smw1` and `smw2`.
- IP addresses of the cluster-specific Ethernet ports for `smw1` and `smw2` (see Table 1).

Procedure 44. Re-enabling an SMW HA cluster

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

1. Log in as `root` to the running SMW (for example, `virtual-smw`).
2. Edit the `/etc/sysconfig/network/ifcfg-eth*` files to restore the cluster-specific IP addresses for the Ethernet ports (`eth0`, `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 `2` or `3`, not `1` (the virtual SMW).

   a. Identify the final digit in the dotted quad of IP addresses for HA cluster.

```
virtual-smw:~# egrep -e '(smw.*smw-net1|smw-net1.*smw)' /etc/hosts /etc/hosts | \
awk '{print $1}' | awk -F"." '{print $4}'
2
```

Note: The returned value is usually 2, which means that the active SMW is `smw1`. The value 3 means that `smw2` is the active SMW; if so, use `3` instead of `2` as the final digit of the dotted quad in the following substeps.

b. In `/etc/sysconfig/network/ifcfg-eth1`, change the `IPADDR` value to `'10.1.1.2/16'`.

c. In `/etc/sysconfig/network/ifcfg-eth2`, change the `IPADDR` value to `'10.2.1.2/16'`.

d. In `/etc/sysconfig/network/ifcfg-eth3`, change the `IPADDR` value to `'10.3.1.2/16'`.

e. In `/etc/sysconfig/network/ifcfg-eth4`, change the `IPADDR` value to `'10.4.1.2/16'`.

f. In `/etc/sysconfig/network/ifcfg-eth0`, change the `IPADDR` value to the cluster's virtual IP address. For example, if the actual IP address for `smw1` is `172.30.49.61`, change this value to `172.30.49.60`.

3. Restore the SMW HA service configuration.

   Important: The command order is important. Do not change the order of these commands.

```
virtual-smw:~# chkconfig rsms off
virtual-smw:~# chkconfig dbMonitor off
virtual-smw:~# chkconfig xinetd on
virtual-smw:~# chkconfig openais on
virtual-smw:~# chkconfig mysql off
virtual-smw:~# chkconfig dhcpd off
```

4. Edit `/etc/hosts` to change the SMW's host name and IP address.

   a. Locate the line that specifies the cluster's virtual IP address and host name, as in this example:

```
172.30.49.160   smw virtual-smw
```

   b. Change this line to the actual IP address and host name of `smw1`, as in this example:

```
172.30.49.161   smw1 virtual-smw1
```

5. Execute the following commands to update `/etc/HOSTNAME` with the actual name for `smw1`.

   Note: This example shows the host names `virtual-smw.us.cray.com` and `smw1.us.cray.com`. Substitute your full host name for `smw1`.

```
virtual-smw:~# cat /etc/HOSTNAME
virtual-smw.us.cray.com
virtual-smw:~# echo smw1.us.cray.com > /etc/HOSTNAME
```

6. Edit `/etc/fstab` to remove the mount points for directories `/var/lib/mysql` (MySQL directory), `/var/opt/cray/disk/1` (Log directory), and `/home` (home directories).

a. If the shared RAID disks are used, remove the mount points for these devices. For more information, see step 8.a in Procedure 43 on page 108.

b. If the local disks are used, comment out (add a comment character to) each line that defines a local disk mount point. For more information, see step 8.b in Procedure 43 on page 108.

7. Change the permission of the `/var/lib/mysql/hssds` directory to `root`.

```
virtual-smw:~# chgrp -R root /var/lib/mysql/hssds
virtual-smw:~# chown -R root /var/lib/mysql/hssds
```

8. Reboot the SMW and wait for it to finish rebooting.

```
virtual-smw:~# reboot
```

When the reboot completes, this SMW is now the active SMW in the cluster (`smw1`).

9. Power on the second SMW (`smw2`) and wait for it to finish rebooting.

The SMW HA configuration is now restored. To verify that the SMW HA cluster is running correctly, see Verifying the Configuration on page 36.