



# SGI® ICE™ X Installation and Configuration Guide

007-5917-003

---

**COPYRIGHT**

© 2013–2014 SGI. All rights reserved; provided portions may be copyright in third parties, as indicated elsewhere herein. No permission is granted to copy, distribute, or create derivative works from the contents of this electronic documentation in any manner, in whole or in part, without the prior written permission of SGI.

---

The SGI Tempo systems management software stack depends on several open source packages which require attribution. They are as follows:

**c3:**

C3 version 3.1.2: Cluster Command & Control Suite Oak Ridge National Laboratory, Oak Ridge, TN, Authors: M.Brim, R.Flanery, G.A.Geist, B.Luethke, S.L.Scott (C) 2001 All Rights Reserved NOTICE Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted provided that the above copyright notice appear in all copies and that both the copyright notice and this permission notice appear in supporting documentation. Neither the Oak Ridge National Laboratory nor the Authors make any representations about the suitability of this software for any purpose. This software is provided "as is" without express or implied warranty. The C3 tools were funded by the U.S. Department of Energy.

**conserver:**

Copyright (c) 2000, conserver.com All rights reserved. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. - Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. - Neither the name of conserver.com nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

---

Copyright (c) 1998, GNAC, Inc. All rights reserved. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met: - Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. - Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. - Neither the name of GNAC, Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

---

Copyright 1992 Purdue Research Foundation, West Lafayette, Indiana 47907. All rights reserved. This software is not subject to any license of the American Telephone and Telegraph Company or the Regents of the University of California. Permission is granted to anyone to use this software for any purpose on any computer system, and to alter it and redistribute it freely, subject to the following

restrictions: 1. Neither the authors nor Purdue University are responsible for any consequences of the use of this software. 2. The origin of this software must not be misrepresented, either by explicit claim or by omission. Credit to the authors and Purdue University must appear in documentation and sources. 3. Altered versions must be plainly marked as such, and must not be misrepresented as being the original software. 4. This notice may not be removed or altered.

---

Copyright (c) 1990 The Ohio State University. All rights reserved. Redistribution and use in source and binary forms are permitted provided that: (1) source distributions retain this entire copyright notice and comment, and (2) distributions including binaries display the following acknowledgment: "This product includes software developed by The Ohio State University and its contributors" in the documentation or other materials provided with the distribution and in all advertising materials mentioning features or use of this software. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

#### **pysqlite:**

Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

---

#### LIMITED RIGHTS LEGEND

The software described in this document is "commercial computer software" provided with restricted rights (except as to included open/free source) as specified in the FAR 52.227-19 and/or the DFAR 227.7202, or successive sections. Use beyond license provisions is a violation of worldwide intellectual property laws, treaties and conventions. This document is provided with limited rights as defined in 52.227-14.

---

#### TRADEMARKS AND ATTRIBUTIONS

Altix, ICE, Performance Co-Pilot, SGI, the SGI logo, and Supportfolio are trademarks or registered trademarks of Silicon Graphics International Corp. or its subsidiaries in the United States and other countries.

Altair is a registered trademark and PBS Professional is a trademark of Altair Engineering, Inc. Intel, Itanium, Phi, and Xeon are trademarks or registered trademarks of Intel Corporation. InfiniBand is a trademark of the InfiniBand Trade Association. InfiniScale is a registered trademark of Mellanox Technologies. Linux is a registered trademark of Linus Torvalds. LSI Logic and MegaRAID are registered trademarks of the LSI Logic Corporation. Red Hat and Red Hat Enterprise Linux are registered trademarks of Red Hat, Inc., in the United States and other countries. SUSE is a registered trademark of SUSE LLC in the United States and other countries.

All other trademarks mentioned herein are the property of their respective owners.



---

## New Features

This revision adds information about the following:

- Information about the new boot loader and partition layout implemented in SGI Tempo 2.9.0.
- Information about how to use the hardware event tracker (HET) has been added to this manual.
- Information about how to configure optional service node features has been moved from this manual to *SGI ICE X Administration Guide*.
- Miscellaneous technical and editorial corrections.



---

## Record of Revision

<b>Version</b>	<b>Description</b>
001	May 2013 Original publication. This revision supports the SGI Management Center 1.7 release.
002	November 2013 Revised to support the SGI Tempo 2.8.0 release.
003	May 2014 Revised to support the SGI Tempo 2.9.0 release.





---

# Contents

<b>About This Guide</b>	<b>xv</b>
Related Publications	xv
Obtaining Publications	xvii
Conventions	xvii
Reader Comments	xvii
<b>1. SGI ICE X System Software Overview</b>	<b>1</b>
SGI ICE X System Component Overview	1
SGI ICE X Networks	3
<b>2. Customizing a Factory-installed SGI ICE X System</b>	<b>7</b>
About Customizing a Factory-installed SGI ICE X System	7
Obtaining Information	8
Changing the Password and Specifying Network Information	9
Completing the Customization	11
Pushing Changes to the Compute Nodes	14
Configuring Additional Features	16
<b>3. Installing and Configuring an SGI ICE X System</b>	<b>17</b>
About Performing a New Installation and Configuring the Software on an SGI ICE X System	18
Planning the Boot Slots and the Disk Partitions	21
Disk Partition Layout and Boot Order Information	23
Boot Parameters	24
Partition Layout for a One-slot SGI ICE X System	27
Partition Layout for a Two-slot SGI ICE X System (Default)	27
<b>007-5917-003</b>	<b>ix</b>

Partition Layout for a Five-slot SGI ICE X System . . . . .	28
Preparing to Install Software on an SGI ICE X System . . . . .	29
(Conditional) Setting a Static IP Address for the Baseboard Management Controller (BMC) in the System Admin Node (SAC) . . . . .	31
(Optional) Configuring a Highly Available System Admin Controller (SAC) or a Highly Available Rack Leader Controller (RLC) . . . . .	33
Booting the System . . . . .	33
Configuring the Operating System on the System Admin Controller (SAC) Node . . . . .	35
Configuring Red Hat Enterprise Linux (RHEL) on the SAC . . . . .	35
Configuring SUSE Linux Enterprise Server (SLES) on the SAC . . . . .	39
Running the Cluster Configuration Tool . . . . .	45
(Conditional) Configuring External Domain Name Service (DNS) Servers . . . . .	55
(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK) . . . . .	56
Verifying the Environment and Installing the DCM SDK . . . . .	57
(Conditional) Installing the SGI Management Center License Key . . . . .	61
Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration . . . . .	61
Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images	64
(Conditional) Downloading the Intel Manycore Platform Software Stack (MPSS) Software and Creating Images . . . . .	66
Downloading the MPSS Software From the Intel Corporation . . . . .	67
Creating Images for the Compute Nodes That Include MIC Devices . . . . .	68
Creating Images for the Service Nodes That Include MIC Devices . . . . .	72
Configuring the Switches . . . . .	76
Verifying the Switch Cabling . . . . .	78
Configuring Management Switches With a MAC File . . . . .	81
Configuring Management Switches Without a MAC File . . . . .	85

(Conditional) Configuring the Cooling Racks and Cooling Distribution Units (CDUs) on the MCell Network's Switch Ports . . . . .	88
Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the <code>discover</code> Command . . . . .	91
(Optional) Configuring a Backup Domain Name Service (DNS) Server . . . . .	97
Configuring the InfiniBand Subnetworks . . . . .	98
<b>4. Configuring Additional Features . . . . .</b>	<b>105</b>
Enabling Hardware Event Tracker (HET) Notifications . . . . .	105
About HET . . . . .	105
Customizing HET Notifications . . . . .	106
HET Examples . . . . .	108
CPU Frequency Scaling . . . . .	109
Enabling or Disabling CPU Frequency Scaling . . . . .	109
(Optional) Changing the Governor Setting and Configuring Turbo Mode . . . . .	111
Installing MPI on a Running SGI ICE X System . . . . .	114
Troubleshooting Configuration Changes . . . . .	116
<b>5. Troubleshooting . . . . .</b>	<b>119</b>
About Troubleshooting . . . . .	119
Using the <code>switchconfig</code> Command . . . . .	120
Repairing a Broken DCM Enablement . . . . .	124
Compute Nodes Are Taking Too Long To Boot . . . . .	125
Verify the Bonding Mode on the Rack Leader Controller (RLC) . . . . .	126
<code>cimage --push-rack</code> Pushes Too Many (or Too Few) Expansions . . . . .	129
Cannot <code>ping</code> the CMCs from the Rack Leader Controller (RLC) . . . . .	130
Restarting the <code>blademon</code> Daemon . . . . .	132
Log Files . . . . .	132
CMC <code>slot_map</code> / <code>blademon</code> Debugging Hints . . . . .	133

Resolving CMC Slot Map Ordering Issues . . . . .	134
In <code>tmpfs</code> Mode, File Has Date in the Future Warnings . . . . .	135
Ensuring Hardware Clock Has the Correct Time . . . . .	136
Troubleshooting a Rack Leader Controller (RLC) With Misconfigured Switch Information . . . . .	137
Switch Wiring Rules . . . . .	139
System Admin Controller (SAC) <code>eth2</code> Link in the Bond is Down . . . . .	140
Installing SGI Tempo Versions Older than 2.9.0 . . . . .	140
<b>Appendix A. YaST2 Navigation . . . . .</b>	<b>143</b>
<b>Appendix B. Subnetwork Information . . . . .</b>	<b>145</b>
About the SCI ICE X Subnetworks . . . . .	145
VLAN Ethernet Network Configurations . . . . .	145
Head VLAN . . . . .	147
Rack VLANs . . . . .	148
MCell Cooling VLAN . . . . .	149
SGI ICE X IP Address Ranges and VLANs for Management and Application Software . . . . .	150
SGI ICE X Subnetworks for Systems That Include MIC Devices . . . . .	153
Component Naming Conventions . . . . .	155
System Control Configuration . . . . .	157
<b>Appendix C. MCell Network IP Addresses . . . . .</b>	<b>161</b>
<b>Appendix D. Legacy Partition Layout Information . . . . .</b>	<b>167</b>
About the Legacy Partition Layout . . . . .	167
Legacy Partition Layout for a One-slot SGI ICE X System . . . . .	167
Legacy Partition Layout for a Two-slot SGI ICE X System . . . . .	168
Legacy Partition Layout for a Five-slot SGI ICE X System . . . . .	169

**Index . . . . . 171**



---

## About This Guide

This guide is a reference document for people who install and configure SGI® ICE™ X systems. It describes how to perform general system discovery, installation, configuration, and operations.

## Related Publications

The release notes for the SGI Foundation Suite and the SGI Performance Suite list SGI publications that pertain to the specific software packages in those products. The release notes reside in a text file in the `/docs` directory on the product media. For example, `SGI-MPI-1.x-readme.txt`. After installation, the release notes and other product documentation reside in the `/usr/share/doc/packages/product` directory.

All SGI publications are available on the Technical Publications Library at the following website:

<http://docs.sgi.com>

The following documentation might be useful to you:

- *SGI ICE X Administration Guide*, publication 007-5918-xxx

This manual explains how to manage an SGI ICE X system.

- *SGI Management Center (SMC) Installation and Configuration*

This manual is intended for system administrators. It describes how to install and configure the SGI Management Center.

- *SGI Management Center (SMC) System Administrator Guide*

This manual describes how you can monitor and control a cluster using the SGI Management Center.

- *Message Passing Toolkit (MPT) User's Guide*

Describes industry-standard message passing protocol optimized for SGI computers. This manual describes how to tune the run-time environment to improve the performance of an MPI message passing application on SGI computers. None of these ways involve application code changes.

- *MPInside Reference Guide*

Documents the SGI MPInside MPI profiling tool.

- SGI hardware documentation.

SGI creates hardware manuals that are specific to each product line. The hardware documentation typically includes a system architecture overview and describes the major components. It also provides the standard procedures for powering on and powering off the system, basic troubleshooting information, and important safety and regulatory specifications.

The following procedure explains how to retrieve a list of hardware manuals for your system.

**Procedure 0-1** To retrieve hardware documentation

1. Type the following URL into the address bar of your browser:

`docs.sgi.com`

2. In the search box on the Techpubs Library, narrow your search as follows:

- In the **search** field, type the model of your SGI system.

For example, type one of the following: "UV 2000", "ICE X", Rackable.

Remember to enclose hardware model names in quotation marks ( " ") if the hardware model name includes a space character.

- Check **Search only titles**.
  - Check **Show only 1 hit/book**.
  - Click **search**.
- In addition to SGI documentation, the following documentation from other sources might interest you:
    - SUSE documentation for SUSE Linux Enterprise Server 11 (SLES 11)
    - Red Hat documentation for Red Hat Linux Enterprise Server 6 (RHEL 6)
    - Intel compiler documentation
    - Intel documentation about Xeon architecture



## Obtaining Publications

You can obtain SGI documentation in the following ways:

- See the SGI Technical Publications Library at the following website:

<http://docs.sgi.com>

Various formats are available. This library contains the most recent and most comprehensive set of online books, release notes, man pages, and other information.

- You can view man pages by typing `man title` on a command line.

## Conventions

The following conventions are used throughout this document:

<b>Convention</b>	<b>Meaning</b>
<code>command</code>	This fixed-space font denotes literal items such as commands, files, routines, path names, signals, messages, and programming language structures.
<i>variable</i>	Italic typeface denotes variable entries and words or concepts being defined.
<b>user input</b>	This bold, fixed-space font denotes literal items that the user enters in interactive sessions. (Output is shown in nonbold, fixed-space font.)
[ ]	Brackets enclose optional portions of a command or directive line.
...	Ellipses indicate that a preceding element can be repeated.

## Reader Comments

If you have comments about the technical accuracy, content, or organization of this publication, contact SGI. Be sure to include the title and document number of the publication with your comments. (Online, the document number is located in the

front matter of the publication. In printed publications, the document number is located at the bottom of each page.)

You can contact SGI in either of the following ways:

- Send e-mail to the following address:

[techpubs@sgi.com](mailto:techpubs@sgi.com)

- Contact your customer service representative and ask that an incident be filed in the SGI incident tracking system:

<http://www.sgi.com/support/supportcenters.html>

SGI values your comments and will respond to them promptly.

