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

S–0044–D

RECORD OF REVISION

S–0044–D Published October 2014 Supports the release of Cray SMW High Availability Extension for SLES 11 SP3 UP01.

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

Procedures

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 XC 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, Install a Cray SMW HA System on page 21.
- Performing site-specific customization (also called *customer configuration*), for a new preinstalled SMW HA system; see Chapter 4, Customize a Preinstalled SMW HA System on page 55.
- Updating or upgrading the operating system and Cray software on the 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 6, Update the Cray SMW HA System on page 81.
- Administering the SMW HA system; see Chapter 7, Manage the Cray SMW HA System on page 97.
- Troubleshooting the SMW HA system; see Chapter 8, Troubleshooting an SMW HA System on page 129.
- Disabling and re-enabling the SMW HA configuration in case of problems; see Appendix A, Disable and Re-enable an SMW HA Cluster on page 139.

This guide also includes an overview of the Cray SMW HA system; see Chapter 2, Overview on page 13.

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 the 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 the 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 SLEHA 11 SP3 UP01 release includes one DVD or ISO file that contains the Cray SMW HA software package. For an initial installation and most upgrade/update installations, you will also need the release media for the operating system, SMW software, and CLE software. For more information, see the *SMW HA README* file provided with the 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, which is based on the SUSE Linux Enterprise High Availability (HA) Extension 11 SP3 software. An additional SMW is required, and the two SMWs must be configured as specified in this document. This feature is supported only for a Cray XC system with two high-end second-generation SMWs (also called *rack-mount SMWs*).

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 fail over 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

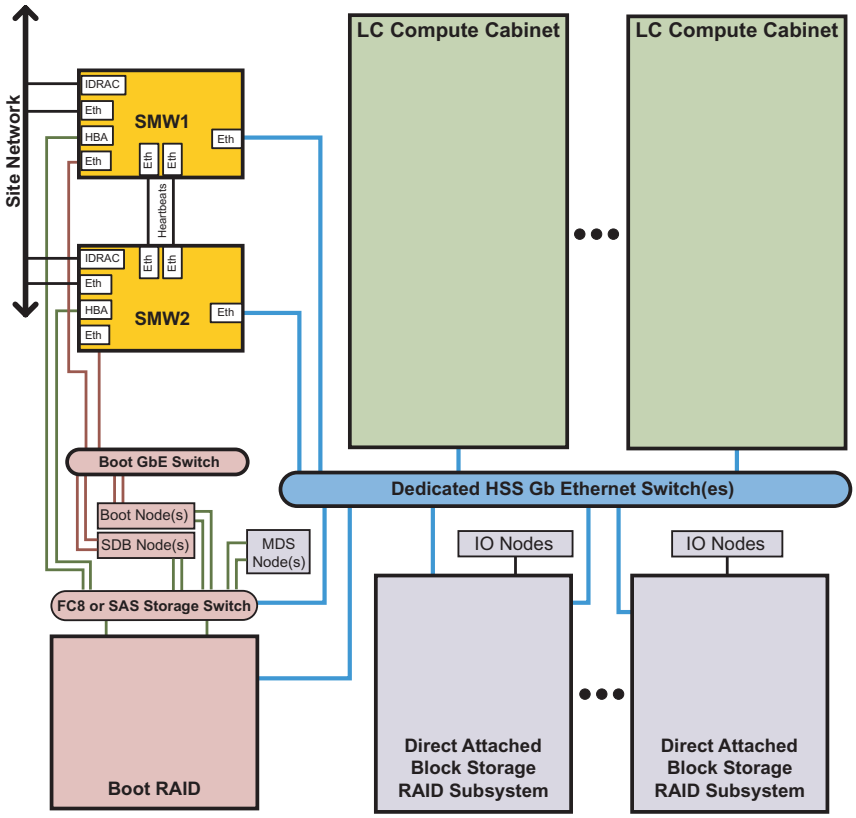
Note: The Pacemaker Cluster Resource Manager uses the term *node* to refer to a host in a CRM cluster. Throughout this document, unless specified otherwise, node refers to an SMW, not to a Cray XC compute or service node.

2.1 SMW Cluster Configuration

Both SMWs are connected to the boot RAID, and are connected to each other 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 Confirm the Required Network Connections on page 23.) An Integrated Dell™ Remote Access Controller (iDRAC) is required on both SMWs.

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 first SMW that is installed and configured becomes the active SMW. The second SMW that is installed and configured becomes the passive 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 Install the First SMW With OS and SMW Software on page 26.

2.2 Shared Storage

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

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 the CLE boot root, the shared root, or any other CLE file systems from the boot RAID on both SMWs at the same time.

The SMW HA system uses shared space on the boot RAID for the following directories:

`/var/opt/cray/disk/1`

Log disk. The following directories symbolically link to the Log disk:

- `/var/opt/cray/debug`
- `/var/opt/cray/dump`
- `/var/opt/cray/log`

`/var/lib/mysql`

MySQL HSS database. Although the database is shared, the HSS database server runs on the active SMW only.

`/home` SMW home directories.

`/var/lib/pgsql`

Power management database (PMDB). The PMDB contains power management data, event router file system (erfs) data, and optional System Environment Data Collections (SEDC) data. Cray strongly recommends using shared storage for the PMDB if sufficient space is available on the boot RAID. If the PMDB is not stored on the boot RAID, the PMDB is split across both SMWs; data collected before an SMW failover will be lost or not easily accessible after failover.

For more information, see Plan Shared Storage for SMW HA on page 22 and Configure the Boot RAID for SMW HA on page 27.

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` lists the Cray-specific files and directories that must be synchronized, as well as small files that are convenient to keep in sync.

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.

The `csync2` utility synchronizes the required files and directories for the SMW HA cluster, such as `/etc/passwd` and `/opt/cray/hss/*/etc/*`. For more information, see Manage Synchronized Files on page 113 or examine the contents of `/etc/csync2/csync2_cray.cfg`.

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

The `csync2` utility is designed to synchronize small amounts of data. If `csync2` must monitor many directories or synchronize a large amount of data, it can become overloaded and failures may not be readily apparent. Cray recommends that you do not change the list of synchronized files (or add only small files); copy large files and directories manually to the other SMW. For more information, see Manage Synchronized Files on page 113.

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: You can reset failcounts with the `clean_resources` or `clear_failcounts` command. For more information, see Resources Are Stopped on page 129.

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 ability to power off that peer at failover time, using the STONITH capability. 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 the shared `/home` directory.

`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.

`ClusterMonitor`

Records failcounts and failed actions in the log file `/var/log/smwha.log` at cluster startup, then clears the failure data from `crm` (for example, in the output of `crm_mon -1`).

`ml-fs` Mounts, unmounts, and monitors the shared log directory, `/var/opt/cray/disk/1`, which symbolically links to the `dump`, `install`, and `log` subdirectories in `/var/opt/cray/`.

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

`mysqld` Controls and monitors MySQL.

`pm-fs` Controls and monitors the power management database (PMDB) file system, `/var/lib/pgsql`.

`postgresqld`

Controls and monitors the power management database (PMDB) PostgreSQL server, `postgresqld`.

2.5 Limitations of SMW Failover

The SMW HA failover feature has the following limitations:

- Both SMWs must run the same versions of SLES and SMW/HSS software.
- System administration of an SMW HA environment is more complex than administration of a system with a single SMW.
- 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.
- 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 capability. Before starting the second SMW, wait until the first SMW has completed startup and initialized all cluster resources. For more information, see Boot an SMW HA Cluster on page 99.
- 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.
- Direct Attached Lustre (DAL) is not supported with the SMW HA failover release.
- If the power management database (PMDB) is not stored on the boot RAID, PMDB data collected before an SMW failover will be lost or not easily accessible after failover.

Install a Cray SMW HA System [3]

A complete initial installation for a new Cray SMW HA system includes 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, Customize a Preinstalled SMW HA System on page 55.

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.

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

3.1 Prepare to Install a New SMW HA System

- 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 Confirm the Required Network Connections on page 23).
  - Identify the configuration values for the system (see Record Configuration Values on page 23).

- Plan space on the boot RAID for the shared storage for the SMW HA system before installing the SMW and CLE software. For more information, see Shared Storage on page 15 and Configure the Boot RAID for SMW HA on page 27.
- For an existing system: Back up the current SMW software before installing the SMW and SMW HA packages.

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

3.1.1 Plan Shared Storage for SMW HA

An SMW HA system requires boot RAID LUNs (Logical Units) for shared storage for the cluster, in addition to the LUNs required for a Cray system with a single SMW. Before installing the SMW and CLE software, plan space on the boot RAID for the shared directories.

Note: Refer to *Installing Cray System Management Workstation (SMW) Software* (S–2480) for the other required LUNs on an SMW.

Table 1 shows the minimum partition sizes for these additional LUNs. A large system may require additional space for the shared directories. Review the requirements of the system in order to determine the appropriate size for these LUNs.

Table 1. Minimum Boot RAID LUN Sizes for SMW Failover

Directory on SMW	Description	Minimum Size
`/var/lib/mysql`	MySQL HSS database	150GB
`/var/opt/cray/disk/1`	Log disk. The following directories symbolically link to the Log disk: • `/var/opt/cray/debug` • `/var/opt/cray/dump` • `/var/opt/cray/log`	500GB
`/home`	SMW home directories	500GB
`/var/lib/pgsql`	Power management database (PMDB)	500GB

Tip: Check the size of `/var/lib/pgsql` and make sure that RAID disk has enough space to hold the PMDB data. Use the following command to display the size of the PMDB:

```
smw1:~ # du -hs /var/lib/pgsql
```

3.1.2 Confirm the Required 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 Record Configuration Values

3.1.3.1 Fixed Configuration Values

An SMW HA cluster uses the following fixed IP addresses. These IP addresses are set by default and are not site dependent.

Table 2. Fixed IP Addresses for an SMW HA system

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 `eth1` connection
10.1.1.2	SMW, `eth1` – Actual `eth1` connection for `smw1`
10.1.1.3	SMW, `eth1` – Actual `eth1` connection for `smw2`

IP Address	Description
10.2.1.1	SMW, `eth2` – Virtual primary heartbeat connection for SMW failover
10.2.1.2	SMW, `eth2` – Actual primary heartbeat connection for `smw1`
10.2.1.3	SMW, `eth2` – Actual primary heartbeat connection for `smw2`
10.2.1.0	Network address to bind to (for `eth2` primary heartbeat connection on `smw2`)
10.3.1.1	SMW, `eth3` – Virtual `eth3` connection
10.3.1.2	SMW, `eth3` – Actual `eth3` connection for `smw1`
10.3.1.3	SMW, `eth3` – Actual `eth3` connection for `smw2`
10.4.1.1	SMW, `eth4` – Virtual redundant heartbeat connection for SMW failover
10.4.1.2	SMW, `eth4` – Actual redundant heartbeat connection for `smw1`
10.4.1.3	SMW, `eth4` – Actual redundant heartbeat connection for `smw2`
10.4.1.0	Network address to bind to (for `eth4` primary heartbeat connection on `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`
1694	Multicast port for primary heartbeat connection (for `eth2` and `eth4` on `smw2`)

3.1.3.2 Site-dependent Configuration Values

An SMW HA system requires the following site-dependent host names and IP addresses. You may find it helpful to record the actual values for the site.

Note: This table lists the HA-specific values only. Refer to *Installing Cray System Management Workstation (SMW) Software*, S–2480) for the other site-dependent values that apply to all systems (with either one or two SMWs).

Table 3. 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 first SMW	`smw1`	
Host name for second SMW	`smw2`	
iDRAC host name on first SMW	`smw1-drac`	
iDRAC host name on second SMW	`smw2-drac`	
Customer network IP address for virtual SMW (the SMW HA cluster)	173.31.73.165	
IP address for first SMW	173.31.73.60	
IP address for second SMW	173.31.73.61	
iDRAC IP address on first SMW	172.31.73.77	
iDRAC IP address on second SMW	172.31.73.79	

Important: The IP addresses for the virtual SMW HA cluster (`virtual-smw`) and the actual SMWs `smw1` and `smw2`) must be on the same subnet.

3.1.4 Passwords For an SMW HA System

On an SMW HA system, the `root` password must be the same on each SMW. In addition, the Integrated Dell™ Remote Access Controllers (iDRAC) password on each SMW must be the same as the `root` password. If you change the default `root` password after installing the SMW software, ensure that the iDRAC `root` accounts use the same password.

The SMW HA configuration process uses the `root` password for the several cluster management accounts (the `hacluster` user, the `stonith-1` resource, and the `stonith-2` resource). When changing passwords after installing the SMW HA software, ensure that these accounts also use the same password as `root` on both SMWs.

Table 4 lists the default values for the passwords that **must** be the same on both SMWs.

Table 4. Default Passwords for an SMW HA System

ID	Default Password
`root` on `smw1`	`initial0`
`root` on `smw2`	`initial0`
`root` (iDRAC) on `smw1`	`initial0`
`root` (iDRAC) on `smw2`	`initial0`
`hacluster` (for logging in to `crm_gui`)	same as `root` password (set during HA configuration)
`stonith-1` resource	same as `root` password (set during HA configuration)
`stonith-2` resource	same as `root` password (set during HA configuration)

For more information, see Change Passwords on an SMW HA System on page 110.

3.2 Install 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 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 first 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 XC 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. Install 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 Prepare to Install a New SMW HA System on page 21 and in section 2.1 of *Installing Cray System Management Workstation (SMW) Software* (S–2480).

Important: Two Ethernet ports are used for heartbeat connections between the two SMWs: `eth2` (on the first quad Ethernet card) and `eth4` (on the second quad Ethernet card), as described in Confirm the Required Network Connections on page 23. 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.

   Tip: You can skip the procedure to back up the newly-installed SMW software (section 2.9 of *Installing Cray System Management Workstation (SMW) Software*, S–2480). The system should be backed up after the SMW HA configuration is complete.

6. If you change the default root password on the SMW, you must also change the default iDRAC password to the same password. For more information, see Passwords For an SMW HA System on page 25.

7. Configure email on the SMW. The SMW HA system uses email for failover notification.

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

3.3 Configure the Boot RAID for SMW HA

After installing the SMW software on the first SMW, configure shared storage on the boot RAID for the MySQL database, log directory, and `/home`. Wait to configure the shared storage for the power management database (PMDB) until after the SMW HA software has been installed.

Before beginning this procedure, determine the space needed for shared storage on the boot RAID, as described in Plan Shared Storage for SMW HA on page 22. For the recommended minimum size for the shared LUNs on the boot RAID, see Table 1.

Procedure 2. Configure the boot RAID for SMW HA

This procedure creates LUNs for the shared directories on the boot RAID.

Note: 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).

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

1. Configure and zone the boot RAID, including the required LUNs for the SMW HA cluster as specified in Table 1. Follow the procedures in chapter 3 of *Installing Cray System Management Workstation (SMW) Software* (S–2480). See *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444) for the recommended boot RAID LUN configuration for a single SMW.
2. Record the LUN numbers for the shared directories so that you can identify the persistent (by-id) device names when you install the CLE software and configure the SMW HA cluster. You may find it helpful to use this table to record this information.

Table 5. Boot RAID LUNs for the Shared Directories

Directory	LUN Number
`/var/lib/mysql`	
`/var/opt/cray/disk/1`	
`/home`	
`/var/lib/pgsql`	

After initial partitioning, 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 *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

Tip: To determine the persistent device name from the LUN number:

1. Use `lsscsi` to show the `/dev/sd*` device name associated with a LUN number (for example, LUN 15). In the first column of the output, the LUN is the final number in the `[n:n:n:n]` value.

   ```
   crayadm@smw1:~> lsscsi
   [0:0:0:0]    disk    ATA      TOSHIBA MK1661GS ME0D  /dev/sda
   [0:0:1:0]    disk    ATA      ST91000640NS     AA03  /dev/sdb
   [0:0:2:0]    disk    ATA      TOSHIBA MK1661GS ME0D  /dev/sdc
   .
   .
   .
   [5:0:0:15]   disk    LSI      INF-01-00        0786  /dev/sdo
   [5:0:0:16]   disk    LSI      INF-01-00        0786  /dev/sdp
   [5:0:0:17]   disk    LSI      INF-01-00        0786  /dev/sdq
   [5:0:0:18]   disk    LSI      INF-01-00        0786  /dev/sdr
   ```

   In this example, LUN 15 is associated with `/dev/sdo`.

2. Use `ls -l` to map the `/dev/sd*` device name to the persistent device name. This example displays the persistent device name for `/dev/sdo` (that is, LUN 15).

```
crayadm@smw1:~> cd /dev/disk/by-id
crayadm@smw1:~> ls -l | grep sdo
lrwxrwxrwx 1 root root 10 Sep  4 00:56 scsi-360080e500037667a000003a2519e3ff2 -> ../../sdo
lrwxrwxrwx 1 root root 10 Sep  4 00:56 wwn-0x60080e500037667a000003a2519e3ff2 -> ../../sdo
```

3.4 Install 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. Install 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 Configure the Boot RAID for SMW HA on page 27.

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 Install 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 XC system.

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

Procedure 4. Install 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 Prepare to Install a New SMW HA System on page 21.

   Note: Skip the step to shut down the Cray system (the first step in section 2.2 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 quad Ethernet card) and `eth4` (on the second quad Ethernet card). Refer to Confirm the Required Network Connections on page 23. 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. Update the controller boot image, as directed in section 2.5 of *Installing Cray System Management Workstation (SMW) Software*. Follow steps 1-2 in procedure 4.

   c. **Skip** the hardware discovery and power-up steps in section 2.5 of *Installing Cray System Management Workstation (SMW) Software*. (Skip step 3 in procedure 4 and all of procedure 5.) The `xtdiscover` and power-up steps were done when installing the first SMW.

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

   e. If you changed the default root and iDRAC passwords on the first SMW, you must also change these passwords on second SMW. Use the same password as on the first SMW. For more information, see Passwords For an SMW HA System on page 25.

      Important: Both SMWs must have exactly the same password for the root and iDRAC accounts.

   f. 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*.

   g. As on the first SMW, you can skip section 2.9 of *Installing Cray System Management Workstation (SMW) Software* (the procedure to back up the newly-installed SMW software). Wait to back up the system until after the SMW HA configuration is complete.

5. Update the `/etc/hosts` file on the second SMW to include information for the Cray system hardware.

a. Create a backup copy of `/etc/hosts` on `smw2`.

```
smw2:~ # cp /etc/hosts /etc/hosts.sav
```

b. Copy `/etc/hosts` from `smw1`.

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

```
smw2:~ # scp smw1:/etc/hosts /etc/hosts
```

6. Configure email on the SMW. The SMW HA system uses email for failover notification.

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

3.6 Install 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. Install CLE on the second SMW

1. Prepare the second SMW (`smw2`) for installing the CLE software, as described in Chapter 5 of *Installing and Configuring Cray Linux Environment (CLE) Software*.
2. Use the following steps to install the Cray CLE software on the SMW, instead of using the full installation procedure in Chapter 5 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

Important: Installing CLE on the second SMW requires the `-X Aries` option to the `CRAYCLEinstall.sh` command.

a. Copy the CLE install directory, `/home/crayadm/install.`*xtrel*, from the first SMW to a local directory on the second SMW (such as `/tmp`). Do not use `/home/crayadm` on the second SMW, because that would create local differences for this shared directory. Replace *xtrel* with the site-determined name specific to the release being installed.

b. Execute following command as `root` to install the Cray CLE software on the second SMW. Include the `-X Aries` option to prevent an `xtdiscover` error on the second SMW.

Note: This example shows `/tmp` as the location of the CLE install directory on the second SMW.

```
smw2:~ # /tmp/install.xtrel/CRAYCLEinstall.sh \
-m /tmp/install.xtrel -X Aries -v -i -w
```

c. 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 can safely ignore these messages.

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

4. Configure firewall and route file changes on the second SMW.

a. Copy the file /var/opt/cray/install/networking_configuration-p0.json from the first SMW to the second SMW.

b. Execute the following command on smw2.

```
smw2:~ # /opt/cray/keystone-cle-config/default/bin/cray_configure_networking.py \
--enable --smw --partition p0 \
--networkjson /var/opt/cray/install/networking_configuration-p0.json \
--servicesjson /opt/cray/keystone-cle-config/default/etc/openstack_services.json \
--nopretend
```

3.7 Install SMW HA Software

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

3.7.1 Verify the SMW and CLE Configuration

Before installing the SMW HA release package, ensure that both SMWs are running the same SMW and CLE software. In addition, the boot RAID must be set up for SMW HA before configuring the SMW HA software release package.

Procedure 6. Verify the SMW and CLE configuration

1. Log in as root to both SMWs.

2. Ensure that both SMWs have the same root password, and that the iDRAC password on each SMW is the same as the root password. For more information, see Passwords For an SMW HA System on page 25.

3. Verify that both SMWs are running the same SMW release.

```
smw1:~ # cat /opt/cray/hss/default/etc/smw-release
7.2.UP02

smw2:~ # cat /opt/cray/hss/default/etc/smw-release
7.2.UP02
```

4. Verify that both SMWs are running the same CLE release.

```
smw1:~ # cat /etc/opt/cray/release/CLEinfo
CLERELEASE=5.2.UP02
INSTALLERVERSION=a16
LUSTRE=yes
NETWORK=ari
XTRELEASE=5.2.40

smw2:~ # cat /etc/opt/cray/release/CLEinfo
CLERELEASE=5.2.UP02
INSTALLERVERSION=a16
LUSTRE=yes
NETWORK=ari
XTRELEASE=5.2.40
```

5. Verify that both SMWs are running the same operating system release.

```
smw1:~ # cat /etc/SuSE-release
USE Linux Enterprise Server 11 (x86_64)
VERSION = 11
PATCHLEVEL = 3

smw2:~ # cat /etc/SuSE-release
USE Linux Enterprise Server 11 (x86_64)
VERSION = 11
PATCHLEVEL = 3
```

6. Ensure that the boot RAID has free LUNS with sufficient space for the shared directories. For more information, see Configure the Boot RAID for SMW HA on page 27.

7. Verify the iDRAC configuration.

   a. Log in as `root` on both SMWs.

   b. Check the iDRAC configuration with the following command:

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

```
smw1:~ # ipmitool -U root -I lan -H smw1-iDRAC-IP-addr -a chassis power status

smw2:~ # ipmitool -U root -I lan -H smw2-iDRAC-IP-addr -a chassis power status
```

      If the iDRAC is configured correctly, these commands return the output `Chassis Power is on` (or `off`). If either of these commands fails, the iDRAC is not configured correctly.