## **SGI ICE X System Software Overview**

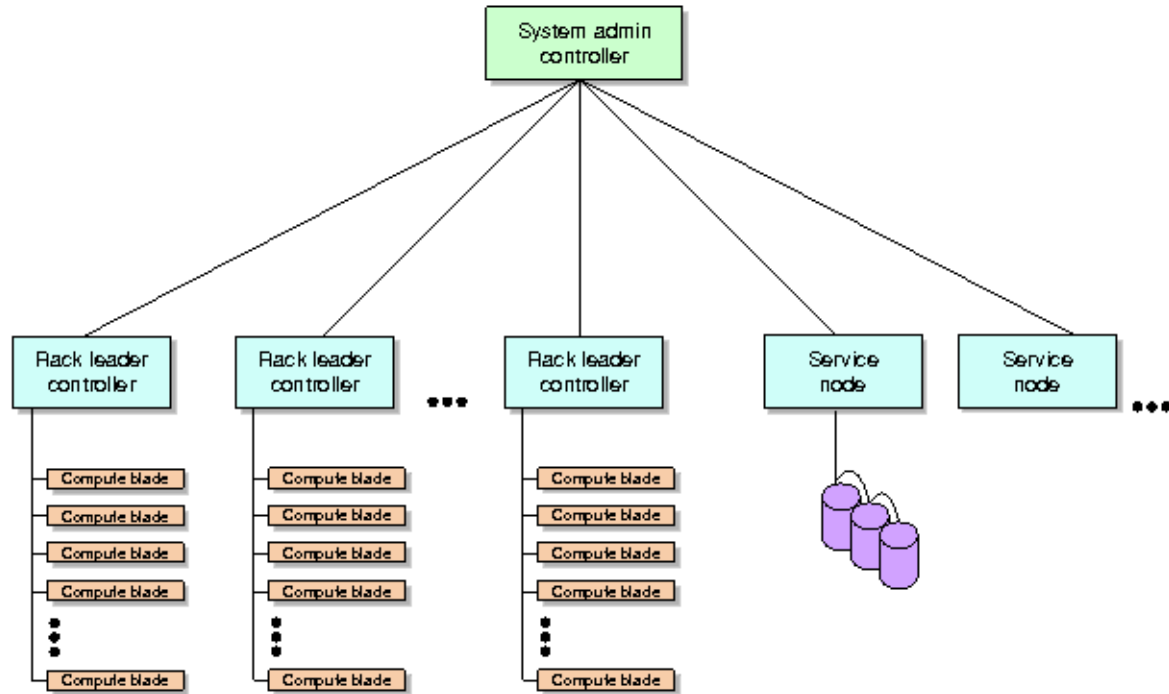
This chapter includes the following topics:

- "SGI ICE X System Component Overview" on page 1
- "SGI ICE X Networks" on page 3

### **SGI ICE X System Component Overview**

The SGI Integrated Compute Environment (ICE) X systems provide an integrated compute node (blade) environment that can include thousands of compute nodes. The SGI Tempo software for SGI ICE X systems enables you to provision, install, configure, and manage your system. This manual describes how to configure the software on the system for use at your site. This manual does not include an SGI ICE X hardware description. For hardware information, see the *SGI ICE X System Hardware User Guide*.

Figure 1-1 on page 2 shows a simple SGI ICE X system.



**Figure 1-1** Example SGI ICE X System

SGI ICE X system processes are distributed across several controllers and service nodes. Each node has a specific role. These components are as follows:

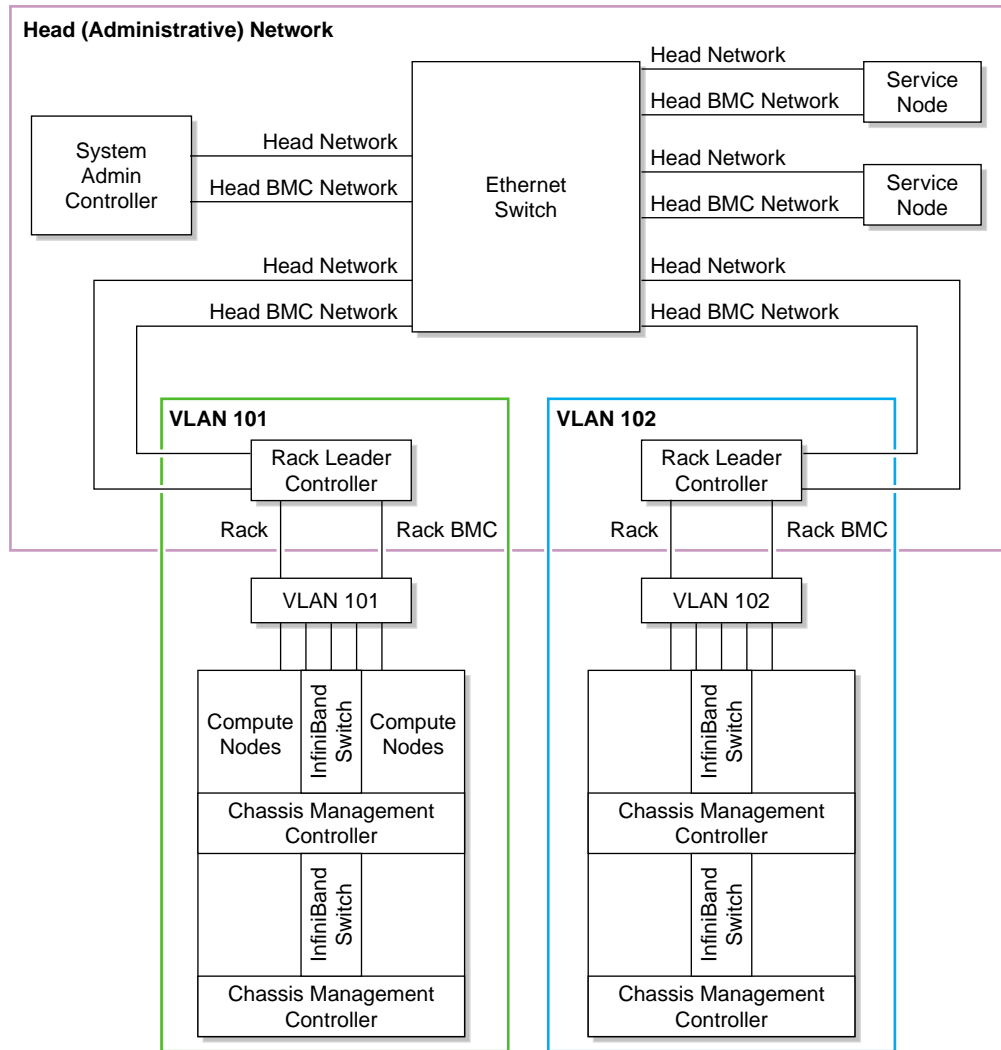
- One system admin controller (SAC). The site system administrator can log into the SAC as the root user to monitor the system, to modify the master software images, and to perform general system administration tasks.
- One rack leader controller (RLC) per compute node (blade) rack. You can have multiple compute nodes in the racks, and you can have multiple racks. Each RLC manages a set of compute nodes in a particular rack.
- One or more rack leader controllers (RLCs). The RLCs manage the compute nodes (blades) in the rack enclosures.

- One or more service nodes. End users can log into a service node to run jobs. There can be several service nodes, and they can host single services or multiple services. The service nodes can host one or more of the following types of services:
  - Login services. These services allow an end user to log in and then, for example, run or monitor MPI jobs.
  - Batch scheduling services. You can install schedulers such as Altair's PBS Professional or TORQUE.
  - I/O gateway. On a small system, you can combine the I/O gateway, login services, and batch scheduling on the same node. The I/O gateway services connect the SGI ICE X system to your house network. You can configure one or more of the following protocols on the service node: network file system (NFS), network address translation (NAT), network information service (NIS).
  - Storage. A storage service node is a network attached (NAS) appliance bundle that provides InfiniBand attached storage for the system.
  - Object storage server. This service is used in Lustre File Storage configurations.
  - Metadata server. This service is used in Lustre File Storage configurations.

The software distribution includes the master system image for the SAC. During installation and configuration process, the installation software creates the master system images for the RLCs, the service nodes, and the compute nodes. As you customize the system for your site, you can modify the node-specific system images on the SAC and push the updated images to the RLCs, to the service nodes, or to the compute nodes.

## SGI ICE X Networks

The SGI ICE X system includes several virtual local area networks (VLANs). Figure 1-2 on page 4 is a logical representation of the SGI ICE X Ethernet network that shows three VLANs in an example SGI ICE X system.



**Figure 1-2** Three VLANs

In Figure 1-2 on page 4, the head network is VLAN 1. The SAC and service nodes are attached to the customer LAN.

The system components are attached to one or more of the following two internal networks:

- The high-performance network, which is designed for computation. This InfiniBand network facilitates communication to all compute nodes from the service nodes. It connects the following:
  - The compute nodes to each other. The IB network connects all the compute nodes (blades) to each other. The compute nodes are not part of the head network.
  - The service nodes to the compute nodes.
- The head network, which is the administrative network. This Ethernet network is designed for communication between the SAC, RLCs, and service nodes. These components communicate to each other directly within the head network. The head network includes the following Ethernet connections:
  - The SAC to the Ethernet switches.
  - The Ethernet switches to the SAC, RLCs, and service nodes.
  - The service nodes to the Ethernet switches.

One or two separate InfiniBand networks segregate traffic within the SGI ICE X system in a way that optimizes computing performance. When there are two InfiniBand networks, communication is segregated by network, as follows:

- `ib0`, which is typically used for Message Passing Interface (MPI) communication.
- `ib1`, which is typically used for storage traffic.

Users can log into the SAC and into the service nodes directly. If you need to reach the RLC, you need to log into the SAC first, and then `ssh(1)` into the RLC.

The RLCs and compute nodes are attached to a separate Ethernet network. An Ethernet network connects each RLC to each of the compute nodes that the RLC controls. Communication to the compute nodes always goes through the RLC. Only the RLC for a particular rack can communicate with the compute nodes on its own rack. Each compute node has its own Ethernet IP address within its rack. If you have multiple racks, compute node IP addresses are duplicated across the racks.





## Customizing a Factory-installed SGI ICE X System

This chapter contains the following topics:

- "About Customizing a Factory-installed SGI ICE X System" on page 7
- "Obtaining Information" on page 8
- "Changing the Password and Specifying Network Information" on page 9
- "Completing the Customization" on page 11
- "Pushing Changes to the Compute Nodes" on page 14
- "Configuring Additional Features" on page 16

### About Customizing a Factory-installed SGI ICE X System

Your SGI ICE X system was tested and configured at the factory. At the factory, SGI configured the following:

- A factory-specified root password. One of the first steps in the configuration procedure is to change this root password on the system admin controller (SAC) node.
- Two slots. SGI configured the operating system that you ordered on slot 1. The operating system can be either Red Hat Enterprise Linux (RHEL) or SUSE Linux Enterprise Server (SLES). Slot 2 is blank.

The SGI ICE X system supports a maximum of five slots. If you need more than two slots, you need to reconfigure the system. During the reconfiguration, you reinstall the operating system and perform many other tasks. For the reconfiguration procedure, see Chapter 3, "Installing and Configuring an SGI ICE X System" on page 17.

- A console configured with a serial port.

If you want to retain the factory-installed configuration, complete the following procedures:

- "Obtaining Information" on page 8

- "Changing the Password and Specifying Network Information" on page 9
- "Completing the Customization" on page 11
- "Pushing Changes to the Compute Nodes" on page 14
- "Configuring Additional Features" on page 16

## Obtaining Information

Your configuration session can proceed more quickly if you gather some information before you start. You need to update the factory-installed, system-wide root password and the time zone. In addition, you need to provide information about your site network for the SAC's `eth0` network interface card (NIC).

The following procedure explains the information that you need to gather.

**Procedure 2-1** To obtain information for the customization

1. Complete the following table:

<b>Information Needed</b>	<b>Specifics for this SGI ICE X System</b>
Factory-installed password	_____
Password for this system at your site	_____
Time zone	_____
IP address	_____
Netmask	_____
Hostname	_____
Default route/Gateway	_____
Fully qualified domain name (FQDN)	_____
House NTP server	_____
First house (site) DNS resolver IP address	_____
(Optional) Second house DNS resolver IP address	_____
(Optional) Third house DNS resolver IP address	_____

House (site) domain \_\_\_\_\_

SGI ICE X system subdomain name \_\_\_\_\_

2. Proceed to the following:

"Changing the Password and Specifying Network Information" on page 9

## Changing the Password and Specifying Network Information

The following procedure explains how to change the password on the SAC and how to update the operating system configuration files with your site's networking information.

**Procedure 2-2** To customize a factory-installed RHEL operating system

1. Use the console attached to the SAC, and log into the SAC as the root user.
2. Type the following command to change the root password on the SAC and on all other nodes:

```
# cpasswd --h
```

If you do not have the current password, you can obtain the factory-installed password from your SGI representative.

For example:

```
SAC:~ # cpasswd
Enter new password:
Enter new password (again):
admin: updating /etc/shadow
rlllead: updating /etc/shadow
service0: updating /etc/shadow
SAC:~ #
```

3. Change the system time zone.

Type the following command:

```
# system-config-date
```

The `system-config-date` command starts a graphical user interface (GUI) tool. Within the GUI tool, change **only** the system time zone. The tool enables you to

change other aspects of the configuration, but for this step, change only the system time zone.

---

**Note:** Do not use this tool to change the NTP server, the time, or other configuration data.

---

4. Use a text editor, such as `vi` or `vim`, to open file `/etc/sysconfig/network-scripts/ifcfg-eth0`.
5. Edit the `/etc/sysconfig/network-scripts/ifcfg-eth0` file.

Add lines for the `IPADDR` and `NETMASK` values appropriate for your public (house) network. Also set `ONBOOT=yes`.

For example:

```
IPADDR=128.162.244.88
NETMASK=255.255.255.0
ONBOOT=yes
```

6. Save and close file `/etc/sysconfig/network-scripts/ifcfg-eth0`.
7. Use a text editor to create file `/etc/sysconfig/network`.
8. Add the following three lines to file `/etc/sysconfig/network`:

```
NETWORKING=yes
HOSTNAME=SAC_hostname
GATEWAY=gateway_IP_address
```

For `SAC_hostname`, type the hostname you want to assign to the SAC.

For `gateway_IP_address`, type the IP address of the gateway for your house network.

For example:

```
NETWORKING=yes
HOSTNAME=my-system-admin
GATEWAY=128.162.244.1
```

9. Save and close file `/etc/sysconfig/network`.
10. Use a text editor to open file `/etc/hosts`.

11. Add a line in the following format to file `/etc/hosts`:

```
SAC_IP SAC_FQDN SAC_hostname
```

The variables in the preceding line are as follows:

- For *SAC\_IP*, type the IP address of the SAC.
- For *SAC\_FQDN*, type the fully qualified domain name (FQDN) of the SAC.
- For *SAC\_hostname*, type the hostname of the SAC.

For example, add the following line:

```
128.162.244.88 acme-admin.acme.usa.com acme-admin
```

12. Save and close file `/etc/hosts`.
13. Type the following command to set the SAC hostname:

```
# hostname SAC_hostname
```

For *SAC\_hostname*, type the hostname of the SAC.

For example:

```
# hostname acme-admin
```

14. Proceed to the following:  
"Completing the Customization" on page 11

## Completing the Customization

The following topic explains how to use the cluster configuration tool to add information about your site's network to the cluster database.

**Procedure 2-3** To customize the cluster database

1. Type the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

For general information about the cluster configuration tool, see "Running the Cluster Configuration Tool" on page 45.

2. On the cluster configuration tool's main menu select **Configure the Time Client/Server (NTP)** and select **OK**.

The system guides you through the process to specify your house NTP server in file `/etc/ntp.conf`. This process differs, depending on your platform, as follows:

- On RHEL platforms, follow the the instructions that the cluster configuration tool presents to you.
- On SLES platforms, the cluster configuration tool opens a YaST2 menu. Follow the prompts in the YaST2 menu to set your NTP server.

3. On the cluster configuration tool's main menu select **Configure House DNS Resolvers** and select **OK**.

You can specify up to three house DNS resolvers.

4. Select **Quit** and select **OK** to log out from the cluster configuration tool.
5. Use the `cadmin` command, in the following format, to set the house (site) domain:

```
# cadmin --set-admin-domain site_domain
```

For *site\_domain*, specify the full name of your house domain. For example, `usa.acme.com`.

6. Use the `cadmin` command, in the following format, to change the subdomain name for the SGI ICE X system:

```
# cadmin --set-subdomain cluster_name
```

For *cluster\_name*, specify the name of the system. For example, `sleet.usa.acme.com`.

For more information about the `cadmin` command, type `cadmin -h` at the system prompt.

7. Type the following command to retrieve the path to the system images:

```
# cinstallman --show-images
```

An SGI ICE X system includes the following major nodes:

- One system admin controller. This node is also called the *SAC* or the *admin node*.

- One or more rack leader controllers. These nodes are also called the *RLCs*. There is one RLC per rack.
- Several compute nodes. These nodes are also called *blades*. The compute nodes are housed in racks, and each rack has one RLC.
- One or more service nodes. Users log into the system through the service nodes to run jobs.

The SAC hosts the master images for each of the preceding node types. If you need to change some aspect of a node's configuration, SGI recommends that you change the configuration in the master image and push out the changed master image to the affected nodes. This practice maintains consistency between the master images on the SAC and the production images on the nodes. For more information about the system images and how they install, see "About Performing a New Installation and Configuring the Software on an SGI ICE X System" on page 18.

The following example shows how to retrieve the paths to the compute node images on the SAC:

```
# cinstallman --show-images
Image Name          BT Path
compute-rhel6.5    1 /var/lib/systemimager/images/compute-rhel6.5
service-rhel6.5    0 /var/lib/systemimager/images/service-rhel6.5
lead-rhel6.5       0 /var/lib/systemimager/images/lead-rhel6.5
```

The following steps explain how to update the master node images with your site's time zone information.

8. Use the `cp(1)` command, in the following format, to set the time zone in the system images.

The format of this command is as follows:

```
cp /etc/localtime /var/lib/systemimager/images/image_name/etc
```

For *image\_name*, type the name of one of the system images you retrieved with the preceding `cinstallman` command. Type one of these commands for each master node system image.

For example, type the command three times, one for each master image, as in the following example:

```
# cp /etc/localtime /var/lib/systemimager/images/compute-rhel6.5/etc
# cp /etc/localtime /var/lib/systemimager/images/service-rhel6.5/etc
# cp /etc/localtime /var/lib/systemimager/images/lead-rhel6.5/etc
```

9. Type the following commands to update the RLC image and service node images with your site's time zone information:

```
# pdcp -g leader /etc/localtime /etc/localtime
# pdcp -g service /etc/localtime /etc/localtime
```

10. Proceed to the following:

"Pushing Changes to the Compute Nodes" on page 14

## Pushing Changes to the Compute Nodes

SGI ICE X systems can host multiple forms of system images for the rack leader controller (RLC) nodes, the service nodes, and the compute nodes. For example, you can have some production images and some test images, and you can push the images to the nodes as needed. The preceding procedure explained how to update the master images for the RLC nodes, the service nodes, and the compute nodes. In addition, the preceding procedure explained how to push the updated the RLC and service nodes images to the ICE X system components.

In this topic's procedure, you push the updated compute node images to the compute nodes. The push action enables the compute nodes to run with the updated networking information that you configured in "Completing the Customization" on page 11.

Procedures that describe system operations, later on in this manual, refer back to this procedure because you need to update system images and push out new images as part of several system administration tasks.

The following procedure explains how to push compute node system images to compute nodes.



**Procedure 2-4** To push software system images

1. Type the following command to stop the compute nodes:

```
# cpower --halt r*i*n*
```

The preceding command stops all the compute nodes. Use the preceding command when you need to push an updated image out to all the nodes.

2. (Optional) Provide information about the number of racks on your system.

Perform this step if you have a small system with fewer than eight IRUs per RLC.

The procedure pushes the updated compute image to all the compute nodes. This process can run for a long time on large systems. If you have a large number of IRUs, you need the system to perform expansions that enable you to change many compute nodes at a time. If you have fewer than eight IRUs per RLC, however, the expansions are not needed.

The following substeps explain how to prepare the system to work on a smaller number of compute nodes:

- Type the following command to retrieve the identifiers for the RLCs on your system:

```
# cnodes --leader
```

- Type the following command one or more times to suppress unnecessary processing:

```
cadmin --set-max-irus --node rlc_id number_of_racks
```

For *rlc\_id*, specify the identifier for one of the RLCs in your system.

For *number\_of\_racks*, specify the number of IRUs associated with this RLC.

For example, the following command specifies that there is only one IRU associated with the RLC identified as *r1lead*:

```
# cadmin --set-max-irus --node r1lead 1
```

3. Use the *cimage* command, in the following format, to push the changes:

```
cimage --push-rack compute-image_name rack
```

For *image\_name*, specify the name of the compute node image that you updated.

For *rack*, specify the nodes. To specify all compute nodes, specify `r\*` or `r*i*n*`. To specify only selected nodes, specify `rxixnx`, and substitute specific integer numbers for the `x` characters.

For example, the following command pushes the time zone changes (from "Completing the Customization" on page 11) to all the compute nodes:

```
# cimage --push-rack compute-rhel6.5 r\*
```

4. Type the following command to power-up the compute nodes:

```
# cpower --boot r*i*n*
```

5. Proceed to the following:

"Configuring Additional Features" on page 16

## Configuring Additional Features

The following procedure explains where you can obtain information about how to configure additional features.

**Procedure 2-5** To configure additional features

1. Configure additional features.

SGI recommends that you configure certain additional features, such as the hardware event tracker (HET). Other features, such as CPU frequency scaling might benefit your installation. For information about these additional features, see the following:

Chapter 4, "Configuring Additional Features" on page 105

2. (Optional) Configure optional features.

The SGI ICE X system supports several optional features, for example, networking features such as network address translation (NAT). For information about how to configure optional features, see the *SGI ICE X Administration Guide*.

## Installing and Configuring an SGI ICE X System

This chapter contains the following topics:

- "About Performing a New Installation and Configuring the Software on an SGI ICE X System" on page 18
- "Planning the Boot Slots and the Disk Partitions" on page 21
- "Preparing to Install Software on an SGI ICE X System" on page 29
- "(Conditional) Setting a Static IP Address for the Baseboard Management Controller (BMC) in the System Admin Node (SAC)" on page 31
- "(Optional) Configuring a Highly Available System Admin Controller (SAC) or a Highly Available Rack Leader Controller (RLC)" on page 33
- "Booting the System" on page 33
- "Configuring the Operating System on the System Admin Controller (SAC) Node" on page 35
- "Running the Cluster Configuration Tool" on page 45
- "(Conditional) Configuring External Domain Name Service (DNS) Servers" on page 55
- "(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK)" on page 56
- "(Conditional) Installing the SGI Management Center License Key" on page 61
- "Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration" on page 61
- "Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images" on page 64
- "(Conditional) Downloading the Intel Manycore Platform Software Stack (MPSS) Software and Creating Images" on page 66
- "Configuring the Switches" on page 76
- "Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the `discover` Command" on page 91

- "(Optional) Configuring a Backup Domain Name Service (DNS) Server" on page 97
- "Configuring the InfiniBand Subnetworks" on page 98

## **About Performing a New Installation and Configuring the Software on an SGI ICE X System**

SGI installs operating system software on each SGI ICE X system before factory shipment occurs. The topics in this chapter include the additional procedures that you need to complete in order to configure the system for your site.

If you want to completely reinstall the operating system and all other software, the topics in this chapter enable you to complete that task. For example, you might need to reinstall the operating system to meet site requirements or to recover a system in case of a disaster.

Figure 3-1 on page 19 depicts the software installation process.

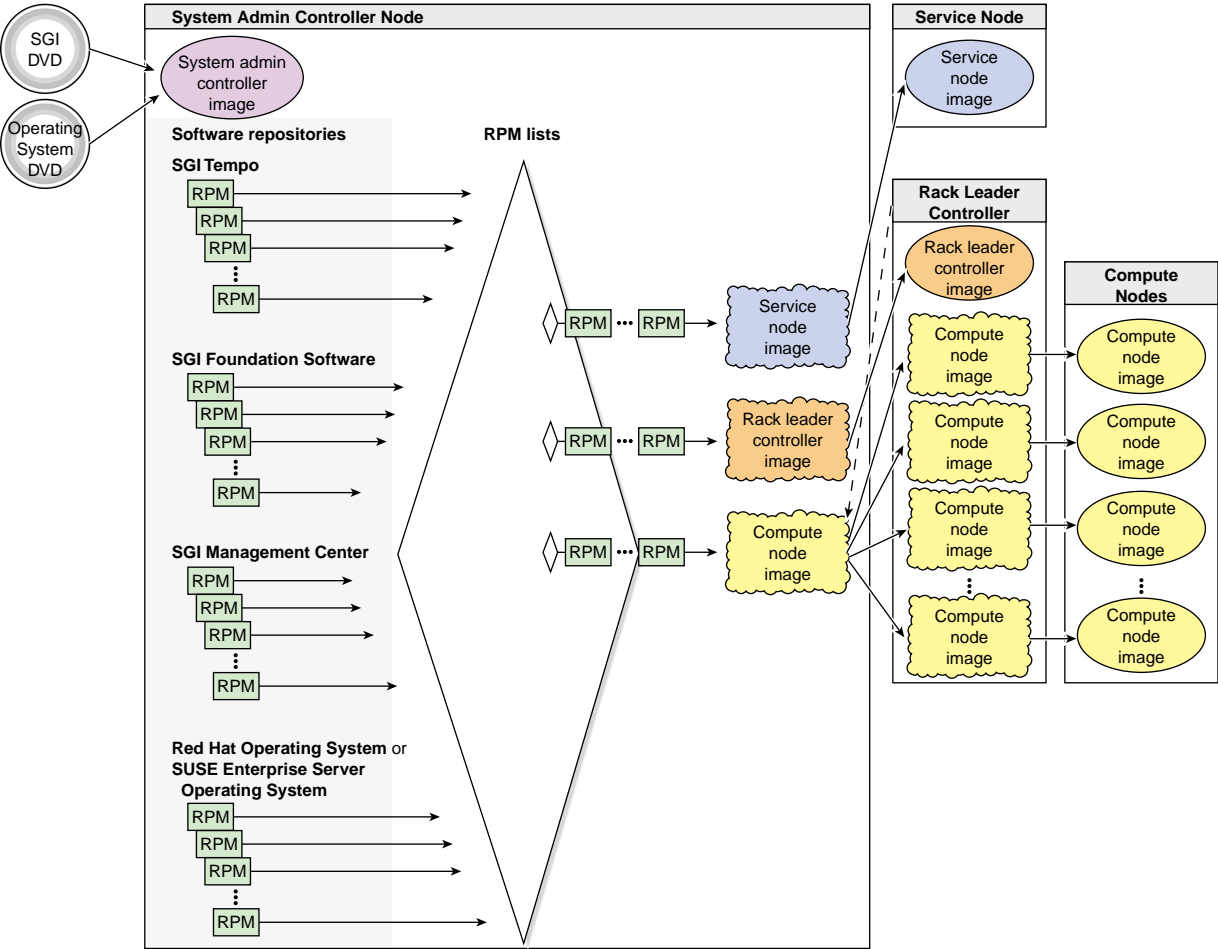


Figure 3-1 SGI ICE X Software Installation Process

Table 3-1 on page 19 shows the installation and configuration procedures to follow if you want to install the SGI ICE X system from scratch. In this case, you reinstall the operating system on the nodes and configure everything yourself.

Table 3-1 SGI ICE X System Installation and Configuration Process

Step	Task	See
1	Plan the boot slots and partitions.	"Planning the Boot Slots and the Disk Partitions" on page 21
2	Prepare to install the SGI ICE X software.	"Preparing to Install Software on an SGI ICE X System" on page 29
3	(Conditional) Configure a static address for the baseboard management controller (BMC) on the system admin controller (SAC). Perform this step only if your site practices require a static IP on the BMC.	"(Conditional) Setting a Static IP Address for the Baseboard Management Controller (BMC) in the System Admin Node (SAC)" on page 31
4	(Optional) Configure a highly available system admin controller (SAC) or a highly available rack leader controller (RLC).	"(Optional) Configuring a Highly Available System Admin Controller (SAC) or a Highly Available Rack Leader Controller (RLC)" on page 33
5	Boot the system.	"Booting the System" on page 33
6	Install the operating system on the system admin controller (SAC) node. You can install either the SUSE Linux Enterprise Server (SLES) or Red Hat Enterprise Server (RHEL) operating system.	"Configuring the Operating System on the System Admin Controller (SAC) Node" on page 35
7	Run the cluster configuration tool. Complete the initial cluster configuration tasks, which include the following: <ul style="list-style-type: none"> <li>• Set up software repositories for required and optional software.</li> <li>• Install the SAC software.</li> <li>• Configure network settings.</li> <li>• Configure the NTP server.</li> <li>• Set up the initial SAC infrastructure.</li> <li>• Configure the house network DNS resolvers.</li> </ul>	"Running the Cluster Configuration Tool" on page 45
8	(Conditional) Configure external domain name service (DNS). If you want to configure network address translation, you also need to configure an external DNS.	"(Conditional) Configuring External Domain Name Service (DNS) Servers" on page 55

Step	Task	See
9	(Optional) Start the Intel Data Center Manager (DCM) Configuration	"(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK)" on page 56
10	Sync the repository updates, apply the latest patches to the newly installed software, and clone the images.	"Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images" on page 64
11	(Conditional) Download the Intel Manycore Platform Software Stack (MPSS).	"(Conditional) Downloading the Intel Manycore Platform Software Stack (MPSS) Software and Creating Images" on page 66
12	Configure the switches.	"Configuring the Switches" on page 76
13	Use the <code>discover</code> command to install and configure the rack leader controller and service node software.	"Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the <code>discover</code> Command" on page 91
14	(Optional) Configure a backup domain name service (DNS) server on a service node.	"(Optional) Configuring a Backup Domain Name Service (DNS) Server" on page 97
15	Configure the InfiniBand subnetworks.	"Configuring the InfiniBand Subnetworks" on page 98
16	Configure optional features.	Chapter 4, "Configuring Additional Features" on page 105

## Planning the Boot Slots and the Disk Partitions

On a multiple-slot SGI ICE X system, the system admin controller (SAC), the rack leader controller (RLC), and the service nodes all have the same disk layout. If you insert an operating system installation disk, power-on the SAC, and type `install` at the `boot:` prompt, the default behavior is for the system to create two slots and to write the initial installation to slot 1. After the system is installed, you cannot increase

the number of slots without destroying the data on the disks. You can configure up to five slots, for a total of five root/boot directory pairs.