8. Check that both SMWs and iDRACs appear in the DNS.

   Note: In the following commands, specify the actual SMW and iDRAC host names.

   ```
   smw1:~ # ping smw2
   smw1:~ # ping smw2-iDRAC-hostname

   smw2:~ # ping smw1
   smw2:~ # ping smw1-iDRAC-hostname
   ```

   If the iDRACs are not in the DNS (or if the DNS is not available), you can add the iDRAC entries to `/etc/hosts` on both SMWs.

3.7.2 Install the SMW HA Release Package on Both SMWs

Procedure 7. Install the SMW HA release package

Important: You **must** start the SMW HA installation on the first 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 second SMW will initially be the passive SMW.

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

1. Log on to the first SMW (`smw1`) as `root`.
2. 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 the 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
```

3. 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
```

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.

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

6. Reboot the first SMW.

7. Log on to the second SMW (smw2) as root to repeat the installation on the other SMW.

8. 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 the 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
```

9. 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
```

10. 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.

11. 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
    ```

12. Reboot the second SMW.

3.8 Configure the Cluster

After the operating system, SMW, CLE, and HA software has been installed on both SMWs, use the procedures in this section to configure the required cluster settings, boot image synchronization, failover notification, and the power management database (PMDB) on the boot RAID.

3.8.1 Configure Required Cluster Settings

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

This procedure requires site-specific configuration information from Table 1, Table 2, and Table 3, and Table 5, including virtual and actual host names, IP addresses, and disk names of the shared file systems on the boot RAID.

Procedure 8. Configure the SMW HA cluster

Important: During this procedure, do not use a typescript session running directly on the SMW. To save the output of this procedure, use the `script` command to start the typescript session on your local workstation **before** logging into the SMW, as in this example:

```
workstation> script -af my_output_file
Script started, file is my_output_file
workstation> ssh root@smw1
```

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

   Important: You **must** log in directly as `root`. 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 `smw1`.

a. Edit the /etc/sysconfig/network/ifcfg-eth2 file.

```
smw1:~ # vi /etc/sysconfig/network/ifcfg-eth2
```

b. Change the NAME value from 'eth2 Reserved' to 'eth2 SMW HA Heartbeat Network 1'.

c. Verify your changes. The file must have the following contents:

```
BOOTPROTO='static'
IPADDR='10.2.1.1/16'
NAME='eth2 SMW HA Heartbeat Network 1'
PREFIXLEN='16'
STARTMODE='auto'
USERCONTROL='no'
```

Note: Do not change the IPADDR value. The HA configuration process determines the actual values automatically.

d. Save your changes and exit the editor.

e. Copy the ifcfg-eth2 file to ifcfg-eth4.

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

f. Edit the /etc/sysconfig/network/ifcfg-eth4 file.

```
smw1:~ # vi /etc/sysconfig/network/ifcfg-eth4
```

g. Change the IPADDR value from '10.2.1.1/16' to '10.4.1.1/16'.

```
IPADDR='10.4.1.1/16'
```

Note: Use the specified IPADDR value on both SMWs. The HA configuration process determines the actual values automatically.

h. Change the NAME value to 'eth4 SMW HA Heartbeat Network 2'.

i. Verify the changes. The file must have the following contents:

```
BOOTPROTO='static'
IPADDR='10.4.1.1/16'
NAME='eth4 SMW HA Heartbeat Network 2'
PREFIXLEN='16'
STARTMODE='auto'
USERCONTROL='no'
```

j. Save your changes and exit the editor.

4. Repeat the previous step on smw2 to update the ifcfg files for eth2 and eth4 on smw2.

a. Edit the /etc/sysconfig/network/ifcfg-eth2 file.

```
smw2:~ # vi /etc/sysconfig/network/ifcfg-eth2
```

b. Change the NAME value from 'eth2 Reserved' to 'eth2 SMW HA Heartbeat Network 1'.

c. Verify your changes. The file must have the following contents:

```
BOOTPROTO='static'
IPADDR='10.2.1.1/16'
NAME='eth2 SMW HA Heartbeat Network 1'
PREFIXLEN='16'
STARTMODE='auto'
USERCONTROL='no'
```

Note: Do not change the IPADDR value. The HA configuration process determines the actual values automatically.

d. Save your changes and exit the editor.

e. Copy the ifcfg-eth2 file to ifcfg-eth4.

```
smw2:~ # cp /etc/sysconfig/network/ifcfg-eth2 /etc/sysconfig/network/ifcfg-eth4
```

f. Edit the /etc/sysconfig/network/ifcfg-eth4 file.

```
smw2:~ # vi /etc/sysconfig/network/ifcfg-eth4
```

g. Change the IPADDR value from '10.2.1.1/16' to '10.4.1.1/16'.

```
IPADDR='10.4.1.1/16'
```

Note: Use the specified IPADDR value on both SMWs. The HA configuration process determines the actual values automatically.

h. Change the NAME value to 'eth4 SMW HA Heartbeat Network 2'.

i. Verify the changes. The file must have the following contents:

```
BOOTPROTO='static'
IPADDR='10.4.1.1/16'
NAME='eth4 SMW HA Heartbeat Network 2'
PREFIXLEN='16'
STARTMODE='auto'
USERCONTROL='no'
```

j. Save your changes and exit the editor.

5. Update the cluster IP addresses.

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

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

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

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

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

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

6. Initialize smw1 as the active SMW.

a. Execute the `sleha-init` command on `smw1`.

```
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, use the IP address of `eth2` (see Table 2). The multicast address and port are fixed values.

```
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).

7. Join `smw2` as the passive SMW.

a. Execute the `sleha-join` command on `smw2`.

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

b. Answer the following questions to join the passive SMW to the cluster. When asked for the password, use the `root` password for the SMWs.

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

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

Password: root-password-for-SMWs
```

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

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

9. Configure eth4 as the redundant heartbeat channel on smw1.

   a. 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, ensure that **Communication Channels** is selected.

   c. In the right panel, check the **Redundant Channel** check box, then enter the following information to configure eth4 as the redundant channel:

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

      Important: Be careful to start the multicast address with 225, not 255. An incorrect multicast address will prevent the cluster from starting.

   d. Click the **Finish** button.

   e. Close the main YaST2 window to exit yast2.

10. Repeat step 9 on smw2 to configure eth4 as the redundant heartbeat channel on smw2.

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

```
smw2:~ # 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, ensure that **Communication Channels** is selected.

    c. In the right panel, check the **Redundant Channel** check box, then enter the following information to configure eth4 as the redundant channel:

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

Important: Be careful to start the multicast address with 225, not 255. An incorrect multicast address will prevent the cluster from starting.

d. Click the **Finish** button.

e. Close the main YaST2 window to exit yast2.

11. On smw1, synchronize the passive SMW.

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

12. 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 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. Reset the login environment on both SMWs by logging out, then logging back in as root.

Important: You **must** log in to the actual (not virtual) SMW as root. Do not use su from a different account.

```
smw1:~ # exit
workstation> ssh root@smw1

smw2:~ # exit
workstation> ssh root@smw2
```

13. Load the ha-smw module on both SMWs.

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

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

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

14. Edit the SMW HA configuration file,

/opt/cray/ha-smw/default/hainst/smwha_args, to configure the site-specific settings. Replace the following default values with the actual values for the site.

Note: For required host names and IP addresses, see Table 3. For the persistent device names for the shared directories on the boot RAID, see Table 5.

```
smw1:~ # vi /opt/cray/ha-smw/default/hainst/smwha_args

--virtual_hostname
cray-smw
--virtual_ip
172.31.73.165
--log_disk_name
/dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9
--db_disk_name
/dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf
--home_disk_name
/dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01
--drac_ip_active
172.31.73.142
--drac_ip_passive
172.31.73.77
--passive_smw_hostname
cray-smw2
--verbose
```

15. Execute the following command on smw1 to ensure that /home/crayadm/.gvfs is **not** mounted.

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

16. Ensure that nothing is mounted on /mnt. The SMWHAconfig script uses /mnt to set up the shared storage.

```
smw1:~ # df -a | grep mnt
smw1:~ #
```

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

    a. Change to the directory containing the SMWHAconfig command.

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

b. Execute SMWHAconfig on smw1 only, using the modified configuration file as an argument (prefaced by the @ character). If necessary, answer a prompt or perform the specified action to complete the ssh connection.

```
smw1:~ # ./SMWHAconfig @smwha_args
Resource ClusterIP not found: No such device or address
Error performing operation: No such device or address
Usage: grep [OPTION]... PATTERN [FILE]...
Try `grep --help' for more information.
2014-08-22 11:1:56,156: INFO     cdir was created
2014-08-22 11:31:56,361: INFO
*************************Starting of HA software installation*************************

2014-08-22 11:31:56,361: INFO     cluster virtual IP = 172.31.73.165
2014-08-22 11:31:56,361: INFO     log disk (/var/opt/cray/disk/1) = /dev/disk/by-id/scsi-360080e500023bff6000006b15
2014-08-22 11:31:56,361: INFO     db disk (/var/lib/mysql)= /dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf
2014-08-22 11:31:56,362: INFO     home disk (/home)= /dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01
2014-08-22 11:31:56,362: INFO     verbose mode =
.
.
.
```

You can safely ignore the error messages about "Resource ClusterIP not found" and "Error performing operation". These errors are expected because the cluster has not been fully configured yet.

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

d. SMWHAconfig automatically loads the HA cluster configuration settings.

Important: SMWHAconfig deletes any existing data in the shared directories on the boot RAID. For an initial installation, existing data is not reused.

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

Note: You can safely ignore the warning message that the CIB has no configuration element, if this message appears in the SMWHAconfig output.

18. Reboot smw1.

```
smw1:~ # reboot
```

Before continuing, wait until smw1 has rejoined the cluster. After the SMW responds to a ping command, log into smw1, sleep for at least 2 minutes, then execute the crm_mon -1 command to verify that smw1 is online and all resources have started.

19. Reboot smw2.

```
smw2:~ # reboot
```

Before continuing, wait until smw2 has rejoined the cluster. After the SMW

responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

3.8.2 Configure 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 whether boot images are specified in the `/etc/sysset.conf` configuration file as files in a */bootimagedir* directory or as images on a raw device (such as `/raw0`).

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 9. Configure boot image synchronization for images in a */bootimagedir* directory

If boot images are specified as files in a */bootimagedir* directory, use this procedure to configure boot image synchronization.

1. Log in as root on the first SMW (`smw1`).
2. Edit the file `/etc/csync2/csync2_cray.cfg`.

   ```
   smw1:~ # vi /etc/csync2/csync2_cray.cfg
   ```

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

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

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

4. Save the changes and exit the editor.
5. If necessary, create the */bootimagedir* directory on `smw2`. The boot image directory must exist on both SMWs.

   ```
   smw2:~ # mkdir -p /bootimagedir
   ```

6. Copy the boot images from `smw1` to `smw2` to initialize boot image synchronization. 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 *smw2* with the host name of the second SMW. Replace *bootimagedir* with the name of the boot image directory.

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

Procedure 10. Configure boot image synchronization for images on a raw device

If the boot image directory is specified as a raw device (for example `/raw0`), use this procedure to configure boot image synchronization.