One feature of a multiple-boot system is that you can install different operating systems, or different operating system versions, into different slots. You can boot the system with the operating system of your choice. This configuration might be useful if you want to test an operating system or other software. The following are some other characteristics of single-boot systems and multiple-boot systems:

**Multiple-boot**

RLCs and service nodes boot from their own disk. Data is retained in the master boot record (MBR).

RLC and service node software is reinstalled from the SAC.

As you increase the number of slots, you decrease the amount of disk space per slot. SGI recommends a minimum of 100 GB per slot.

After an upgrade, your SAC might have a partition layout that differs from the layouts shown in this topic.

**Procedure 3-1** To plan the boot slots and disk partitions

1. Select your boot parameters and plan the disk partitions.

The following topics describe the boot parameters you can specify, disk partitioning, and operations such as cloning:

- "Disk Partition Layout and Boot Order Information" on page 23
- "Boot Parameters" on page 24
- "Partition Layout for a One-slot SGI ICE X System" on page 27
- "Partition Layout for a Two-slot SGI ICE X System (Default)" on page 27
- "Partition Layout for a Five-slot SGI ICE X System" on page 28

2. Proceed to the following:

**Single-boot**

RLCs and service nodes boot from the boot partition in the slot that is currently configured as the boot slot. Only the SAC retains data in the MBR.

Software on the RLCs and service nodes is reinstalled over the network.

A single slot uses all available disk space.



"Preparing to Install Software on an SGI ICE X System" on page 29

## Disk Partition Layout and Boot Order Information

The SGI Tempo 2.9.0 release partition layout uses the GUID partition table (GPT) and a new boot system, which is GRUB version 2. In previous Tempo versions, the partition layout was the MSDOS layout, and the boot system was GRUB version 1. These changes affect system operations in the following ways:

- If your SGI ICE X system is new, then the Tempo 2.9.0 software is the first Tempo version installed on this system. All system disks use the new partition layout. You can use the `cadmin` command, as shown in the SGI documentation, to manage all the slots.
- If you upgrade a slot from a previous Tempo release to the Tempo 2.9.0 release, note the following:
  - If you upgrade any slot in the system admin controller (SAC) to Tempo 2.9.0, the SAC converts the boot loader installed in the master boot record (MBR) to GRUB version 2. This boot loader chooses a slot when you boot the SAC.
  - If you do not upgrade all the slots to Tempo 2.9.0, the only administrative action you can perform on the slots with the older Tempo software is to boot the slot. You can boot the slots that host older versions of Tempo at the console or by selecting the slot on the GRUB version 2 boot menu. Because the SAC MBR is now upgraded to GRUB version 2, you can no longer use the `cadmin` command to manage any slots that host older versions of Tempo.
  - You can no longer install Tempo releases older than Tempo 2.9.0 on any of the system's slots. If you perform any future, from-scratch installations, these installations must be to Tempo 2.9.0 or later. You can, however, use the `updatetempo` upgrade script to upgrade slots to releases other than Tempo 2.9.0.

Disks with the new GPT layout can exist on the system alongside disks with the MSDOS partition layout. On an SGI ICE X system newly installed with Tempo 2.9.0, the software configures the GPT layout on all system disks and all disks that you clear, except for the compute nodes. After you upgrade one or more slots to Tempo 2.9.0, if you add new system disks or you clear existing system disks, the Tempo software configures the GPT layout on system disks that you add and on all disks that you clear, except for the compute nodes.

- For SAC, rack leader controller (RLC), and service nodes, when you install Tempo 2.9.0 with blank system disks, or with system disks that are cleared, Tempo 2.9.0 uses the new GPT layout. If the disks have the old MSDOS partition table format, Tempo continues to use the old MSDOS partition table format until you manually clear the disks and reinstall (if desired).

If you need to downgrade an SGI ICE X system from Tempo 2.9.0 to a previous release, see the following:

"Installing SGI Tempo Versions Older than 2.9.0" on page 140

## Boot Parameters

When you use the SGI system admin controller (SAC) installation DVD to boot an SGI ICE X system, you can specify parameters at the `boot:` prompt. By default, if you type `install` at the boot prompt and the installer detects a SAC with exactly one blank disk, the installer partitions the SAC with two slots, and the installer writes the initial installation to slot 1.

If you specify more than one boot parameter, use a space to separate each parameter.

The following list shows the result of pressing Enter at the boot prompt and shows the result of the different boot options:

Parameter	Effect
-----------	--------

(None)	
--------	--

If you do not specify any boot options, that is, if you press Enter at the `boot:` prompt without specifying any boot parameters, the installer creates two slots. It installs a root directory on only one slot, however. In this case, the software repository is not configured, so when you run the `configure-cluster` command, you need to use the GUI tool to manually enter the distribution ISO images in the `/tftpboot` directory on the SAC.

You can use the second slot for a clone or preproduction testing.

<code>console=specs</code>	
----------------------------	--

Specify console characteristics. Use when you connect to the compute node through a serial console. By default, this is set to `ttyS1,38400n8`.

The BIOS console and the IPMI console settings are configured in the SAC BIOS. If you configure the SAC BIOS to be different from the default, use this boot option to specify your console's characteristics.

`destructive=1`

Permits partitioning operations that are potentially destructive. Use this parameter only if you want to repartition a disk that contains data. Use in conjunction with the `re_partition_with_slots` parameter.

If the system encounters data in a partition, nothing destructive happens unless you also specify the `destructive=1` parameter.

`install`

Creates two slots on each compute node and installs the operating system from the DVD to slot 1. Both slots contain a root directory, and the installer writes the software distribution ISO images to each slot.

This is the recommended boot option for an initial configuration.

`install_slot=slot`

Specifies the slot into which the operating system is installed first. Specify 1 (default), 2, 3, 4, or 5 for *slot*. By default, slot 1 is installed first. To direct the initial installation to a slot other than slot 1, specify this parameter.

If you specify a *slot* that appears to have data on it, no repartitioning is performed unless you also specify `destructive=1`.

Example: You specify `install_slot=2` to direct the first installation into slot 2. In this case, after the boot completes, type the following `cadmin` command to set the default slot to slot 2:

```
# cadmin --set-default-root --slot 2
```

In this example, if you do not specify the default slot (slot 2) during the first boot of the SAC, then the SAC attempts to boot from an empty slot 1. The boot fails. If this happens, restart the boot, and select the install slot during the boot.

For information about how to manage or change the boot slot, see the *SGI ICE X Administration Guide*.

`netinst=path`

Specifies the NFS path to an ISO image for a network installation.

`re_partition_with_slots=slots`

Partitions the SAC system drive with between one and five slots. Specify 1, 2 (default), 3, 4, or 5 for *slots*.

The SAC system drive must be blank in order for this parameter to have an effect. If the SAC system drive contains data and you want to reconfigure the system, also specify the `destructive=1` parameter.

For example, the following parameter creates five slots:

```
re_partition_with_slots=5
```

The installer creates partitions on the rack leader controllers (RLCs) and service nodes to mimic the SAC. If an RLC or service node is discovered to have a slot count that does not match the SAC, the system reinitializes the partitions on the RLC and service nodes. Likewise, if you change the number of slots on the SAC, the system updates the disk partitioning on the RLC and service nodes, too.

`rescue=1`

Creates a troubleshooting environment.

`serial`

Configures the system for serial console operations for the installation and later operations. If you do not specify `serial`, the system sends output to the VGA.

Also see `vga` in this list.

`vga`

Configures the system so that you can use the VGA screen for the installation and later operations.

If you press Enter at the boot prompt, the result is the same as if you had typed `vga`.

Also see `serial` in this list.

### Partition Layout for a One-slot SGI ICE X System

Table 3-2 on page 27 shows the partition layout for a one-slot SGI ICE X system. This layout yields one boot partition. If you configure a single-slot system and later decide to add another partition, the addition process destroys all the data on your system.

**Table 3-2** Partition Layout for a Single-boot SGI ICE X System

Partition	File System Type	File System Label	Notes
1	ext3	sgidata	Contains slot information. On the SAC, contains GRUB version 2 data for choosing root slots at boot time.
2	swap	sgiswap	Swap partition.
3-10	N/A	N/A	Reserved for future use.
11	ext3	sgiboot1	Slot 1 /boot partition.
12-30	N/A	N/A	Reserved for future use.
31	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot1	Slot 1 / partition.

### Partition Layout for a Two-slot SGI ICE X System (Default)

Table 3-3 on page 28 shows the partition layout for a two-slot SGI ICE X system. This layout yields two boot partitions.

**Table 3-3** Partition Layout for a Dual-boot SGI ICE X System (Default Layout)

Partition	File System Type	File System Label	Notes
1	ext3	sgidata	Contains slot information. On the SAC, contains GRUB version 2 data for choosing root slots at boot time.
2	swap	sgiswap	Swap partition.
3-10	N/A	N/A	Reserved for future use.
11	ext3	sgiboot1	Slot 1 /boot partition.
12	ext3	sgiboot2	Slot 2 /boot partition.
13-30	N/A	N/A	Reserved for future use.
31	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot1	Slot 1 / partition.
32	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot2	Slot 2 / partition.

### Partition Layout for a Five-slot SGI ICE X System

Table 3-4 on page 28 shows the partition layout for a five-slot SGI ICE X system. This layout yields five boot partitions.

**Table 3-4** Partition Layout for a Quintuple-boot SGI ICE X System

Partition	File System Type	File System Label	Notes
1	?	sgidata	Contains slot information. On the SAC, contains GRUB version 2 for choosing root slots at boot time.
2	swap	sgiswap	Swap partition.
3-10	N/A	N/A	Reserved for future use.

Partition	File System Type	File System Label	Notes
11	ext3	sgiboot1	Slot 1 /boot partition.
12	ext3	sgiboot2	Slot 2 /boot partition.
13	ext3	sgiboot3	Slot 3 /boot partition.
14	ext3	sgiboot4	Slot 4 /boot partition.
15	ext3	sgiboot5	Slot 5 /boot partition.
16-30	N/A	N/A	Reserved for future use.
31	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot1	Slot 1 / partition.
32	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot2	Slot 2 / partition.
33	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot3	Slot 3 / partition.
34	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot4	Slot 4 / partition.
35	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot5	Slot 5 / partition.

## Preparing to Install Software on an SGI ICE X System

The following procedure explains the information you need to obtain before you begin working with the SGI ICE X system. Your installation session can proceed more quickly if you gather information before you begin.

### Procedure 3-2 To prepare for an installation

1. Contact your site's network administrator to obtain network information.

Obtain the following information to use when you configure the baseboard management controller (BMC):

- (Optional) The current IP address of the BMC on the system admin node (SAC). You can set the BMC address from a serial console if you do not have this information.
- The address you want to set for the BMC.
- The netmask you want to set for the BMC.
- The default gateway you want to set for the BMC.

Your network administrator can provide an IP address, a hostname, or a fully qualified domain name (FQDN) for each of the preceding addresses.

Obtain the following information to use when you configure the network for the SGI ICE X system:

- Hostname
- Domain name
- IP address
- Netmask
- Default route
- Root password

Obtain the following information about your site's house network:

- IP addresses of the domain name servers (DNSs)
2. Familiarize yourself with the boot parameters, and determine which boot parameters you want to use.

You can configure your SGI ICE X system to boot from one, two (default), three, four, or five partitions. This enables you to configure your SGI ICE X system as either a single-boot computer system or as a multiple-boot computer system. A multiple-boot computer system has two or more partitions, so it has more than one root directory (/) and more than one boot directory (/boot). In an SGI ICE X system, these root and boot directories are paired into multiple *slots*. A multiple-slot disk layout is also called a *cascading dual-root layout* or a *cascading dual-boot layout*.

The installation procedure explains how to create a default, two-slot SGI ICE X system and directs you to use the `install` boot parameter. If you want to create



only one slot, or if you want to create three or more slots, you need to specify different boot parameters.

The installer creates the same disk layout on all nodes. For more information about boot parameters, disk layouts, and so on, see "Planning the Boot Slots and the Disk Partitions" on page 21.

3. (Optional) Obtain the MAC file for your system from your SGI representative.

The MAC file contains MAC address information for the nodes. If you have these addresses, the node discovery process can complete more quickly.

## **(Conditional) Setting a Static IP Address for the Baseboard Management Controller (BMC) in the System Admin Node (SAC)**

Perform the procedure in this topic only if your site practices require a static IP address for the BMC.

When you set the IP address for the BMC on the SAC, you ensure access to the SAC when the site DHCP server is inaccessible. This procedure is required if you want to enable high availability. If you want to configure a highly available SAC, make sure to perform this topic's procedure on the BMCs on each of the two SACs.

The following procedures explain how to set a static IP address.

**Procedure 3-3 Method 1** — To change from the BIOS

1. Use the BIOS documentation for the SAC.

**Procedure 3-4 Method 2** — To change the IP address from the SAC

1. Log into the SAC as the root user.
2. Type the following command to retrieve the current network settings:

```
# ipmitool lan print 1
```

3. In the output from the preceding command, look for the IP Address Source line and the IP Address line.

For example:

```
IP Address Source      : DHCP Address
IP Address             : 128.162.244.59
```

Note the IP address in this step and decide whether or not this IP address is acceptable. The rest of this procedure explains how to keep this IP address or to set a different static IP address.

4. Type the following command to specify that you want the BMC to have a static IP address:

```
# ipmitool lan set 1 ipsrc static
```

This step specifies that the IP address on the BMC is a static IP address, and this step sets the IP address to the IP address that is currently assigned to the BMC. If you want to set the IP address to a different IP address, proceed to the following step. If the current IP address is acceptable, you do not need to perform the next step.

5. (Optional) Set a different IP address.

Perform this step if you want to set the static IP address to be different from the IP address that is set currently.

Type `ipmitool` commands in the following format:

```
ipmitool lan set 1 ipaddr ip_addr  
ipmitool lan set 1 netmask netmask  
ipmitool lan set 1 defgw gateway
```

The arguments are as follows:

<b>Argument</b>	<b>Specification</b>
-----------------	----------------------

<i>ip_addr</i>	The IP address you want to assign to the BMC.
<i>netmask</i>	The netmask you want to assign to the BMC.
<i>gateway</i>	The gateway you want to assign to the BMC.

For example, you can type the following commands to set the IP address to 100.100.100.100:

```
# ipmitool lan set 1 ipaddr 100.100.100.100  
# ipmitool lan set 1 netmask 255.255.255.0  
# ipmitool lan set 1 defgw 128.162.244.1
```

6. Proceed to one of the following:

- If you want to configure a highly available SAC, proceed to the following:

"(Optional) Configuring a Highly Available System Admin Controller (SAC) or a Highly Available Rack Leader Controller (RLC)" on page 33

- If you want to configure a traditional SAC, proceed to the following:  
"Booting the System" on page 33

## (Optional) Configuring a Highly Available System Admin Controller (SAC) or a Highly Available Rack Leader Controller (RLC)

SGI supports the ability to configure the SAC and rack leader controllers (RLCs) as highly available nodes in an SGI ICE X system. If you want to enable high availability (HA) on the SAC or on the RLCs, contact your SGI representative.