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

1. Log in as `root` on the second SMW (`smw2`).
2. Create a symbolic link from the physical device name to the raw device.

   Note: In the following command, 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-id/xxxx /rawdevice
```

3.8.3 Configure 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 an SMW, see http://www.postfix.org/BASIC_CONFIGURATION_README.html.

If email is not configured on the SMW, you can skip this procedure and continue with the SMW HA configuration. After SMW HA is configured and running, you can configure email and failover notification as described in Configure Failover Notifications on page 112.

Procedure 11. Configure failover notification

1. Execute the `crm resource` command.

```
smw1:~ # crm resource param Notification set email address@thedomain.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@thedomain.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
```

3.8.4 Move the PMDB to the Shared Boot RAID

If you have created a LUN on the boot RAID for the power management database (PMDB), move the PMDB to the boot RAID after you have configured the SMW HA system.

Important: Check the size of `/var/lib/pgsql` and make sure that RAID disk has enough space to hold the data. Use the following command to display the size of the local data:

```
smw1:~ # du -hs /var/lib/pgsql
```

Procedure 12. Move the PMDB to the shared boot RAID

Use these steps to configure the RAID disk and transfer the power management data to the power management disk on the shared boot RAID.

Important: During this procedure, do not use a typescript session running directly on the SMW. To save the output of this procedure, use the `script` command to start the typescript session on your local workstation **before** logging into the SMW, as in this example:

```
workstation> script -af my_output_file
Script started, file is my_output_file
workstation> ssh crayadm@smw1
```

1. Shut down the Cray system by typing the following command as `crayadm` on the active SMW (`smw1`).

```
crayadm@smw1:~>xtbootsys -s last -a auto.xtshutdown
```

2. Log into the active SMW as `root`, either at the console or by using the actual (not virtual) host name.

   Important: You **must** log in directly as `root`. Do not use `su` from a different SMW account such as `crayadm`.

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

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

4. Use the `SMWHAconfig` command to move the PMDB and configure the required HA resources.

   Note: Replace *scsi-xxxxxxxxx* with the persistent device name for the PMDB directory on the boot RAID.

```
smw1:~ # ./SMWHAconfig --add_disk=pm-fs \
--device=/dev/disk/by-id/scsi-xxxxxxxxx --directory=/var/lib/pgsql
```

   This command mounts the PMDB directory (`/var/lib/pgsql`) to the boot RAID, copies the PMDB data, and configures the HA resources `pm-fs` and `postgresqld`.

5. Reboot the active SMW (`smw1`).

```
smw1:~ # reboot
```

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw1` is online and all resources have started.

6. Reboot the passive SMW (`smw2`).

```
smw2:~ # reboot
```

   Before continuing, wait until `smw2` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

7. Verify that the power management database is on the boot RAID and that the required resources are running.

   a. Examine the log file

`/opt/cray/ha-smw/default/hainst/SMWHAconfig.out` to verify that the power management database disk appears in the `Cluster RAID Disks` section (at the end of the file), as in this example.

```
---------------- Cluster RAID Disks -----------------
07-07 20:47 INFO     MYSQL Database disk = /dev/disk/by-id/scsi-360080e5...xxx
07-07 20:47 INFO     Log disk            = /dev/disk/by-id/scsi-360080e5...xxx
07-07 20:47 INFO     /home disk          = /dev/disk/by-id/scsi-360080e5...xxx
07-07 20:47 INFO     PM database disk    = /dev/disk/by-id/scsi-360080e5...xxx
07-07 20:47 INFO     ***************** Ending of HA software add_disk *****************
```

b. Ensure that the power management file system is mounted by checking for `/var/lib/pgsql` in the output of the `df` command.

```
smw1:~ # df
     Filesystem     1K-blocks      Used Available Use% Mounted on
     /dev/sda2      120811676  82225412  32449332  72% /
     udev            16433608       756  16432852   1% /dev
     tmpfs           16433608     37560  16396048   1% /dev/shm
     /dev/sdo       483807768 197536596 261695172  44% /var/opt/cray/disk/1
     /dev/sdp       100791728  66682228  28989500  70% /home
     /dev/sdq       100791728    484632  95187096   1% /var/lib/mysql
     /dev/sdr        30237648    692540  28009108   3% /var/lib/pgsql
```

c. Check the output of `crm_mon` to ensure that the `pm-fs` and `postgresqld` resources are running.

```
smw1:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

Tip: To move the PMDB back to the original disk from the shared boot RAID, see Move the PMDB Off the Shared Boot RAID on page 119.

3.8.5 Verify the Cluster 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, wait for 30 to 60 seconds for the cluster system to come up.

Procedure 13. Verify the SMW HA configuration

1. Log in as `root` to the active SMW 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.

   ```
   workstation> ssh root@virtual-smw
   .
   .
   .
   smw1:~ #
   ```

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: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

Note: `crm_mon` may display different resource names, group names, or resource order on the 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 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
```

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

   Tip: You can check cluster status with the `crm_mon -1` command. If the output of this command shows only a subset of the SMW HA services, wait for another minute, then check again.

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

3.9 Back Up a Newly-installed SMW HA Software

After installing and configuring the system, back up the 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).

3.10 Change Default Passwords After Installation

During HA configuration, the passwords for `hacluster` and the `stonith` resources are set to the `root` password for the SMWs. If you changed the default `root` and iDRAC passwords after installing the SMW software, you do not need to change the passwords again. Otherwise, use the following procedure to change the passwords.

Important: The `hacluster`, `stonith`, and iDRAC passwords **must** be the same as the SMW `root` password. For more information, see Table 4.

Procedure 14. Change SMW, iDRAC, and STONITH passwords

1. Log on to the active SMW (`smw1`) 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 on `smw1`:

   ```
   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. Change the SMW `root` and `hacluster` passwords on `smw2`:

   ```
   smw1:~# passwd root
   smw1:~# passwd hacluster
   ```

   Important: Use the same `root` password as on `smw1`. The `hacluster` password **must** be the same as the `root` password.

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

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

Customize 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 3 when configuring an SMW HA cluster:

- 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.
- 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 iDRAC on each SMW.
- 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.

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

Procedure 15. Customize a preinstalled SMW HA system

Before you start this procedure, ensure that the preinstalled system is backed up, as specified in Back Up a Newly-installed SMW HA Software on page 53.

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`.

1. Log in as `root` on each SMW console. Because this procedure changes host names and IP addresses, you must execute this procedure on the SMW consoles rather than logging in remotely.
2. Shut down both SMWs, if they are not already shut off.

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

3. Connect the Ethernet cables to the network.
4. Power on `smw1-default`.
5. Log in as `root` on the `smw1-default` console.
6. Execute `yast2` to open the **YaST2 Control Center**.

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

7. In the right panel, scroll to the **Network Devices** section and select **Network Settings**.
8. In the **Network Settings** window, select the **Overview** tab.
9. 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.
10. 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 name in the **Domain Name** box to the actual name for the system.

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

11. Change the route settings.

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

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

12. If necessary, change the time zone.

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

14. Exit `yast2`.

15. 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 the system, as in this example:

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

    d. Save the changes and exit the editor.

16. 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 the 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 the system, as in this example:

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

f. Save the changes and exit the editor.

17. 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 the 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 the changes and exit the editor.

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

19. Power on `smw2-default`.

20. Log in as `root` on the `smw2-default` console.

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

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

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

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

24. 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.

25. 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 name in the **Domain Name** box to the actual name for the system.

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

26. Change the route settings.

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

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

27. If necessary, change the time zone.

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

29. Exit `yast2`.

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

31. 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`).

32. 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 can safely ignore the first message that the node is not found.

33. Restart the OpenAIS service on `smw1`.

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

34. 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`).

35. 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/
```

36. 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.

37. Display the cluster status.

```
smw1-new:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2-new
Stack: classic openais (with plugin)
Current DC: smw1-new - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1-new smw2-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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1-new
 Notification   (ocf::heartbeat:MailTo):        Started smw1-new
 dhcpd  (lsb:dhcpd):    Started smw1-new
 fsync  (ocf::smw:fsync):       Started smw1-new
 hss-daemons    (lsb:rsms):     Started smw1-new
 stonith-1      (stonith:external/ipmi):        Started smw2-new
 stonith-2      (stonith:external/ipmi):        Started smw1-new
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1-new
     cray-syslog        (lsb:cray-syslog):      Started smw1-new
     homedir    (ocf::heartbeat:Filesystem):    Started smw1-new
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1-new
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1-new
     postgresqld        (lsb:postgresql):       Started smw1-new
     mysqld     (ocf::heartbeat:mysql): Started smw1-new
```

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

38. Verify that all resources have started. If necessary, see Procedure 13 on page 51 for additional steps to examine cluster status and fix problems with stopped resources or failed actions.

39. Back up the 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).

40. Change the cluster passwords as described in Change Default Passwords After Installation on page 53.

For information on optional configuration changes, see the following sections:

- See Customize the SMW HA Cluster on page 112 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.
- See Move the PMDB to the Boot RAID on page 117 to configure the shared boot RAID for the power management database (PMDB).

Upgrade the Cray SMW HA System [5]

Cray provides periodic upgrades to the SMW, CLE, and SMW HA software releases, as well as infrequent upgrades to the base operating system version running on the SMW. A software upgrade installation on an SMW HA system involves moving to the next release for all software components. In an upgrade release, the major version number changes. For example, if your system is currently running the CLE 5.1, SMW 7.1, and SMW HA SLEHA 11 SP2 release, you can upgrade to the CLE 5.1, SMW 7.2, and SMW SLEHA 11 SP3 base releases.

Important: For the SMW SLEHA 11 SP3 release, you must upgrade to the SUSE Linux Enterprise Server version 11 Service Pack 3 (SLES 11 SP3) SMW base operating system before upgrading the SMW and SMW HA software.

The procedures in this chapter describe how to install these upgrades.

Important: You must upgrade the base operating system software, SMW software, and SMW HA software together. Cray recommends upgrading in the following order:

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

You can upgrade the CLE software either before or after the OS, SMW, and SMW HA upgrades.

For each upgrade release package, upgrade the active SMW first, then upgrade the passive SMW. Do not upgrade both SMWs at the same time.

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.

Note: In this document, some examples are left-aligned to better fit the page. Left alignment has no special significance.

5.1 Before You Start

- Read the *SMW HA Release Notes* and the *SMW HA README* provided with the 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.
- Identify any 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 installing the upgrade packages, back up the 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.
- Cray recommends checking all file systems with `fsck` before beginning an upgrade, because upgrading an SMW HA system requires several reboots.

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

5.2 Upgrade 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.

Procedure 16. Upgrade 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*. This step summarizes the tasks required to upgrade the operating system on `smw1`.

   a. Back up the current software.

   b. Shut down the Cray system.

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

   c. Upgrade 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 upgrading the operating system on `smw2` before upgrading the SMW, CLE, and SMW HA software. If necessary, however, you can complete all SMW, CLE and SMW HA software installation on `smw1` before installing the software on `smw2`. Wait until all software is installed on **both** SMWs before configuring the upgraded SMW HA cluster. Do **not** upgrade the SMW HA cluster configuration (as described in Upgrade SMW HA Software on page 73) 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*. This step summarizes the tasks required to upgrade the operating system on `smw2`.

   a. Back up the current software.

   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. 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.

5. When you have confirmed that the upgrade was successful, create a single bootable backup drive as described in appendix A of *Upgrading the SMW Base Operating System to SLES 11 SP3*.

5.3 Upgrade SMW Software

For a system running the SMW 7.1 release 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) release software, use the procedures in Chapter 6, Update the Cray SMW HA System on page 81.

Important: Before upgrading the SMW HA software, ensure that the operating system and SMW software has been upgraded to the required releases.

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 Prepare for an SMW Upgrade

Procedure 17. Prepare for an SMW 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. Determine whether the `postgresql` service is currently on or off, and record this state. After completing the upgrade, you will return the `postgresql` service to the same state.

   ```
   smw1:~ # chkconfig postgresql
   postgresql state
   ```

4. 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.

5. Record the iDRAC IP address of both SMWs in case you need to power-cycle either SMW.

Usually, the iDRAC host name follows the naming convention *hostname*`-drac`. For example, if the host names are `smw1` and `smw2`, the iDRAC host names would be `smw1-drac` and `smw2-drac`. Use the following `ping` commands to display the iDRAC IP addresses.

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

6. Shut down the Cray system by typing the following command as `crayadm` on the active SMW (`smw1`).