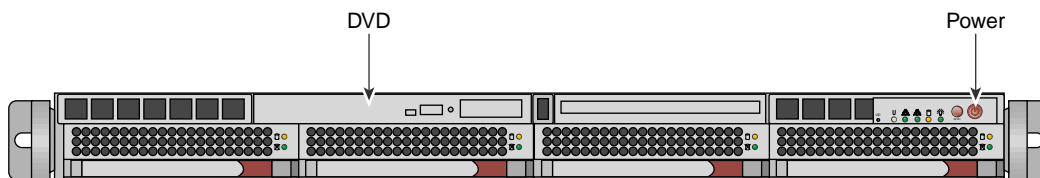
### Booting the System

The following procedure explains how to boot the system and begin the installation.

**Procedure 3-5** To boot the system

1. Power-on the SAC.

As Figure 3-2 on page 33 shows, the power-on button is on the right of the SAC.



**Figure 3-2** SAC Power On Button and DVD Drive

2. (Optional) Configure the system so that you can perform the installation from a VGA screen and can perform later operations from a serial console.

If you want to enable this capability, perform the following steps:

- Use a text editor to open file `/boot/grub/menu.lst`.

- Search the file for the word `kernel` at the beginning of a line.
- Add the following to the kernel line: `console=type`.

For example:

```
kernel /boot/vmlinuz-2.6.16.56-0.12-smp root=/dev/disk/by-label/sgiroot console=ttyS1,38400n8 splash=silent showopts
```

- Add the `console=type` parameter to the end of every kernel line. By default, this is set to `ttyS1,38400n8`. You might have `ttys2`, for example.

Later, if you want to access the SAC from only a VGA, you can remove the `console=` parameters.

3. Insert the SGI Admin Node Autoinstallation DVD into the DVD drive on the SAC.

The autoinstallation message appears, and at the end is the `boot:` prompt.

4. At the `boot:` prompt, type `install` and, optionally, other boot parameters.

"Planning the Boot Slots and the Disk Partitions" on page 21 explains the other, optional, boot parameters.

Monitor the installation. This can take several minutes.

5. Remove the operating system installation DVD.

6. At the `#` prompt, type `reboot`.

This is the first boot from the SAC's hard disk.

7. (Optional) Suppress log messages.

If you want to suppress the SAC's log message output to the screen during the boot, edit file `/etc/syscontrol.conf` and add the following line to the top of the file (line 1):

```
kernel.printk = 2 4 1 7
```

In the preceding `kernel.printk` line, the spaces between the numbers 2 4 1 7 are Tab characters.

8. Proceed to the following:

"Configuring the Operating System on the System Admin Controller (SAC) Node" on page 35

## Configuring the Operating System on the System Admin Controller (SAC) Node

The SGI ICE X platform supports both the Red Hat Enterprise Linux (RHEL) operating system and the SUSE Linux Enterprise Server (SLES) and operating system. Use one of the following procedures to install your operating system software on the SAC node:

- "Configuring Red Hat Enterprise Linux (RHEL) on the SAC" on page 35
- "Configuring SUSE Linux Enterprise Server (SLES) on the SAC" on page 39

### Configuring Red Hat Enterprise Linux (RHEL) on the SAC

This section describes how to configure Red Hat Enterprise Linux 6 on the SAC.

**Procedure 3-6** To install RHEL 6 on an SGI ICE X SAC

1. Connect to the system admin controller (SAC) by one of the following methods:
  - Through the intelligent platform management interface (IMPI) tool
  - Through the console attached to the SGI ICE X system
  - Through a separate keyboard, video display terminal, and mouse
2. Use a text editor, such as `vi` or `vim`, to open file `/etc/sysconfig/network-scripts/ifcfg-eth0`.
3. Add lines for the `IPADDR`, `NETMASK`, and `NETWORK` values appropriate for your public (house) network to file `/etc/sysconfig/network-scripts/ifcfg-eth0`.

For example:

```
IPADDR=128.162.244.88
NETMASK=255.255.255.0
NETWORK=128.162.244.0
```

4. Save and close file `/etc/sysconfig/network-scripts/ifcfg-eth0`.
5. Use a text editor to create file `/etc/sysconfig/network`.

6. Add the following three lines to file `/etc/sysconfig/network`:

```
NETWORKING=yes
HOSTNAME=SAC_hostname
GATEWAY=gateway_IP_address
```

For *SAC\_hostname*, type the hostname you want to assign to the SAC.

For *gateway\_IP\_address*, type the IP address of the gateway for your house network.

For example:

```
NETWORKING=yes
HOSTNAME=my-system-admin
GATEWAY=128.162.244.1
```

7. Save and close file `/etc/sysconfig/network`.
8. Use a text editor to open file `/etc/hosts`.
9. Add a line in the following format to file `/etc/hosts`:

```
SAC_IP SAC_FQDN SAC_hostname
```

The variables in the preceding line are as follows:

- For *SAC\_IP*, type the IP address of the SAC.
- For *SAC\_FQDN*, type the fully qualified domain name (FQDN) of the SAC.
- For *SAC\_hostname*, type the hostname of the SAC.

For example, add the following line:

```
128.162.244.88 my-system-admin.domain-name.mycompany.com my-system-admin
```

10. Save and close file `/etc/hosts`.
11. Type the following command to set the SAC hostname:

```
# hostname SAC_hostname
```

For *SAC\_hostname*, type the hostname of the SAC.

For example:

```
# hostname my-system-admin
```

12. Use a text editor to create file `/etc/resolv.conf`.
13. Add lines to file `/etc/resolv.conf` that specify the search domain and the domain name service (DNS) servers at your site.

Later in the configuration process, when you run the cluster configuration tool, the tool uses the DNS servers you specify in this step for its defaults.

Specify lines with the following format:

```
search search_domain
nameserver name_server_IP
nameserver name_server_IP
```

The following is an example `resolv.conf` file:

```
search mydomain.com
nameserver 192.168.0.1
nameserver 192.168.0.25
```

14. Type the following `nscd(8)` command to force the invalidation of the name service cache daemon:

```
# nscd -i hosts
```

15. Type the following commands, in the order shown, to restart services:

```
# /etc/init.d/network restart
# /etc/init.d/rpcbind start
# /etc/init.d/nfslock start
```

16. Type the following command to retrieve the SAC's current time zone information:

```
# strings /etc/localtime | tail -1
CST6CDT,M3.2.0,M11.1.0
```

The previous output shows the SAC set to US Central time. If the output you see is not correct for this SGI ICE X system, perform the following steps:

- Type the following command to change to the directory that contains the time zone configuration files:

```
# cd /usr/share/zoneinfo
```

- Select a file from that directory that describes the time zone for the SAC.
- Type the following commands to enable the new time zone configuration file.

For example:

```
# /bin/cp -l /usr/share/zoneinfo/time_zone_file /etc/localtime.$$  
# /bin/mv /etc/localtime.$$ /etc/localtime
```

For *time\_zone\_file*, type the name of the time zone file that you need from the /usr/share/zoneinfo directory.

For example, type the following commands to change the SAC's time zone to US Pacific time:

```
# /bin/cp -l /usr/share/zoneinfo/PST8PDT /etc/localtime.$$  
# /bin/mv /etc/localtime.$$ /etc/localtime
```

- Type the following command to confirm the time zone:

```
# strings /etc/localtime | tail -1  
PST8PDT,M3.2.0,M11.1.0
```

17. (Conditional) Edit file /etc/ntp.conf to direct requests to the network time protocol (NTP) server at your site.

Complete the following steps if you want to direct requests to your site's NTP server instead of to the public time servers of the pool.ntp.org project:

- Use a text editor to open file /etc/ntp.conf.
- Insert a pound character (#) into column 1 of each of each line that includes rhel.pool.ntp.org.

---

**Note:** Do not edit or remove entries that serve the cluster networks.

---

- At the end of the file, add a line that points to your site's NTP server.

The following is an example of a correctly edited file:

```
# Use public servers from the pool.ntp.org project.  
# Please consider joining the pool (http://www.pool.ntp.org  
# server 0.rhel.pool.ntp.org  
# server 1.rhel.pool.ntp.org  
# server 2.rhel.pool.ntp.org  
server ntp.mycompany.com
```



The preceding output has been truncated at the right for inclusion in this guide.

- Type the following command to restart the NTP server:

```
# /etc/init.d/ntpd restart
```

18. (Conditional) Type a tilde character (~) and then a period character (.) to exit from the IPMI tool.

Perform this step if you connected to the system through the IPMI tool.

19. (Optional) Configure the system so that you can perform the installation from a VGA screen and can perform later operations from a serial console.

If you want to enable this capability, perform the following steps:

- Use a text editor to open file `/boot/grub/menu.lst`.
- Search the file for the word `kernel` at the beginning of a line.
- Add the following to the kernel line: `console=type`.

For example:

```
kernel /boot/vmlinuz-2.6.16.56-0.12-smp root=/dev/disk/by-label/sgiroot console=ttyS1,38400n8 splash=silent showopts
```

- Add the `console=type` parameter to the end of every kernel line. By default, this is set to `ttyS1,38400n8`. You might have `ttys2`, for example.

Later, if you want to access the SAC from only a VGA, you can remove the `console=` parameters.

20. Proceed to the following:

"Running the Cluster Configuration Tool" on page 45

## Configuring SUSE Linux Enterprise Server (SLES) on the SAC

The SLES YaST2 interface enables you to install the SLES operating system on the SGI ICE X system. To navigate the YaST2 modules, use key combinations such as the following:

- The Tab key moves the cursor forward, and the Shift + Tab keys move the cursor backward.

- The arrow keys move the cursor up, down, left, and right.
- To use shortcuts, press the Alt key + the highlighted letter.
- Press Enter to complete or confirm an action.
- Press Ctrl + L to refresh the screen.

For more information about navigation, see Appendix A, "YaST2 Navigation" on page 143.

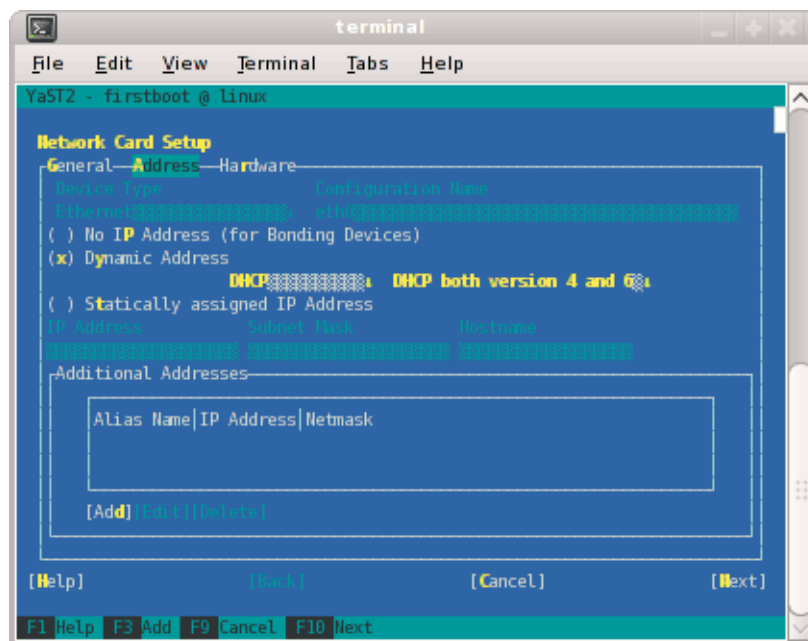
The following procedure explains how to use YaST2 to install SLES 11 on an SGI ICE X system.

**Procedure 3-7** To install SLES 11 on an SGI ICE X SAC

1. Connect to the system admin controller (SAC) by one of the following methods:
  - Through the intelligent platform management interface (IMPI) tool
  - Through the console attached to the SGI ICE X system
  - Through a separate keyboard, video display terminal, and mouse
2. On the **Language and Keyboard Layout** screen, complete the following steps:
  - Select your language
  - Select your keyboard layout
  - Select **Next**.
3. On the **Welcome** screen, select **Next**.
4. On the **Hostname and Domain Name** screen, complete the following steps:
  - Type the hostname for this SGI ICE X system.
  - Type the domain name.
  - Clear the box next to **Change Hostname via DHCP**. The box appears with an x in it by default, but you need to clear this box.
  - Select **Assign Hostname to Loopback IP**. Put an x in this box.
  - Select **Next**.
5. On the **Network Configuration** screen, complete the following steps:

- Select **Change**. A pop-up window appears.
  - On the pop-up window, choose **Network Interfaces**.
6. On the **Network Settings** screen, complete the following steps:
- Highlight the first network interface card that appears underneath **Name**.
  - Select **Edit**.
7. On the **Network Card Setup** screen, specify the SAC's house/public network interface.

Figure 3-3 on page 41 shows the **Network Card Setup** screen.



**Figure 3-3 Network Card Setup Screen**

Complete the following steps:

- Select **Statically Assigned IP Address**. SGI recommends a static IP address, not DHCP, for the SAC.

- In the **IP Address** field, type the system's IP address.
- In the **Subnet Mask** field, type the system's subnet mask.
- In the **Hostname** field, type the system's fully qualified domain name (FQDN). SGI requires you to type an FQDN, not the system's shorter hostname, into this field. For example, type `mysystem-admin.mydomainname.com`. Failure to supply an FQDN in this field causes the `configure-cluster` command to fail.
- Select **Next**.

You can specify the default route, if needed, in a later step.

8. On the **Network Settings** screen, complete the following steps:

- Select **Hostname/DNS**.
- In the **Hostname** field, type the system's fully qualified domain name (FQDN).
- In the **Domain Name** field, type the domain name for your site.
- Put an **x** in the box next to **Assign Hostname to Loopback IP**.
- In the **Name Servers and Domain Search List**, type the name servers for your house network.
- Back at the top of the screen, select **Routing**.

The **Network Settings > Routing** screen appears.

- In the **Default Gateway** field, type your site's default gateway.
- Select **OK**.

9. On the **Network Configuration** screen, click **Next**.

The **Saving Network Configuration** screen appears and saves your configuration.

10. On the **Clock and Time Zone** screen, complete the following steps:

- Select your region.
- Select your time zone.
- (Optional) In the **Hardware Clock Set To** field, choose **Local Time** or accept the default of **UTC**.

- Select **Next**.

This step synchronizes the time in the BIOS hardware with the time in the operating system. Your choice depends on how the BIOS hardware clock is set. If the clock is set to GMT, which corresponds to UTC, your system can rely on the operating system to switch from standard time to daylight savings time and back automatically.