```
crayadm@smw1:~>xtbootsys -s last -a auto.xtshutdown
```

7. Stop file synchronizing.

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

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

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

Next, you will upgrade the SMW software on the active SMW.

5.3.2 Upgrade SMW Software on the Active SMW

Procedure 18. Upgrade 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 device name for the mount command.

   Note: In the following command, replace `scsi-`*xxxxxxxxx* with the persistent 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. Upgrade the SMW software 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 and all cluster services have started.

3. Change 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 upgrade 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: Check cluster status with the crm_mon -1 command.

Next, you will upgrade the SMW software on the passive SMW.

5.3.3 Upgrade SMW Software on the Passive SMW

Procedure 19. Upgrade 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 device name for the mount command.

Note: In the following command, replace scsi-*xxxxxxxxx* with the persistent 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. Upgrade 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 the 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, including the reboot of smw2.

d. Update the controller image, as described by step 1 in procedure 23, For Cray XC Systems Only: Updating the BC and CC Firmware. Skip the remaining steps in the SMW update procedure.

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

f. **Skip** section 4.9, Updating SMW Software on the Boot Root and Shared Root. This procedure was already done for the first SMW.

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

6. Update the `/etc/hosts` file on the second SMW to include information for the Cray system hardware.

   a. Create a backup copy of `/etc/hosts` on `smw2`.

   ```
   smw2:~ # cp /etc/hosts /etc/hosts.sav
   ```

   b. Copy `/etc/hosts` from `smw1`.

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

   ```
   smw2:~ # scp smw1:/etc/hosts /etc/hosts
   ```

   c. Edit `/etc/hosts` to change the final digit in the dotted quad (shown below as *N*) for the following items.

   - If *N* is 2, change it to 3.
   - If *N* is 3, change it to 2.

   ```
   10.1.1.N        smw smw-net1 dbhost
   10.2.1.N        smw-net2
   10.3.1.N        smw-net3
   10.4.1.N        smw-net4
   ```

   d. Save your changes and exit the editor.

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

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

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

```
smw2:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

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

Next, you will finish the SMW software upgrade.

5.3.4 Finish the SMW Upgrade

Procedure 20. Finish 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 at the end of Procedure 19 on page 69. 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
   ```

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

   ```
   smw2:~ # reboot
   ```

   Before continuing, wait until `smw2` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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).

8. Return the `postgresql` service to its pre-upgrade state (either on or off), as recorded in Procedure 17 on page 66(see step 3).

   ```
   smw1:~ # chkconfig postgresql state
   ```

5.4 Upgrade 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 82.

Procedure 21. Upgrade 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. Upgrade the SMW HA software 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 the 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. Upgrade the SMW HA software 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 the 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. Reboot the first SMW (`smw1`) and wait for the reboot to finish.

```
smw1:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

```
smw2:~ # reboot
```

Before continuing, wait until `smw2` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

7. 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. Execute the following commands to correct the `crm configure` directives in the `SMWHAconfig` script.

```
smw1:~ # cp SMWHAconfig SMWHAconfig.sav
smw1:~ # sed -i 's/crm configure verify/crm -c cib_update configure verify/' ./SMWHAconfig
```

e. On the active SMW, execute the `SMWHAconfig` command with the `--update` option.

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

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

f. When `SMWHAconfig` runs, it prompts for the virtual host name if the system is being updated from an older version of the release (such as from UP00 to UP01). Enter the virtual host name for the SMW HA cluster.

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

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

```
smw1:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

```
smw2:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

8. If you have made local changes to the list of synchronized files and directories in `/etc/csync2/csync2_cray.cfg`, restore the local changes to the upgraded 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. On `smw1`, navigate to the `/etc/csync2` directory.

b. Edit the files `csync2_cray.cfg` and `csync2_cray.cfg.sav`.

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.

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

d. Save your changes to `csync2_cray.cfg` and exit the editor for both files.

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

```
smw1:~ # clear_failcounts
```

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

```
smw1:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

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

11. 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.

12. 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 Must Restore a Previous SMW HA Configuration on page 137 for information on how to restore the previous configuration.

5.5 Upgrade CLE Software on Both SMWs

When upgrading the SMW HA system, Cray recommends upgrading the CLE software after you have upgraded the operating system, SMW, and SMW HA software. If necessary, you can upgrade the CLE software before you start upgrading the operating system, SMW, and SMW HA 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 22. Upgrade 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 active SMW, and replace *smw2* with the host name of the passive SMW. Replace *bootimagedir* with the name of the boot image directory, and replace *file* with the name of the boot image.

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

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

3. Copy the CLE install directory, `/home/crayadm/install.`*xtrel*, from the first SMW to a local directory on the second SMW (such as `/tmp`). Do not use `/home/crayadm` on the second SMW, because that would create local differences for this shared directory. Replace *xtrel* with the site-determined name specific to the release being installed.

4. 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. Add the following files to the synchronization list:

   ```
   /var/opt/cray/install/shell_bootimage*
   /var/opt/cray/install/networking_configuration-p*.json
   ```

   For this procedure, see Add Site-specific Files to the Synchronization List on page 115.

5.6 Make Optional Configuration Changes After an Upgrade

After upgrading the software on an SMW HA system, you may choose to make the following configuration change.

- Move the power management database (PMDB) to the shared boot RAID; for more information, see Move the PMDB to the Boot RAID on page 117.

Update the Cray SMW HA System [6]

Cray provides periodic updates to the SMW, CLE, and SMW HA software releases. A software update installation on an SMW HA system involves moving to the next update release for all necessary software components (the SMW, CLE, and SMW HA software). In an update release, the minor version number changes; for example, from SMW SLEHA 11 SP3 to SMW SLEHA 11 SP3 UP01.

The procedures in this chapter describe how to install these updates. The procedures provided in this chapter do not change the base operating system version running on your SMW.

Caution: You must be running the SUSE Linux Enterprise Server version 11 Service Pack 3 (SLES 11 SP3) SMW base operating system and a release of SMW 7.2 or later on your SMW in order to perform the procedures in this chapter.

Important: You must update the SMW software and SMW HA software together. Cray recommends updating in the following order:

1. SMW software
2. SMW HA software

You can update the CLE software either before or after the SMW and SMW HA updates.

For each update release package, update the active SMW first, then update the passive SMW. Do not update both SMWs at the same time.

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.

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

6.1 Before You Start

- Read the *SMW HA Release Notes* and the *SMW HA README* provided with the SMW HA release package to confirm the required versions for the operating system, SMW, and SMW HA software release, as well as the supported update 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.
- Identify any 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 installing the update packages, back up the 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.
- Cray recommends checking all file systems with `fsck` before beginning an update, because updating an SMW HA system requires several reboots.

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

6.2 Update SMW Software

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, you must upgrade the software using the procedures in Chapter 5, Upgrade the Cray SMW HA System on page 63.

Important: Before updating the SMW HA software, ensure that the operating system and SMW software has been updated to the required releases.

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).

6.2.1 Prepare for an SMW Update

Procedure 23. Prepare for an SMW 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. Determine whether the `postgresql` service is currently on or off, and record this state. After completing the update, you will return the `postgresql` service to the same state.

   ```
   smw1:~ # chkconfig postgresql
   postgresql state
   ```
4. 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.
5. Record the iDRAC IP address of both SMWs in case you need to power-cycle either SMW.

   Usually, the iDRAC host name follows the naming convention *hostname*`-drac`. For example, if the host names are `smw1` and `smw2`, the iDRAC host names would be `smw1-drac` and `smw2-drac`. Use the following `ping` commands to display the iDRAC IP addresses.

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

6. Shut down the Cray system by typing the following command as `crayadm` on the active SMW (`smw1`).

```
crayadm@smw1:~>xtbootsys -s last -a auto.xtshutdown
```

7. Stop file synchronizing.

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

8. 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.

Next, you will update the SMW software on the active SMW.

6.2.2 Update SMW Software on the Active SMW

Procedure 24. Update 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. Update the SMW software 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 30
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).

Next, you will update the SMW software on the passive SMW.

6.2.3 Update SMW Software on the Passive SMW

Procedure 25. Update 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. Update 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 the 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, including the reboot of smw2.

   d. Update the controller image, as described in procedure 23, For Cray XC Systems Only: Updating the BC and CC Firmware (see step 1). Skip the remaining steps in the SMW update procedure.

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

   f. **Skip** section 4.9, Updating SMW Software on the Boot Root and Shared Root. This procedure was already done for the first SMW.

7. Update the /etc/hosts file on the second SMW to include information for the Cray system hardware.

a. Create a backup copy of `/etc/hosts` on `smw2`.

```
smw2:~ # cp /etc/hosts /etc/hosts.sav
```

b. Copy `/etc/hosts` from `smw1`.

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

```
smw2:~ # scp smw1:/etc/hosts /etc/hosts
```

c. Edit `/etc/hosts` to change the final digit in the dotted quad (shown below as *N*) for the following items.

- If *N* is 2, change it to 3.
- If *N* is 3, change it to 2.

```
10.1.1.N        smw smw-net1 dbhost
10.2.1.N        smw-net2
10.3.1.N        smw-net3
10.4.1.N        smw-net4
```

d. Save your changes and exit the editor.

Next, you will finish the SMW update.

6.2.4 Finish the SMW Update

Procedure 26. Finish 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
   ```

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

   ```
   smw2:~ # reboot
   ```

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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).

8. Return the `postgresql` service to its pre-update state (either on or off), as recorded in Procedure 23 on page 83 (see step 3).

   ```
   smw1:~ # chkconfig postgresql state
   ```

6.3 Update 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 82.

Procedure 27. Update 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. Update the SMW HA software 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 the 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. Update the SMW HA software 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 the 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. Reboot the first SMW (`smw1`) and wait for the reboot to finish.

```
smw1:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

```
smw2:~ # reboot
```

Before continuing, wait until `smw2` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

7. 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. Execute the following commands to correct the `crm configure` directives in the `SMWHAconfig` script.