11. On the **Password for System Administrator “root”** screen, complete the following steps:

- In the **Password for root User** field, type the password you want to use for the root user.

This password becomes the root user’s password for all the nodes on the ICE X system. These nodes are as follows:

- SAC
- Rack leader controller (RLC)
- Service nodes
- Compute nodes (blades)

- In the **Confirm password** field, type the root user’s password again.
- In the **Test Keyboard Layout** field, type a few characters.

For example, if you specified a language other than English, type a few characters that are unique to that language. If these characters appear in this plain text field, you can use these characters in passwords safely.

- Select **Next**.

12. On the **User Authentication Method** screen, select one of the authentication methods and select **Next**.

Typically, users accept the default (**Local**).

13. On the **New Local User** screen, create additional user accounts or select **Next**.

If you do not create additional users, select **Yes** on the **Empty User Login** warning pop-up window, and select **Next**.

14. On the **Installation Completed** screen, select **Finish**.

15. Type a tilde character (~) and then a period character (.) to exit from the IPMI tool.
16. Log into the SAC, open file `/etc/hosts` within a text editor, and verify that the SAC's fully qualified domain name (FQDN) and hostname are entered correctly.

For example, the following `/etc/hosts` file entry contains the correct data in the three required fields and is correct for a SAC with an IP address of `100.100.100.100`, an FQDN of `mysystem-admin.mydomain.com`, and a hostname of `mysystem-admin`:

```
100.100.100.100    mysystem-admin.mydomain.com    mysystem-admin
```

Make sure that the `/etc/hosts` file on the SAC contains the required information. If it does not, edit the `/etc/hosts` file to contain the three required fields as the preceding example shows.

17. Confirm that the system is working as expected.  
If necessary, restart YaST2 to correct settings.
18. (Optional) Configure the system so that you can perform the installation from a VGA screen and can perform later operations from a serial console.

If you want to enable this capability, perform the following steps:

- Use a text editor to open file `/boot/grub/menu.lst`.
- Search the file for the word `kernel` at the beginning of a line.
- Add the following to the `kernel` line: `console=type`.

For example:

```
kernel /boot/vmlinuz-2.6.16.56-0.12-smp root=/dev/disk/by-label/sgiroot console=ttyS1,38400n8 splash=silent showopts
```

- Add the `console=type` parameter to the end of every `kernel` line. By default, this is set to `ttyS1,38400n8`. You might have `ttys2`, for example.

Later, if you want to access the SAC from only a VGA, you can remove the `console=` parameters.

19. Proceed to the following:  
"Running the Cluster Configuration Tool" on page 45

## Running the Cluster Configuration Tool

The cluster configuration tool enables you to configure, or reconfigure, your SGI ICE X system. The procedure in this topic explains the general, required configuration steps for SGI ICE X systems. If your SGI ICE X system includes optional components, or if your site has specific requirements, later procedures explain how to use the cluster configuration tool to create a more customized environment.

The following procedure explains how to complete the following required cluster configuration steps:

- Create repositories for software installation files and updates.
- Install the system admin node (SAC) cluster software.
- Configure the cluster subdomain and examine other network settings. The cluster subdomain is likely to be different from the `eth0` domain on the SAC itself.
- Configure the NTP server.
- Install the cluster's software infrastructure. This step can take 30 minutes.
- Configure the house network's DNS resolvers.

### **Procedure 3-8** To run the cluster configuration tool

1. Locate your site's SGI software distribution DVDs or verify the path to your site's online software repository.

You can install the software from either physical media or from an ISO on your network.

2. From the VGA screen, or through an `ssh` connection, log into the system admin controller (SAC) as the root user.

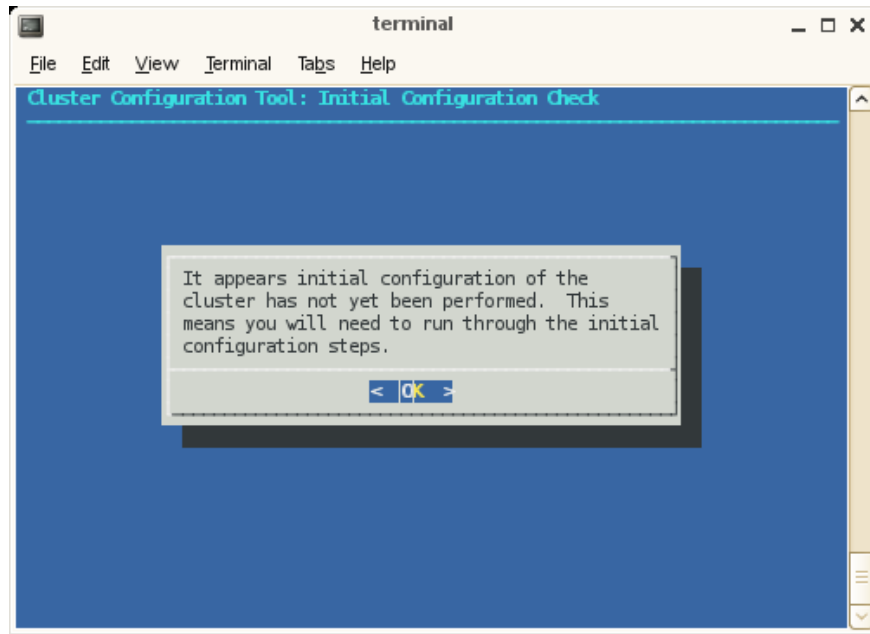
SGI recommends that you run the cluster configuration tool either from the VGA screen or from an `ssh` session to the system admin controller (SAC). Avoid running the `configure-cluster` command from a serial console.

3. Type the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

4. On the cluster configuration tool's **Initial Configuration Check** screen, select **OK** on the initial window.

Figure 3-4 on page 46 shows the initial window.



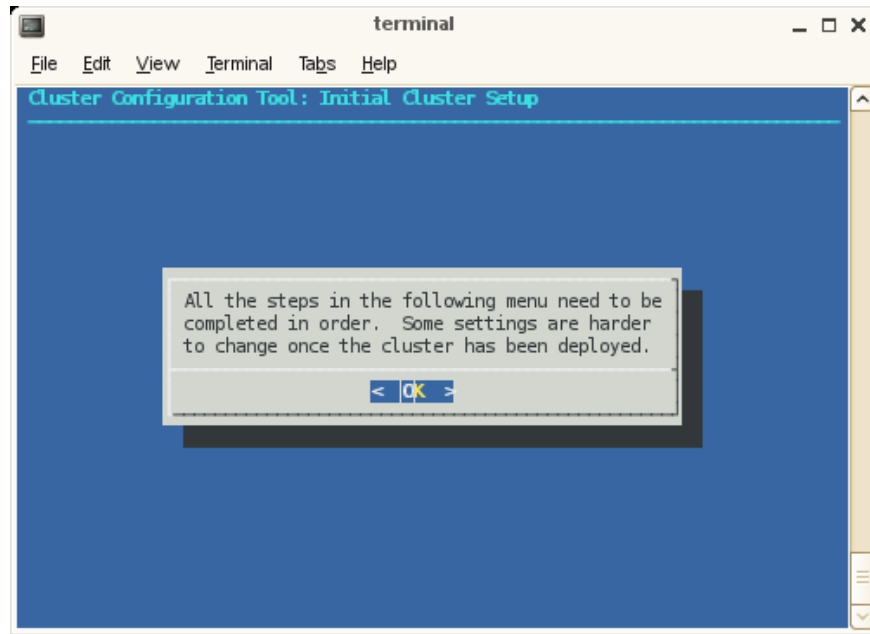
**Figure 3-4 Initial Configuration Check Screen**

The cluster configuration tool recognizes a configured cluster. If you start the tool on a configured SGI ICE X system, it opens into the **Main Menu**.

5. On the **Initial Cluster Setup** screen, select **OK** on the screen.

Figure 3-5 on page 47 shows the window.

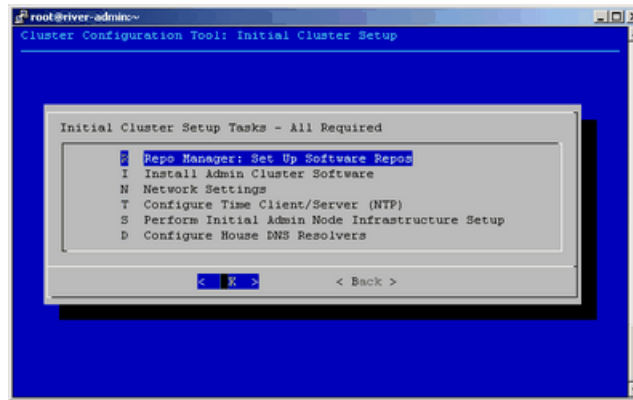




**Figure 3-5 Initial Cluster Setup** Screen with the initial screen

6. On the **Initial Cluster Setup** screen, select **R Repo Manager: Set Up Software Repos**, and click **OK**.

Figure 3-6 on page 48 shows the **Initial Cluster Setup** screen with the task menu. This procedure guides you through the tasks you need to perform for each of the menu selections on the **Initial Cluster Setup** screen.



**Figure 3-6 Initial Cluster Setup Tasks** Screen

The next few steps create software repositories for the initial installation packages and for updates. You need to create repositories for the following software:

- The operating system software, either RHEL or SLES
  - SGI Foundation Suite
  - SGI Tempo
  - Additional software packages for which you hold licenses, such as the Message Passing Toolkit, the SGI Performance Suite, and any others
7. On the **One or more ISOs were embedded on the ...** screen, select **Yes**.
  8. On the **Repositories are created ...** screen, press **Enter**.
  9. On the **You will now be prompted to add additional media ...** screen, select **OK**.
  10. On the **Would you like to register media with Tempo? ...** screen, select **Yes**.
  11. On the **Please either insert the media in your DVD drive ...** screen, select either **Insert DVD** or **Use Custom path/url**.

Proceed as follows:

- To install the software from DVDs, perform the following steps:
  - Insert a DVD.

- Select **Mount inserted DVD**.
- On the **Media registered successfully with crepo ...** screen, select **OK**, and eject the DVD.
- On the **Would you like to register media with Tempo? ...** screen, select **Yes** if you have more software that you need to register.

If you select **Yes**, repeat the preceding tasks in this sequence for the next DVD.

If you select **No**, proceed to the next step.

- To install the software from a network location, perform the following steps:
  - Select **Use custom path/URL**.
  - On the **Please enter the full path to the mount point or the ISO file ...** screen, type the full path in *server\_name:path\_name/iso\_file* format. This field also accepts a URL or an NFS path. Select **OK** after typing the path.
  - On the **Media registered successfully with crepo ...** screen, select **OK**.
  - On the **Would you like to register media with Tempo? ...** screen, select **Yes** if you have more software that you need to register.

If you select **Yes**, repeat the preceding tasks in this sequence for the next DVD.

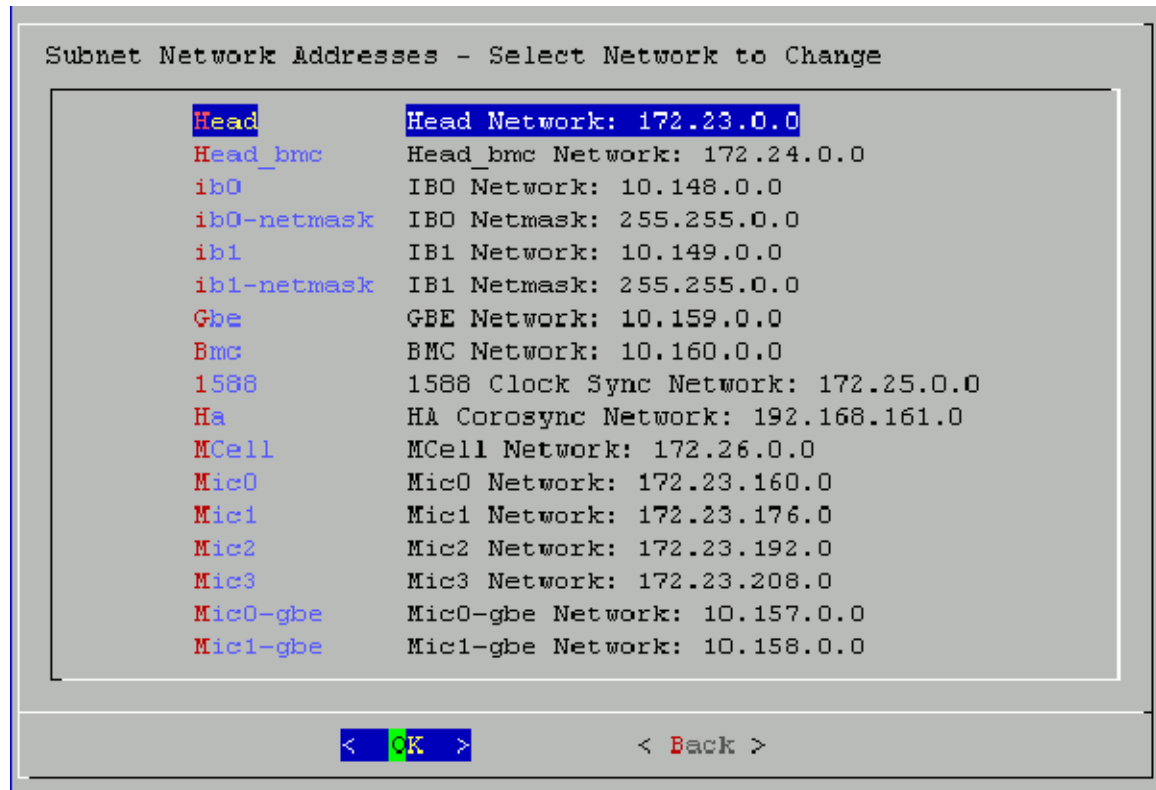
If you select **No**, proceed to the next step.

12. On the **Initial Cluster Setup Tasks** screen, select **I Install Admin Cluster Software**, and select **OK**.

This step installs the cluster software that you wrote to the repositories.

13. On the **Initial Cluster Setup Tasks** screen, select **N Network Settings**, and select **OK**.
14. On the **Cluster Network Settings** screen, select **S Configure Subnet Addresses**, and select **OK**.
15. On the **Warning: Changing the subnet IP addresses ...** screen, click **OK**.
16. Review the settings on the **Subnet Network Addresses** screen, and modify these settings only if absolutely necessary.

Figure 3-7 on page 50 shows the **Subnet Network Addresses** screen. This screen displays the default networks and netmasks that reside within the SGI ICE X cluster.



**Figure 3-7** Subnet Network Addresses Screen

If you accept the defaults, select **OK**.

If you do not accept the defaults, you can change the network settings. For example, it is possible that your site has existing networks or conflicting network requirements. For additional information about the IP address ranges, see Appendix B, "Subnetwork Information" on page 145. Complete the following steps if you need to change the network settings:

- Highlight the setting you want to change, and select **OK**.

- Type in a new IP address, and select **OK**.
- Press Enter.