```
smw1:~ # cp SMWHAconfig SMWHAconfig.sav
smw1:~ # sed -i 's/crm configure verify/crm -c cib_update configure verify/' ./SMWHAconfig
```

e. On the active SMW, execute the `SMWHAconfig` command with the `--update` option.

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

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

f. When `SMWHAconfig` runs, it prompts for the virtual host name if the system is being updated from an older version of the release (such as from UP00 to UP01). Enter the virtual host name for the SMW HA cluster.

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

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

```
smw1:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

```
smw2:~ # reboot
```

Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

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

8. If you have made local changes to the list of synchronized files and directories in `/etc/csync2/csync2_cray.cfg`, restore the 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 files `csync2_cray.cfg` and `csync2_cray.cfg.sav`.

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.

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

d. Save your changes to `csync2_cray.cfg` and exit the editor for both files.

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

   ```
   smw1:~ # clear_failcounts
   ```

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

```
smw1:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

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

11. 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.

12. 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 Must Restore a Previous SMW HA Configuration on page 137 for information on how to restore the previous configuration.

6.4 Update CLE Software on Both SMWs

When updating the SMW HA system, Cray recommends updating the CLE software after you have upgraded the SMW and SMW HA software. If necessary, you can update the CLE software before you start updating the SMW and SMW HA 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 28. Update 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 9 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 active SMW, and replace *smw2* with the host name of the passive SMW. Replace *bootimagedir* with the name of the 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. Copy the CLE install directory, `/home/crayadm/install.`*xtrel*, from the first SMW to a local directory on the second SMW (such as `/tmp`). Do not use `/home/crayadm` on the second SMW, because that would create local differences for this shared directory. Replace *xtrel* with the site-determined name specific to the release being installed.

4. Update the CLE release software on the passive SMW (`smw2`) by following the procedures in section 9.2 of *Installing and Configuring Cray Linux Environment (CLE) Software* (S–2444).

   Important: Skip the other procedures in chapter 9 of *Installing and Configuring Cray Linux Environment (CLE) Software*.

5. Add the following files to the synchronization list:

   ```
   /var/opt/cray/install/shell_bootimage*
   /var/opt/cray/install/networking_configuration-p*.json
   ```

   For this procedure, see Add Site-specific Files to the Synchronization List on page 115.

6.5 Make Optional Configuration Changes After an Update

After updating the software on an SMW HA system, you may choose to make the following configuration change:

- Move the power management database (PMDB) to the shared boot RAID; for more information, see Move the PMDB to the Boot RAID on page 117.

Part II: Administration Guide

Manage the Cray SMW HA System [7]

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:

- Boot an SMW HA Cluster on page 99
- Log in to the active SMW; see Log In to the SMW HA Cluster on page 99
- Identify the Active SMW on page 100
- Monitor the SMW HA Cluster on page 106
- Change Passwords on an SMW HA System on page 110
- Check Shared File Systems Manually with `fsck` on page 111
- Customize the SMW HA Cluster on page 112
- Handle Failover on page 122

This chapter also describes the SMW HA commands; see Understand SMW HA Commands on page 100.

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 virtual (active) SMW, which could be either `smw1` or `smw2`. This virtual host name was defined during initial installation.

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

7.1 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`.

- Key system services (also called *resources*) are controlled by the cluster manager (see Cluster Resources on page 16). Do not start or stop these services individually. Instead, use cluster management tools to start and stop these services. For more information, see Understand SMW HA Commands on page 100.

- 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 the CLE boot root, the shared root, or any other CLE file systems from the boot RAID on both SMWs at the same time.

- An SMW HA system disables the automatic `fsck` for shared file systems at system start time, because the checks could delay failover by several minutes or hours. Cray recommends manually checking each shared file system on a regular basis, such as during periodic maintenance. For more information, see Check Shared File Systems Manually with `fsck` on page 111.

For additional differences, see Limitations of SMW Failover on page 18.

7.2 Boot 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 capability. Before starting the second SMW, wait until the first SMW has completed startup and initialized all cluster resources.

Procedure 29. Boot 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).

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all HA resources have started.

   Tip: You can check the status of the SMW HA services with the `crm_mon -1` command. For more information, see Display Cluster Status on page 106.

2. Boot `smw2` (or the SMW that you want to be passive).

   Before continuing, wait until `smw2` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw2`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

3. Verify that both SMWs are online and that all cluster resources have started. For more information, see Display Cluster Status on page 106.

7.3 Log In to the SMW HA Cluster

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 30. Log into the SMW HA Cluster

- 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 the 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`).

7.4 Identify the Active SMW

Procedure 31. Identify 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 30 on page 99.
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
```

7.5 Understand 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
```

7.5.1 `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
 ClusterMonitor (ocf::smw:ClusterMonitor): 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
```

To display the same information with the interactive interface:

```
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. 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*.

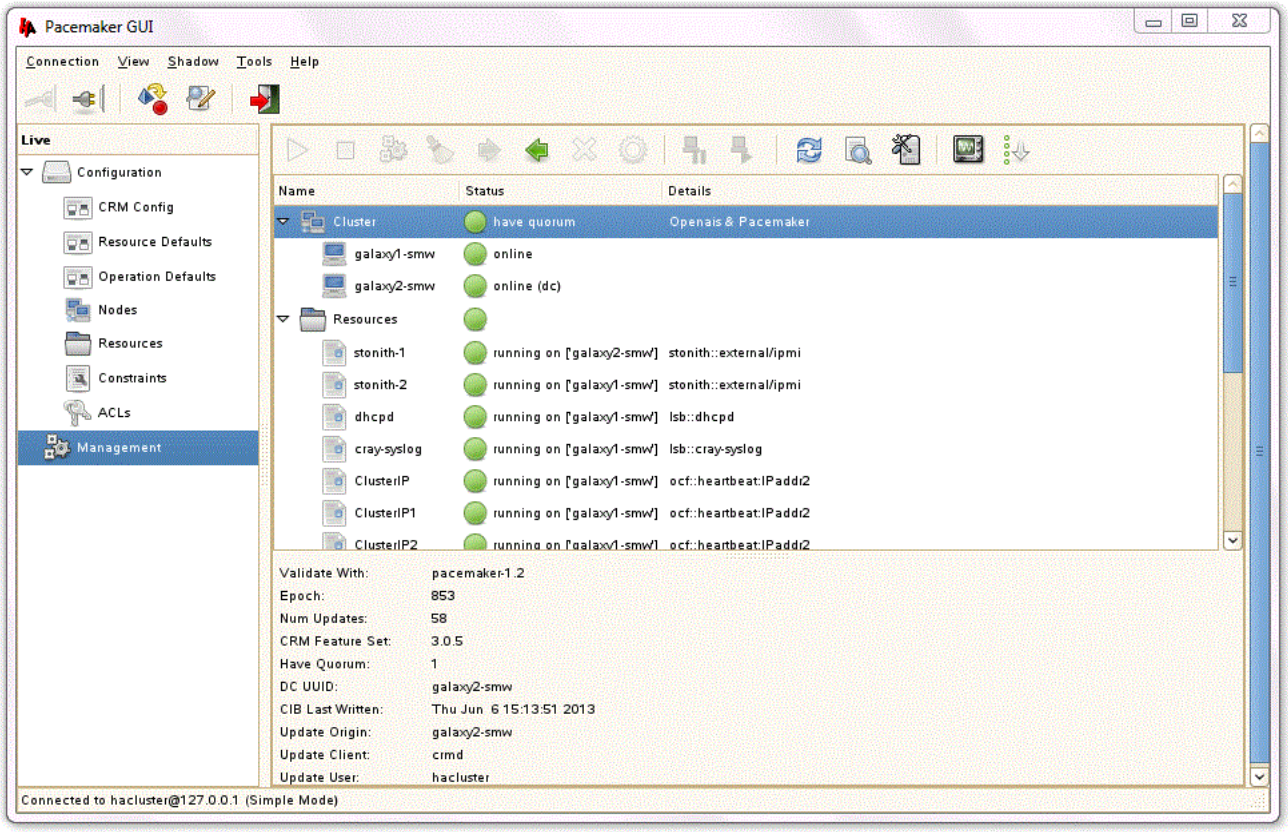
7.5.2 `crm_gui`

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. Use the same password as `root` on the SMW (see Passwords For an SMW HA System on page 25).

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*.

7.5.3 crm_mon

The SUSE crm_mon command helps monitor cluster status and configuration. The output includes the number of nodes, host names, SMW status, the resources configured in the cluster, the current status of each resource, and any failed actions.

By default (if no options are specified), crm_mon runs continuously, redisplaying the cluster status every 15 seconds. To specify the number of repeats, enter a number as an option. This example displays one snapshot of cluster status.

```
smw1:~ # crm_mon -1
Last updated: Sun Oct 26 23:54:38 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

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

Tip: Use the -r option to include inactive resources.

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*.

7.5.4 crm_resource

The SUSE crm_resource command displays resource information for the cluster (see Cluster Resources on page 16). You can use the -l (lower-case L) option to list the name of each instantiated cluster resource. For example, enter the following command as root on either SMW.

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

Note: crm_resource may display different resource names or resource order on your system.

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

7.5.5 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 and failed action data for all SMW failover resources. This command resets the resource failcounts (number of failures) and list of failed actions on both SMWs in an SMW HA cluster.

  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, 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.

7.6 Monitor the SMW HA Cluster

7.6.1 Display Cluster Status

Procedure 32. Display cluster and resource status

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: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): 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 the 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 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 Resources Are Stopped on page 129.

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-ClusterMonitor 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 Resources Are Stopped on page 129.

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.

7.6.2 Display SMW Power Status

Procedure 33. Display 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 Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 ClusterIP      (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP1     (ocf::heartbeat:IPaddr2):       Started smw1
 ClusterIP2     (ocf::heartbeat:IPaddr2):       Started smw1
   .
   .
   .
```

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

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

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

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

   At the password prompt, enter the `root` password for the iDRAC.

For the procedure to restore power and join the SMW to the cluster, see An SMW Is Powered Off on page 131.

7.7 Change Passwords on an SMW HA System

On an SMW HA system, the `root` password **must** be the same on both SMWs. In addition, the `hacluster`, `stonith`, and Integrated Dell™ Remote Access Controller (iDRAC) passwords **must** be the same as the SMW `root` password. For more information, see Passwords For an SMW HA System on page 25.

When you change the SMW `root` password, you must also change the passwords for the iDRAC, `hacluster`, and the `stonith` resources to use the same password as `root` on both SMWs.

Procedure 34. Change SMW, iDRAC, 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 on the active SMW, execute the following commands on `smw1`:

```
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 SMW `root` and `hacluster` passwords on the passive SMW, execute the following commands on `smw2`:

```
smw1:~# passwd root
smw1:~# passwd hacluster
```

   Important: Use the same `root` password as on `smw1`. The `hacluster` password **must** be the same as the `root` password.

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

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

7.8 Check Shared File Systems Manually with `fsck`

Administrators should schedule time to run `fsck` to check each shared file systems on the boot RAID.

An SMW HA system does not run `fsck` for the shared file systems at system start time, because the checks could delay failover by several minutes or hours. (Note that even if a file system's `fsck` parameter is set to `auto`, the HA resource agent does not run `fsck`.) These shared file systems are `ext3` journal file systems, so inconsistencies after a power outage, STONITH failover, or forced unmount will usually be detected when the file system is mounted at system start time. However, Cray recommends checking each file system on a regular basis, such as during periodic maintenance.