On the **Update Subnet Addresses** screen, the **Head Network** field shows the SAC's IP address. SGI recommends that you do not change the IP address of the SAC or rack leader controllers (RLCs) if at all possible. You can change the IP addresses of the InfiniBand network (**IB0** and **IB1**) to match the IP requirements of the house network, and then select **OK**.

17. On the **Cluster Network Settings** screen, select **D Configure Cluster Domain Name**, and select **OK**.
18. On the **Please enter the domain name for this cluster** pop-up window, type the domain name, and select **OK**.

The domain you type becomes a subdomain to your house network.

For example, type `ice.americas.sgi.com`.

19. On the **Cluster Network Settings** screen, select **Back**.
20. On the **Initial Cluster Setup** screen, select **T Configure Time Client/Server (NTP)**, and select **OK**.
21. Configure your NTP server.

On the subsequent screens, you set the SAC as the time server to the SGI ICE X system. For this step, the installer screens differ on RHEL platforms and SLES platforms.

On RHEL platforms, complete the following step:

- On the **A new ntp.conf has been put in to position ...** screen, select **OK**.

On SLES platforms, complete the following steps:

- On the **A new ntp.conf has been put in to position ...** screen, select **OK**.
  - Use the YaST interface and the SLES documentation to guide you through the NTP configuration.
  - On the **This procedure will replace your ntp configuration file ...** screen, select **Yes**.
22. On the **Initial Cluster Setup Tasks** menu, select **S Perform Initial Admin Node Infrastructure Setup**, and select **OK**.

- 23. On the **A script will now perform the initial cluster ...** screen, select **OK**.

This step runs a series of scripts that configure the SAC on the SGI ICE X system. The scripts also create the root images for the RLCs, service nodes, and compute nodes. The scripts run for approximately 30 minutes. At the end, the script issues a line that includes **install-cluster completed** in its output.

The final output of the script is as follows:

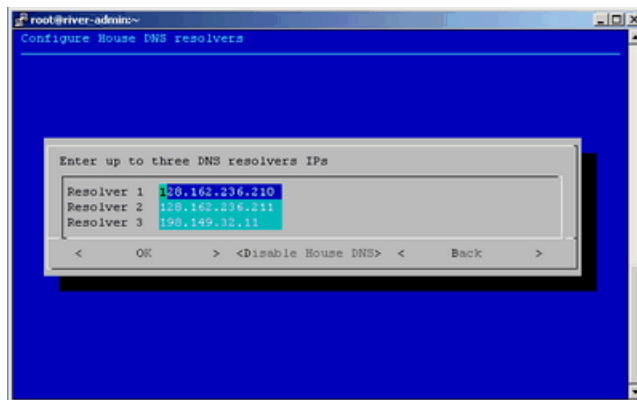
```
/opt/sgi/sbin/create-default-sgi-images Done!
```

The output of the `mkssiimage` commands are stored in a log file at the following location:

```
/var/log/cinstallman
```

- 24. On the **Initial Cluster Setup Complete** window, select **OK**.
- 25. On the **One or more ISOs were embedded on the admin install DVD and copied to ...**, screen, select **OK**.
- 26. On the **Initial Cluster Setup** menu, select **D Configure House DNS Resolvers**, and select **OK**.

Figure 3-8 on page 52 shows the **Configure House DNS Resolvers** screen.



**Figure 3-8 Configure House DNS Resolvers Screen**

The system autopopulates the values on the **Configure House DNS Resolvers** screen to match the DNS specifications on the SAC. The DNS resolvers you specify here enable the service nodes to resolve host names on your network. You can set the DNS resolvers to the same name servers used on the SAC itself.

Perform one of the following actions:

- To accept these settings, select **OK**, and then select **Yes**.
- To change the settings, type in different IP addresses, select **OK**, and then select **Yes**.
- To disable house network resolvers, select **Disable House DNS**.

On the **Setting DNS Forwarders to ...** screen, select **Yes**.

27. On the **Initial Cluster Setup** screen, select **Back**.

This action returns you to the cluster configuration tool main menu.

28. On the **Main Menu**, select **S Configure Switch Management Network (optional)**, and select **OK**.

On an SGI ICE X system, the switch management network enables the Ethernet switch to control all VLANs and trunking.

29. On the pop-up window that appears, select **Y yes**, and select **OK**.

Figure 3-9 on page 53 shows the selection pop-up window:

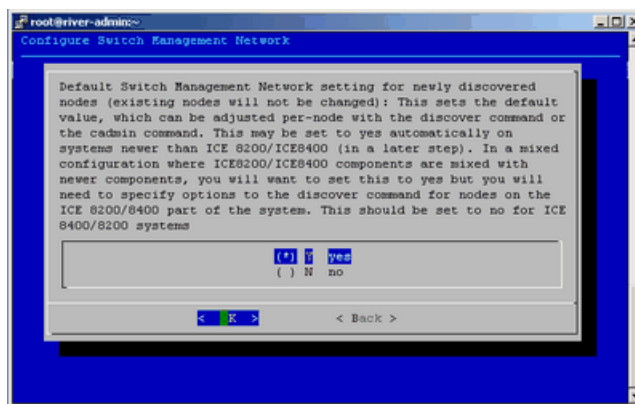


Figure 3-9 Configure Switch Management Network screen

30. (Conditional) On the **Main Menu**, select **N Configure MCell Network (optional)**, and select **OK**.

Perform this step if your SGI ICE X system contains MCells.

31. (Conditional) On the screen that appears, select **Y yes**, and select **OK**.

Perform this step if your SGI ICE X system contains MCells.

32. Select **Quit**.

33. Type the `cattr list -g` command to verify the features you configured with the cluster configuration tool.

The following example output is appropriate for an SGI ICE X system with MCells. If your system does not include MCells, the `mcell_network` value should display `no`. The output is as follows:

```
# cattr list -g
global
  redundant_mgmt_network : yes
  switch_mgmt_network   : yes
  mcell_network          : yes
  discover_skip_switchconfig : no
  max_rack_irus          : 8
  blademon_d_scan_interval : 120
  my_sql_replication     : no
  rack_vlan_start        : 101
  rack_vlan_end          : 1100
  head_vlan              : 1
  mcell_vlan             : 3
  replication_file       : mysql-bin.000132
  replication_position   : 6580967
```

If you need to respecify any global values, start the cluster configuration tool again, and correct your specifications. To start the cluster configuration tool, type the following command:

```
# /opt/sgi/sbin/configure-cluster
```

34. Proceed to one of the following:



- To configure one or more external Domain Name Service (DNS) servers, proceed to "(Conditional) Configuring External Domain Name Service (DNS) Servers" on page 55.
- To enable the Intel Data Center Manager (DCM) Software Development Kit (SDK), proceed to "(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK)" on page 56.

## (Conditional) Configuring External Domain Name Service (DNS) Servers

Perform the procedure in this section if you want to enable network address translation (NAT) gateways for the SGI ICE X system. A later procedure explains how to configure NAT on one of your service nodes. If you want to enable NAT, perform the procedure in this topic at this time.

When external DNS and NAT are enabled, the host names for the compute nodes (blades) in the cluster resolve through external DNS servers. The compute nodes need to be able to reach your house network.

---

**Note:** You cannot configure this feature after you run the `discover` command. If you configure this feature after you run the `discover` command, the IP addresses assigned previously on the configured nodes remain.

---

The following procedure explains how to configure external DNS servers.

**Procedure 3-9** To configure external DNS servers

1. Obtain a large block of IP addresses from your network administrator.

This feature requires you to reserve a block of IP addresses on your house network. If you want to use external DNS servers, all nodes on the InfiniBand networks, both the `ib0` and `ib1` networks are included. The external DNS is enabled to provide addresses for all rack leader controllers (RLCs), all service nodes, and all compute nodes.

2. Through an `ssh` connection, log into the system admin controller (SAC) as the root user.
3. Type the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

4. Select **E Configure External DNS Masters (optional)**, and select **OK**.
5. On the **This option configures SGI Tempo to look up the IP addresses for the InfiniBand networks from external DNS servers ...** screen, select **Yes**.
6. On the **Enter up to five external DNS master IPs** screen, type the IP addresses of up to five external DNS servers on your house network, and select **OK**.
7. On the **Setting external DNS masters to ip\_addr**, select **Yes**.
8. Proceed to one of the following:
  - To enable the Intel Data Center Manager (DCM) Software Development Kit (SDK), proceed to the following:

"(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK)" on page 56.
  - To synchronize the software repository, proceed to the following:

"Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images" on page 64.

## **(Optional) Configuring the Intel Data Center Manager (DCM) Software Development Kit (SDK)**

The Intel DCM SDK enables you to monitor and manage energy consumption in your data center. This feature is easier to enable if you have not yet installed the SGI Management Center (SMC) license key because in this situation, the system creates DCM entries in the `/opt/sgi/sgimc/etc/site-site.profile` for you. If you want to enable DCM on a production system, however, you can use this procedure because it includes steps that explain how to enable DCM on a deployed system with installed license keys.

The following procedures explain how to configure the DCM SDK:

- "Verifying the Environment and Installing the DCM SDK" on page 57
- "(Conditional) Installing the SGI Management Center License Key" on page 61
- "Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration" on page 61

## Verifying the Environment and Installing the DCM SDK

The following procedure explains how to verify your SGI ICE X environment and determine if you can enable the DCM SDK.

**Procedure 3-10** To verify your environment

1. Verify that your SGI ICE X system supports the following:

- The Intel Intelligent Power Node Manager (IPNM)-capable compute nodes. Consult your SGI technical support representative if you are unsure.
- The SMC DCM plugin

If you are unsure about the preceding requirements, contact your SGI representative or review your hardware specifications. You can configure DCM at any time, but it is easier to configure at this point in the installation process because you have not yet installed the SGI Management Center (SMC) license key.

2. Verify that your SGI Management Center (SMC) license includes both of the following keys:

- A premium license
- A power opt license

If you are installing a completely new SGI ICE X system, there is no SMC license installed on the system at this time, and you need information about what your license includes, contact your SGI representative to verify your licenced features.

If you are enabling the DCM feature on a production SGI ICE X system that already has an SMC license installed upon it, use the `lk_verify` command to verify your system's license key. The following example shows these keys:

```
# lk_verify
1 /etc/lk/keys.dat:007      product=SGIMC, version=1.000, count=0, begDate=1374482126, \
  expDate=1375765199, licenseID=0, key=Iym9+MynzDO2R0ABCcqJjgrXYqgu6rwp, \
  info='SMC 1.7 PREM SVR/CL',attr='ST=CL NC=0 LC=24FF', \
  vendor='Silicon Graphics International',ref_id='279071'
  Verdict:      SUCCESS.

2 /etc/lk/keys.dat:018      product=SGIMC, version=1.000, count=0, begDate=1373396727, \
  expDate=1381208399, licenseID=b9e99f78, key=xybn6FAs4rx2UoC7WzclQE1AQ7DVbc+6, \
  info='SGIMC POWER OPT',attr='LC=0200', vendor='Silicon Graphics International', \
  ref_id='278444'
```

Verdict: SUCCESS.

In the preceding `lk_verify` output, license 1 is the premium license, and license 2 shows the power option in the `info` and `attr` fields.

3. (Conditional) Back up the current DCM configuration files on the system admin controller (SAC) and on all rack leader controller (RLC) nodes.

Perform this step if you already have the DCM SDK installed on this SGI ICE X system and you are reinstalling the DCM SDK. This can be the case if you are troubleshooting an incomplete installation.

The steps are as follows:

- Log into the SAC or into one of the RLCs as the root user.
- Type the following command to stop the DCM daemon:

```
# service mcdcm stop
```

- Type the following command to create a scratch directory and write the existing DCM configuration files to that directory:

```
# mkdir scratch; cp -a /opt/sgi/mcdcm/conf /scratch/mcdcm-conf
```

- Type the following command to remove the DCM package:

```
# rpm -e mcdcm
```

- Type the following command to remove any existing initialization scripts:

```
# rm -rf /etc/init.d/dcm* /etc/init.d/mcdcm*
```

- Type the following command to remove the `mcdcm` daemon:

```
# rm -rf /opt/sgi/mcdcm
```

- Repeat the preceding steps on the other computers in the SGI ICE X cluster.

4. Type the following commands to install the SMC DCM package:

```
# cinstallman --yum-node --node admin install mcdcm
# cinstallman --yum-node --node r1lead install mcdcm
```

If you have multiple RLCs, repeat the preceding command on other RLCs as needed.

5. On the SAC, type the following command to verify the installation:

```
# cat /opt/sgi/mcdcm/logs/dcmcmdstatus.log
```

The command returns information such as the following if the installation was successful:

```
Success. You can now start the database server using:
[...]
waiting for server to start.... done
server started
```

6. On the SAC, type the following commands to change localhost to admin in the system files:

```
# sed -i 's/localhost/admin/g' \
/opt/sgi/mcdcm/conf/web.config.xml
# sed -i 's/localhost/admin/g' \
/opt/sgi/mcdcm/external/apache-tomcat/webapps/DCMWSdl/internal.wsdl
# sed -i 's/localhost/admin/g' \
/opt/sgi/mcdcm/external/apache-tomcat/webapps/DCMWSdl/dcm.wsdl
```

Note that the preceding commands use the backslash character (\) because the command continues to a second line.

7. On each RLC, type the following commands to change localhost to admin in the system files:

```
# pdsh -g leader 'sed -i "s/localhost/${HOSTNAME}/g" \
/opt/sgi/mcdcm/conf/web.config.xml'
# pdsh -g leader 'sed -i "s/localhost/${HOSTNAME}/g" \
/opt/sgi/mcdcm/external/apache-tomcat/webapps/DCMWSdl/internal.wsdl'
# pdsh -g leader 'sed -i "s/localhost/${HOSTNAME}/g" \
/opt/sgi/mcdcm/external/apache-tomcat/webapps/DCMWSdl/dcm.wsdl'
```

Note that the preceding commands use the backslash character (\) because the command continues to a second line.

8. Type the following commands to start the SMC DCM daemons on the SAC and on the RLCs:

```
# service mcdcm start
# pdsh -g leader service mcdcm start
```

9. Type the following commands on the SAC and on the RLCs, and observe the output:

```
# ps -ef | grep jsvc | grep -i "startup.Bootstrap"  
# ps -ef | grep jsvc | grep -i "DaemonImpl"
```

In the output, make sure that the `com.intel.dcm.daemon.DaemonImpl` and `org.apache.catalina.startup.Bootstrap` daemons started.

10. On the SAC, and on each RLC, type the following command to make sure that the DCM daemon is started:

```
# grep "process started" /opt/sgi/mcdcm/logs/DatacenterManager.log  
Intel(R) Data Center Manager 2.x.xxxxx process started!
```

11. On the SAC, type the following command to install the `sgimc-dcm` daemon:

```
# cinstallman --yum-node --node admin install sgimc-dcm
```

12. On the SAC, use a text editor to open file `/opt/sgi/sgimc/etc/dcm.properties`, search for the string `dcm.bmc`, and modify the existing lines to be as follows:

```
dcm.bmc.user = ADMIN  
dcm.bmc.password = ADMIN
```

The preceding lines are case sensitive.