Use this procedure to check the maximum mount count and last-check date for the shared file systems.

Procedure 35. Check the `fsck` status of shared file systems

1. Determine the `/dev/sd` name for each shared file system.

```
smw1:~ # df
     Filesystem     1K-blocks      Used Available Use% Mounted on
     /dev/sda2      120811676  82225412  32449332  72% /
     udev            16433608       756  16432852   1% /dev
     tmpfs           16433608     37560  16396048   1% /dev/shm
     /dev/sdo       483807768 197536596 261695172  44% /var/opt/cray/disk/1
     /dev/sdp       100791728  66682228  28989500  70% /home
     /dev/sdq       100791728    484632  95187096   1% /var/lib/mysql
     /dev/sdr        30237648    692540  28009108   3% /var/lib/pgsql
```

2. Check the mount count and last-checked date of each shared file system, as in this example.

```
smw1:~ # tune2fs -l /dev/sdo | egrep -i "mount|check"
Last mounted on:          <not available>
Default mount options:    (none)
Last mount time:          Wed Feb 26 16:33:10 2014
Mount count:              352
Maximum mount count:      38
Last checked:             Sun Jul 21 15:57:55 2013
Check interval:           15552000 (6 months)
Next check after:         Fri Jan 17 14:57:55 2014
```

3. If the mount count exceeds the maximum mount count, schedule time to manually check the file system with `fsck`.

7.9 Customize 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 104.
- Do not attempt to migrate a single resource. All resources must migrate as a group. For more information, see Cluster Resources on page 16.
- 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 Manage Synchronized Files on page 113.

7.9.1 Configure 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 36. Change 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@thedomain.com
```

2. Verify the setting.

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

7.9.2 Manage 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 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).

7.9.2.1 Monitor the `fsync` Resource

The `fsync` resource controls file synchronized operations. Every 100 seconds, `fsync` checks for files that need to be synchronized. If `fsync` stops, no file synchronization occurs.

Procedure 37. Monitor the `fsync` resource

1. Check the status of the `fsync` resource by executing the following command as `root` on either SMW:

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

2. If `fsync` is stopped, display the failcount data for this resource. The status `Stopped` is usually caused by exceeding the failcount for a resource.

```
smw1:~ # show_failcounts | grep fsync
node=smw1 scope=status  name=fail-count-fsync value=13
node=smw2 scope=status  name=fail-count-fsync value=0
```

3. If necessary, clear the failcount data for the `fsync` resource.

```
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=ClusterMonitor
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
 ClusterMonitor (ocf::smw:ClusterMonitor):      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
Cleaning resource on node=smw2 for resource=ClusterMonitor
```

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

7.9.2.2 Add Site-specific Files to the Synchronization List

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.

Important: The `csync2` utility is designed to synchronize small amounts of data. If `csync2` must monitor many directories or synchronize a large amount of data, it can become overloaded and failures may not be readily apparent. Cray recommends that you add only small files to `/etc/csync2/csync2_cray.cfg`. 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`, if SEDC does not use the PMDB
- `/etc/hosts`
- Very large files

Tip: You can use `scp` to copy a large, static file to the passive SMW, as in this example:

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

For directories and files that may change during the copy operation, you can use the `rsync` command.

Procedure 38. Add site-specific files to the synchronization list

1. For each file or directory on the active SMW that you want to synchronize,

ensure that the parent directory exists 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.

2. Edit the file `/etc/csync2/csync2_cray.cfg` as `root` on the active SMW.
3. To add a 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: For a symbolic link, only the link itself is synchronized, not the content (destination) of the symbolic link.
4. Save your changes and exit the editor.

   The `fsync` resource will synchronize the additional files and directories the next time it runs.

7.9.3 Set 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 39. Set 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.

7.9.4 Move the PMDB to the Boot RAID

If you have created a LUN on the boot RAID for the power management database (PMDB), move the PMDB to the boot RAID after you have configured the SMW HA system.

Important: Check the size of `/var/lib/pgsql` and make sure that RAID disk has enough space to hold the data. Use the following command to display the size of the local data:

```
smw1:~ # du -hs /var/lib/pgsql
```

Procedure 40. Move the PMDB to the shared boot RAID

Use these steps to configure the RAID disk and transfer the power management data to the power management disk on the shared boot RAID.

Important: During this procedure, do not use a typescript session running directly on the SMW. To save the output of this procedure, use the `script` command to start the typescript session on your local workstation **before** logging into the SMW, as in this example:

```
workstation> script -af my_output_file
Script started, file is my_output_file
workstation> ssh crayadm@smw1
```

1. Shut down the Cray system by typing the following command as `crayadm`.

```
crayadm@smw1:~>xtbootsys -s last -a auto.xtshutdown
```

2. Log into the active SMW as `root`, either at the console or by using the actual (not virtual) host name.

   Important: You **must** log in directly as `root`. Do not use `su` from a different SMW account such as `crayadm`.

3. Change to the directory containing the SMWHAconfig command.

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

4. Use the SMWHAconfig command to move the PMDB and configure the required HA resources.

   Note: Replace scsi-*xxxxxxxxx* with the persistent device name for the PMDB directory on the boot RAID.

```
smw1:~ # ./SMWHAconfig --add_disk=pm-fs \
--device=/dev/disk/by-id/scsi-xxxxxxxxx --directory=/var/lib/pgsql
```

   This command mounts the PMDB directory (/var/lib/pgsql) to the boot RAID, copies the PMDB data, and configures the HA resources pm-fs and postgresqld.

5. Reboot the active SMW (smw1).

```
smw1:~ # reboot
```

   Before continuing, wait until smw1 has rejoined the cluster. After the SMW responds to a ping command, log into smw1, sleep for at least 2 minutes, then execute the crm_mon -1 command to verify that smw2 is online and all HA resources have started.

6. Reboot the passive SMW (smw2).

```
smw2:~ # reboot
```

   Before continuing, wait until smw2 has rejoined the cluster. After the SMW responds to a ping command, log into smw2, sleep for at least 2 minutes, then execute the crm_mon -1 command to verify that smw2 is online and all HA resources have started.

7. Verify that the power management database is on the boot RAID and that the required resources are running.

   a. Examine the log file /opt/cray/ha-smw/default/hainst/SMWHAconfig.out to verify that the power management database disk appears in the Cluster RAID Disks section (at the end of the file), as in this example.

```
---------------- Cluster RAID Disks -----------------
07-07 20:47 INFO     MYSQL Database disk = /dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf
07-07 20:47 INFO     Log disk            = /dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9
07-07 20:47 INFO     /home disk          = /dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01
07-07 20:47 INFO     PM database disk    = /dev/disk/by-id/scsi-360080e500023bff6000006b5515d9d01
07-07 20:47 INFO     ***************** Ending of HA software add_disk *****************
```

b. Ensure that the power management file system is mounted by checking for `/var/lib/pgsql` in the output of the `df` command.

```
smw1:~ # df
     Filesystem     1K-blocks      Used Available Use% Mounted on
     /dev/sda2      120811676  82225412  32449332  72% /
     udev            16433608       756  16432852   1% /dev
     tmpfs           16433608     37560  16396048   1% /dev/shm
     /dev/sdo       483807768 197536596 261695172  44% /var/opt/cray/disk/1
     /dev/sdp       100791728  66682228  28989500  70% /home
     /dev/sdq       100791728    484632  95187096   1% /var/lib/mysql
     /dev/sdr        30237648    692540  28009108   3% /var/lib/pgsql
```

c. Check the output of `crm_mon` to ensure that the `pm-fs` and `postgresqld` resources are running.

```
smw1:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

7.9.5 Move the PMDB Off the Shared Boot RAID

If the power management database (PMDB) is stored on the boot RAID, you can use `SMWHAconfig` to move the PMDB directory, `/var/lib/pgsql` to local disk on both SMWs.

Important: If the PMDB is not stored on the boot RAID, the PMDB is split across both SMWs. Data collected before an SMW failover will be lost or not easily accessible after failover.

Procedure 41. Move the PMDB off the shared boot RAID

Use these steps to move the PMDB directory, `/var/lib/pgsql`, from the shared boot RAID to local disk on both SMWs.

Note: During this procedure, do not use a typescript session running directly on the SMW. To save the output of this procedure, use the `script` command to start the typescript session on your local workstation **before** logging into the SMW.

1. Log into the active SMW as `root`.

   Important: You **must** log in directly as `root` (via `ssh`). Do not use `su` from a different SMW account such as `crayadm`.

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

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

3. Use the `SMWHAconfig` command to remove the PM database disk from the boot RAID.

```
smw1:~ # ./SMWHAconfig --remove_disk=pm-fs
```

   This command moves the PMDB directory, `/var/lib/pgsql`, to the original location on the active SMW (smw1) and removes the power management resources (`pm-fs` and `postgresqld`) from the HA configuration.

   Important: This command does **not** copy the PMDB data from the boot RAID.

4. Reboot the active SMW and wait for it to boot completely.

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

5. Reboot the passive SMW and wait for it to boot completely.

   Before continuing, wait until `smw1` has rejoined the cluster. After the SMW responds to a `ping` command, log into `smw1`, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that `smw2` is online and all resources have started.

6. Verify that the power management database is not on the shared boot RAID and that the PMDB resources are not running.

   a. Examine the log file

`/opt/cray/ha-smw/default/hainst/SMWHAconfig.out` to verify that the power management database disk does **not** appear in the `Cluster RAID Disks` section (at the end of the file), as in this example.

```
---------------- Cluster RAID Disks -----------------
07-07 20:47 INFO     MYSQL Database disk = /dev/disk/by-id/scsi-360080e500023bff6000006b3515d9bdf
07-07 20:47 INFO     Log disk            = /dev/disk/by-id/scsi-360080e500023bff6000006b1515d9bc9
07-07 20:47 INFO     /home disk          = /dev/disk/by-id/scsi-360080e500023bff6000006b5515d9c01
07-07 20:47 INFO     ***************** Ending of HA software add_disk *****************
```

b. Check the output of `crm_mon` to ensure that the `pm-fs` and `postgresqld` resources are **not** running.

```
smw1:~ # crm_mon -1
Last updated: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

7. If necessary, manually copy the PMDB data from the shared boot RAID to the original location of the PMDB for each SMW.

8. If the original location for the PMDB was not on local disk, you can use the `xtmvpmdb` command to move the PMDB to another location such as a dedicated disk. For more information, see the `xtmvpmdb`(8) man page and *Monitoring and Managing Power Consumption on the Cray XC System* (S-0043).

7.10 Handle 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 capability) 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).
- Failcounts and failed actions are written to the log file `/var/log/smw.ha` on the newly active SMW.

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 18.
- 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 103.

7.10.1 Restore 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 42. Restore 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). Use the following methods to help diagnose problems:

   a. Examine the log file `/var/log/smwha.log` on the new active SMW. For more information, see Examine the SMW HA Log File on page 126.

   b. Execute the `show_failcounts` command and note any resources with non-zero failcounts.

   c. From the active SMW, examine `/var/opt/cray/log/smwmessages-`*yyyymmdd* for relevant messages.

   d. 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 44 on page 129.

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 44 on page 129. 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, and passwordless access between the boot node and SMW is not configured, the mount point on the boot node to the SMW is stale. To refresh the mount point:

   a. Log into the boot node.

b. Unmount then remount the SMW directories.

c. Restart `bnd`.

```
boot:~ # /etc/init.d/bnd restart
```

7.10.2 Manually Cause 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 43. Perform 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: Mon Oct 27 01:19:23 2014
Last change: Thu Oct 23 15:15:04 2014 by root via crm_attribute on smw2
Stack: classic openais (with plugin)
Current DC: smw1 - partition with quorum
Version: 1.1.9-2db99f1
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 Notification   (ocf::heartbeat:MailTo):        Started smw1
 dhcpd  (lsb:dhcpd):    Started smw1
 fsync  (ocf::smw:fsync):       Started smw1
 hss-daemons    (lsb:rsms):     Started smw1
 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 Resource Group: HSSGroup
     ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
     cray-syslog        (lsb:cray-syslog):      Started smw1
     homedir    (ocf::heartbeat:Filesystem):    Started smw1
     md-fs      (ocf::heartbeat:Filesystem):    Started smw1
     pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
     postgresqld        (lsb:postgresql):       Started smw1
     mysqld     (ocf::heartbeat:mysql): Started smw1
```

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

4. If `crm_mon` shows resource problems, see the troubleshooting tips in Resources Are Stopped on page 129.

7.10.3 Examine the SMW HA Log File

The log file `/var/log/smwha.log` contains cluster status and resource failure data that can help determine the cause of a failover. At system startup (such as after a failover), the `ClusterMonitor` resource records failcounts and failed actions in the log file, then clears this failure information from `crm` (for example, in the output of `crm_mon -1`).

Note: The log file `/var/log/smwha.log` is not shared; entries are recorded only on the active SMW.

This example shows the format of entries in the log file.

```
************************************************************
smw1 acted as active SMW at Wed Jul 23 08:45:42 CDT 2014
************************************************************
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-ClusterIP value=0
node=smw1 scope=status  name=fail-count-ClusterIP1 value=0
node=smw1 scope=status  name=fail-count-ClusterIP2 value=0
node=smw1 scope=status  name=fail-count-ClusterIP3 value=0
node=smw1 scope=status  name=fail-count-ClusterIP4 value=0
node=smw1 scope=status  name=fail-count-fsync value=0
node=smw1 scope=status  name=fail-count-hss-daemons value=24
node=smw1 scope=status  name=fail-count-Notification value=0
node=smw1 scope=status  name=fail-count-ClusterMonitor value=0
node=smw1 scope=status  name=fail-count-ml-fs value=0
node=smw1 scope=status  name=fail-count-cray-syslog value=0
node=smw1 scope=status  name=fail-count-homedir value=0
node=smw1 scope=status  name=fail-count-md-fs value=0
node=smw1 scope=status  name=fail-count-pm-fs value=0
node=smw1 scope=status  name=fail-count-postgresqld value=0
node=smw1 scope=status  name=fail-count-mysqld value=0
node=smw2 scope=status  name=fail-count-stonith-1 value=0
node=smw2 scope=status  name=fail-count-stonith-2 value=0
node=smw2 scope=status  name=fail-count-dhcpd value=0
node=smw2 scope=status  name=fail-count-ClusterIP value=0
node=smw2 scope=status  name=fail-count-ClusterIP1 value=0
node=smw2 scope=status  name=fail-count-ClusterIP2 value=0
node=smw2 scope=status  name=fail-count-ClusterIP3 value=0
node=smw2 scope=status  name=fail-count-ClusterIP4 value=0
node=smw2 scope=status  name=fail-count-fsync 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-ClusterMonitor value=0
node=smw2 scope=status  name=fail-count-ml-fs value=0
node=smw2 scope=status  name=fail-count-cray-syslog value=0
node=smw2 scope=status  name=fail-count-homedir value=0
node=smw2 scope=status  name=fail-count-md-fs value=0
node=smw2 scope=status  name=fail-count-pm-fs value=0
node=smw2 scope=status  name=fail-count-postgresqld value=0
node=smw2 scope=status  name=fail-count-mysqld value=0

smw1:~ # crm status
Last updated: Wed Jul 23 08:45:48 2014
Last change: Wed Jul 23 08:45:09 2014 by root via crm_shadow on smw1
Current DC: smw1 - partition with quorum
2 Nodes configured, 2 expected votes
19 Resources configured.


Online: [ smw1 smw2 ]

 stonith-1      (stonith:external/ipmi):        Started smw2
 stonith-2      (stonith:external/ipmi):        Started smw1
 dhcpd  (lsb:dhcpd):    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
hss-daemons    (lsb:rsms):     Started smw1
Notification   (ocf::heartbeat:MailTo):        Started smw1
Resource Group: HSSGroup
    ml-fs      (ocf::heartbeat:Filesystem):    Started smw1
    cray-syslog        (lsb:cray-syslog):      Started smw1
    homedir    (ocf::heartbeat:Filesystem):    Started smw1
    md-fs      (ocf::heartbeat:Filesystem):    Started smw1
    pm-fs      (ocf::heartbeat:Filesystem):    Started smw1
    postgresqld        (lsb:postgresql):       Started smw1
    mysqld     (ocf::heartbeat:mysql): Started smw1
```

Troubleshooting an SMW HA System [8]

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

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

8.1 Resources Are Stopped

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 44. Restart 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 101.

- 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
 ClusterMonitor (ocf::smw:ClusterMonitor):      Started smw1
 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=ClusterMonitor
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
 ClusterMonitor (ocf::smw:ClusterMonitor):      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
Cleaning resource on node=smw2 for resource=ClusterMonitor
```

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

8.2 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 45. Return 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 smw1
...
```

   Note: `crm_mon` may display different resource names, group names, or resource order on the 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-iDRAC-IP-addr* with the passive SMW's iDRAC IP address.

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

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

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

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

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

6. Verify the changed power status.

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

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

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

7. Wait for the SMW to reboot.

   Before continuing, wait until the SMW has rejoined the cluster. After the SMW responds to a `ping` command, log into the SMW, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that the active SMW is online and all resources have started.

8. Join the passive SMW to the cluster.

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

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

8.3 Cluster Manager Repeatedly Kills an SMW

If the cluster manager repeatedly kills one or both SMWs with the STONITH capability, 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 Confirm the Required Network Connections on page 23.

8.4 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 46. Clear 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
   ```

8.5 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 47. Recover 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`.

8.6 Files Are Not Correctly Synchronized

If file synchronization does not seem to be working, use the following procedure to check for basic problems and stop any extra `corosync` processes.

Procedure 48. Check file synchronization and stop extra corosync processes

1. Check the current /var/opt/cray/log/smwmessages-*timestamp* file for the following error (or 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.
   ```

   If no explanation can be found in the log file, continue with the following steps.

2. Check for the corosync process on each SMW by executing the following ps command on each SMW.

   ```
   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
   ```

   Each SMW must have one (and only one) corosync process. The remaining steps describe how to stop extra corosync processes.

3. If you see multiple corosync processes, 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
   ```

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

```
smw1:~ # ps h -C corosync

smw2:~ # ps h -C corosync
```

5. If `corosync` is still running on either SMW, use the `killall` command to kill the process manually.

```
smw1:~ # killall -9 corosync

smw2:~ # killall -9 corosync
```

6. 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
```

8.7 Must Restore a Previous SMW HA Configuration

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

Caution: If you reinstall the SMW HA software on an existing SMW HA cluster, `SMWHAconfig` automatically deletes any existing data in the shared directories on the boot RAID. When redoing an initial installation, existing data is not reused.

Procedure 49. Restore a previous SMW HA configuration

If there is a problem with the update, use this procedure to restore the previous configuration.

1. Log in as `root` on the active SMW (`smw1`).
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. Erase the cluster resources.

```
smw1:~ # crm configure erase
```

4. Locate the previous configuration 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`).

5. Restore the previous configuration. In the following command, replace *YYYY-MM-DD-hh:mm* with the actual timestamp of the file.

```
smw1:~ # crm configure load replace \
/opt/cray/ha-smw/default/hainst/_CLUSTER_CONFIG_BACKUP_YYYY-MM-DD-hh:mm 2> /dev/null
.
.
.
```

6. Take both SMWs out of standby mode (put them back online).

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

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

Disable and Re-enable 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.

Note: In this chapter, some examples are left-aligned to better fit the page. Left alignment has no special significance.

A.1 Disable 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 3).
- Virtual host name and virtual IP address of the cluster (see Table 3). The examples in this procedure use the virtual host name `virtual-smw`; substitute the actual host name for the system.
- Host names of the active and passive SMWs (see Table 3). The examples in this procedure use the host names `smw1` and `smw2`; substitute the actual host names for the system.
- IP addresses of the original (virtual) Ethernet ports (see Table 2).
- Device names of the shared directories on the boot RAID. You will need the persistent device names for the following directories:
  - `/var/lib/mysql` (MySQL database)
  - `/var/opt/cray/disk/1` (Log directory)
  - `/home` (home directories)
  - `/var/lib/pgsql` (power management database, PMDB)

  For more information, see Configure the Boot RAID for SMW HA on page 27.

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=/dev/disk/by-id/scsi-360080e500023bff6000006b5515d9d01 directory=/var/lib/pgsql
```

- 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`
  - `/var/lib/pgsql` (if the PMDB is on the shared boot RAID)

  For more information, see Configure the Boot RAID for SMW HA on page 27.

Procedure 50. Disable 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 the 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 the site uses the standard fixed IP addresses for these ports. If the 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
smw1:~ # chkconfig postgresql on
```

5. Power off the passive SMW (`smw2`). For *drac-ip-address*, specify the passive SMW's iDRAC IP address.

```
smw2:~ # /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 the 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. For each shared directory, check whether the shared RAID disk is mountable. In the following commands, replace `scsi-`*xxxxxxx* with the persistent device name for a 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:~ # umount /mnt/test
```

   If the `echo` command displays the value `1` (as the error status for the `mount` command), the shared RAID disk is not mountable. Skip to step 10.

9. If the shared RAID disk is mountable, define the mount points in `/etc/fstab` for the MySQL database and the Log directory on the boot RAID.

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

   b. 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
```

   c. Continue to step 11.

10. If the shared RAID disk is **not** mountable, define the mount points in `/etc/fstab` for the MySQL database and the Log directory on local disk.

    a. 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 ...
```

    b. 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 ...
```

11. 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-enable an SMW HA Cluster

To re-enable a disabled SMW HA cluster, undo the changes to the active SMW (`smw1`) that were made in Procedure 50 on page 140. 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 3).
- Host names of the active and passive SMWs (see Table 3). 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 2).

Procedure 51. Re-enable 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 the 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 the site uses the standard fixed IP addresses for these ports. If the 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 the 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 9 in Procedure 50 on page 140.

   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 10 in Procedure 50 on page 140.

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
   ```

   Before continuing, wait until the SMW has rejoined the cluster. After the SMW responds to a `ping` command, log into the SMW, sleep for at least 2 minutes, then execute the `crm_mon -1` command to verify that the SMW is online and all resources have started.

   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.

10. If necessary, restore the original `postgresql` service state on `smw1`.

    a. Check the original PMDB configuration on the passive SMW (`smw2`).

       ```
       2~ # chkconfig postgresql
       ```

    b. If the service state on `smw1` is different, change the state on this system to match the state on `smw2`.

The SMW HA configuration is now restored. To verify that the SMW HA cluster is running correctly, see Verify the Cluster Configuration on page 50.