13. On the SAC, use a text editor to open file `/opt/sgi/sgimc/etc/tempo.properties`, search for the string `dcm.enabled`, and modify the existing line (or add it) to be as follows:

```
dcm.enabled = true
```

14. Proceed to one of the following:

- If you need to install SMC license keys, for example on a new system, proceed to the following:

"(Conditional) Installing the SGI Management Center License Key" on page 61

- If the correct SMC license keys are already installed on your system, proceed to the following:

"Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration" on page 61

## (Conditional) Installing the SGI Management Center License Key

Perform the procedure in this topic if you need to install an SMC license key on your SGI ICE X system.

The SGI Management Center (SMC) software runs on the system admin controller (SAC). SMC provides a graphical user interface for system configuration, operation, and monitoring. For more information about using SMC, see *SGI Management Center (SMC) System Administrator Guide*. For more information about licensing, see the licensing FAQ on the following website:

<http://www.sgi.com/support/licensing/faq.html>

The following procedure explains how to obtain and install the license key for SMC.

**Procedure 3-11** To license the SMC software

1. Use a text editor to open file `/etc/lk/keys.dat`.
2. Copy and paste the license key exactly as it was given to you.

There can be line breaks in the license key.

3. Save the file.
4. Type the following command to restart the SMC daemon:

```
# service mgr restart
```

5. Type the following command to start SMC:

```
# mgrclient
```

6. Proceed to "Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration" on page 61.

## Completing the Intel Data Center Power Manager (DCM) Software Development Kit (SDK) Configuration

The following procedure explains how to complete the DCM enablement procedure.

---

**Note:** Some output lines in this procedure were wrapped for inclusion in this documentation.

---

**Procedure 3-12** To complete the Intel DCM enablement procedure

1. Log into the system admin controller (SAC) and type the following command to restart the mgr daemon:

```
# service mgr restart
```

2. On the SAC, use a text editor to open file `/opt/sgi/sgimc/etc/perfhost_options.profile`, and search for `tempo.dcm`.

Verify that lines similar to the following appear in the `perfhost_options.profile` file:

```
tempo.dcmBmc.r1i0n1=192.168.160.101
tempo.dcmBmc.r1i0n0=192.168.160.100
```

3. On the SAC, verify the contents of file `/opt/sgi/sgimc/etc/site-site.profile`.

At this time, the `site-site.profile` file should contain a line for each node that declares the rack leader controller (RLC) for each node. These lines are similar to the following:

```
PlatformManagementService.hostsettings.r1i0n0:BEGIN1,CPOWER,false,localhost,4,END1,BEGIN2,CONSERVER,false,
localhost,admin,END2,BEGIN3,DCM,false,r1lead,400,0,,4,IPNM2,END3,1,1,0,2,0
PlatformManagementService.hostsettings.r1i0n1:BEGIN1,CPOWER,false,localhost,4,END1,BEGIN2,CONSERVER,false,
localhost,admin,END2,BEGIN3,DCM,false,r1lead,400,0,,4,IPNM2,END3,1,1,0,2,0
```

If the `site-site.profile` file does not contain a line for each node, you need to add these lines manually. This can be the case if you installed the SMC license and are now using this procedure to enable DCM. You can use a loop, such as the following, to add these lines:

```
for node in `cnodes --compute --rack 1`; do
    echo PlatformManagementService.hostsettings.$node:BEGIN1,CPOWER,false,localhost,4,END1,BEGIN2,CONSERVER,false,
localhost,admin,END2,BEGIN3,DCM,false,r1lead,400,0,,4,IPNM2,END3,1,1,0,2,0 >> /opt/sgi/sgimc/etc/site-site.profile
done
```

4. (Optional) Add entries to the `/opt/sgi/sgimc/etc/site-site.profile` file for the RLCs and service nodes.

Perform this step if you want to enable DCM on nondefault nodes such as the RLCs and service nodes.



Edit the following configuration files to define DCM for the RLCs and service nodes:

- Open file `site-site.profile` and add IP address information, such as the following examples show, to the file:

```
tempo.dcmBmc.rllead=172.24.0.2
tempo.dcmBmc.service0=192.168.160.109
```

- Open file `/opt/sgi/sgimc/etc/site-site.profile` and add information such as the following:

```
PlatformManagementService.hostsettings.rllead:BEGIN1,CPOWER,false,localhost,4,END1,BEGIN2,CONSERVER,false,localhost,admin,END2,BEGIN3,DCM,false,admin,400,0,,4,IPNM2,END3,1,1,0,2,0
PlatformManagementService.hostsettings.service0:BEGIN1,CPOWER,false,localhost,4,END1,BEGIN2,CONSERVER,false,localhost,admin,END2,BEGIN3,DCM,false,admin,400,0,,4,IPNM2,END3,1,1,0,2,0
```

5. Type the `mgrclient` command to start the SGI Management Center GUI.
6. Click **Edit > Preferences > Platform Management**.

In the window that appears, select or set the following:

- To the right of **Management Device 3**, click **Configure**.  
Select **DCM** as the third platform device.
- Verify the defaults and change defaults if needed. The defaults are as follows:

```
DCM Primary Host Name: localhost
Derated Power: 400
Nameplate Power 450
Unmanaged Power 100
Cooling Power Multiplier: 0.5
Power Cost Per KWh: 1
Device Connector Type: IPNM2
Platform Management User: ADMIN
```

For information about the defaults, see the *SGI Management Center System Administrator's Guide*.

- Select the **policy** button.
- Click **OK**.

- Click **OK**.
7. On the SAC, type the following command to restart the `mgr` daemon:  

```
# service mgr restart
```
  8. Wait for the changes to be applied to your SGI ICE X system.  
On a large cluster, these changes can take 30 minutes or more to complete.
  9. Type the `mgrclient` command to restart the SGI Management Center GUI.
  10. In the SGI Management Center GUI, click **Admin > Instrumentation > Power**.  
Visually inspect the information on the **Power** tab to verify that it presents data for the cluster and for the individual nodes.
  11. Proceed to the following:  
"Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images" on page 64
- For DCM enablement troubleshooting information, see "Repairing a Broken DCM Enablement" on page 124.

## Synchronizing the Software Repository, Installing Software Updates, and Cloning the Images

The following procedure explains how to update the software in the repositories that you created with the cluster configuration tool. The following procedure assumes that the SGI ICE X system has a connection to the internet. If you need to perform this procedure on a secure SGI ICE X system, you need to modify this procedure. For a secure system, obtain the software updates from SGI SupportFolio manually and use the `crepo` command to install the software manually.

**Procedure 3-13** To update the software

1. Through an `ssh` connection, log into the system admin controller (SAC) as the root user.
2. Type the following command to retrieve information about the network interface card (NIC) bonding method on the SAC:

```
# cadmin --show-mgmt-bonding --node admin
```

If bonding has been set appropriately, the command returns 802.3ad.

If the command does not return 802.3ad, type the following commands to set the bonding appropriately and reboot the system:

```
# cadmin --set-mgmt-bonding --node admin 802.3ad
# reboot
```

3. Type the following command to retrieve the new images from SGI SupportFolio and the operating system vendor:

```
# sync-repo-updates
```

For RHEL-based systems, make sure the system is subscribed as `rhel-x86_64--server-6`.

This step requires that the system be connected to the internet. Contact your SGI representative if this update method is not acceptable for your site.

4. Type the `cinstallman --show-images` command to retrieve the image names.

For example:

```
# cinstallman --show-images
Image Name          BT Path
compute-rhel6.5    1 /var/lib/systemimager/images/compute-rhel6.5
service-rhel6.5    0 /var/lib/systemimager/images/service-rhel6.5
lead-rhel6.5       0 /var/lib/systemimager/images/lead-rhel6.5
```

The preceding output includes a line for the MPSS images if you have downloaded MPSS software for Intel Many Integrated Core Architecture (Intel MIC Architecture) based products.

5. (Optional) Clone the images.

Perform this step if you want to back up the current images before they are installed.

Type the following command:

```
cinstallman --create-image --clone --source src_image_name --image image
```

For `src_image_name`, specify the name of the source image. For example:  
lead-rhel6.5.

For *image*, specify a file name for the copied file (the clone). For example:  
lead-rhel6.5.backup

For example:

```
# cinstallman --create-image --clone --source compute-rhel6.5 --image compute-rhel6.5.backup
# cinstallman --create-image --clone --source service-rhel6.5 --image service-rhel6.5.backup
# cinstallman --create-image --clone --source lead-rhel6.5 --image lead-rhel6.5.backup
```

6. Type a series of `cinstallman --update-image` commands to update the software images.

For *image*, specify the software packages displayed in the previous step.

For example, to install the packages shown in Procedure 3-13, step 4 on page 65, type the following commands:

```
# cinstallman --update-image --image compute-rhel6.5
# cinstallman --update-image --image service-rhel6.5
# cinstallman --update-image --image lead-rhel6.5
```

7. Proceed to one of the following:

- If your SGI ICE X system contains Intel® Many Integrated Core Architecture (Intel MIC Architecture) devices on the compute nodes and/or the service nodes, proceed to "(Conditional) Downloading the Intel Manycore Platform Software Stack (MPSS) Software and Creating Images" on page 66.
- If your SGI ICE X system does not contain MIC devices on the compute nodes and/or the service nodes, proceed to "Configuring the Switches" on page 76.

## **(Conditional) Downloading the Intel Manycore Platform Software Stack (MPSS) Software and Creating Images**

Perform the procedures in this topic if your SGI ICE X system is equipped with Intel Many Integrated Core Architecture (Intel MIC Architecture) based products. The Intel Many Integrated Core (MIC) devices are part of the Intel Manycore Platform Software Stack that runs on the Intel Xeon Phi Coprocessors found on SGI ICE X compute nodes and service nodes.

Intel Corporation provides software for its Intel MIC architecture products, and you need to download this software for use on your SGI ICE X system. The MPSS package that you download contains the software packages for the MIC devices on the compute blades and service nodes. The procedures in this topic explain how to download the RPMs from the Intel Corporation website and how to create images for compute nodes and service nodes that are equipped with MIC devices. Your system might have MIC devices on the compute nodes, on the service nodes, or both. Complete the procedures that are appropriate for your hardware configuration.

The following procedures explain how to obtain and deploy the MPSS software from Intel Corporation:

- "Downloading the MPSS Software From the Intel Corporation" on page 67
- "Creating Images for the Compute Nodes That Include MIC Devices" on page 68
- "Creating Images for the Service Nodes That Include MIC Devices" on page 72

## Downloading the MPSS Software From the Intel Corporation

Complete the procedure in this topic if you have any MIC devices on your SGI ICE X system.

The following procedure explains how to download the MPSS software from Intel Corporation.

**Procedure 3-14** To download the MPSS package

1. Open a browser, and navigate to the following website:

<http://software.intel.com/mic-developer>

2. Click the **Tools & Downloads** tab.
3. Click the **Software Drivers: Intel Manycore Platform Software Stack (Intel MPSS)** link.

Follow the instructions on the website to download the Linux software version for your operating system platform. The download comes in the form of a `tar(1)` file.

4. Use the instructions from Intel to build the RPM files that you need.

A later procedure explains how to transfer these files to the SGI ICE X system and build new images.

5. Proceed to one of the following topics:

- "Creating Images for the Compute Nodes That Include MIC Devices" on page 68
- "Creating Images for the Service Nodes That Include MIC Devices" on page 72

Perform both of the preceding procedures if you have MIC devices on both compute nodes and service nodes.

## Creating Images for the Compute Nodes That Include MIC Devices

The following procedure explains how to create compute node images that include MIC device software.

**Procedure 3-15** To create compute node images for compute nodes with MIC devices

1. On the SAC, use the `mkdir(1)` command, in the following format, to create a directory for the RPM repository:

```
mkdir -p /tftpboot/intel/mpss_repository_directory
```

For *mpss\_repository\_directory*, type a name for the directory that is to contain the MPSS repository. For convenience, make sure to include an identifier for the MPSS release level you downloaded.

For example:

```
# mkdir -p /tftpboot/intel/mpss_u3-2.1.6720-19
```

2. Use operating system commands to copy the RPM files you downloaded to the `/tftpboot/intel/mpss_repository_directory` directory on the SAC node.

For example, use `cp(1)`, `ftp(1)`, `rsync(1)`, `scp(1)`, or another method.

3. Use the `crepo` command, in the following format, to specify a custom repository for the MPSS RPMs:

```
crepo --add rpm_repo_directory -custom rpm_repo_name
```

The variables in this command are as follows:

- For *rpm\_repo\_directory*, specify the full path to the directory that contains the RPM files.

- For *rpm\_repo\_name*, create a name for the image. You can specify the same name for both *rpm\_repo\_directory* and *rpm\_repo\_name*. After the image is built, the `cinstallman --show-images` command returns this label in the **Image Name** column of its output.

For example:

```
# crepo --add /tftpboot/intel/mpss_u3-2.1.6720-19 --custom mpss_u3-2.1.6720-19
```

4. Type the following command to confirm that the MPSS image is in the correct repository:

```
# crepo --show
```

5. Use the `crepo` command, in the following format, to add the custom repository to the generated RPM list in `/etc/opt/sgi/rpmlists`:

```
crepo --select rpm_repo_name
```

For *rpm\_repo\_name*, create a name for the image. You can specify the same name for both *rpm\_repo\_directory* and *rpm\_repo\_name*. After the image is built, the `cinstallman --show-images` command returns this label in the **Image Name** column of its output. This is the same *rpm\_repo\_name* that you specified in the following step:

Procedure 3-15, step 3 on page 68

For example, the following command adds the custom repository and displays the content of the repository:

```
# crepo --select mpss_u3-2.1.6720-19
Selecting: mpss_u3-2.1.6720-19
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-lead-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
```

6. Type the following command to confirm that you selected the new repository that contains the MPSS RPMs:

```
# crepo --show
* mpss_u3-2.1.6720-19 : /tftpboot/intel/mpss_u3-2.1.6720-19
* Red-Hat-Enterprise-Linux-6.5 : /tftpboot/distro/rhel6.5
* SGI-MPI-1.7-rhel6 : /tftpboot/sgi/SGI-MPI-1.7-rhel6
* SGI-Management-Center-1.9-rhel6 : /tftpboot/sgi/SGI-Management-Center-1.9-rhel6
* SGI-Foundation-Software-2.9-rhel6 : /tftpboot/sgi/SGI-Foundation-Software-2.9-rhel6
```

```
* SGI-Tempo-2.9-rhel6 : /tftpboot/sgi/SGI-Tempo-2.9-rhel6
```

The asterisk character (\*) in column 1 indicates an image that is selected.

7. Type the following command to display the images that are available for installation on the compute nodes:

```
# cinstallman --show-images
Image Name          BT Path
lead-rhel6.5        0 /var/lib/systemimager/images/lead-rhel6.5
service-rhel6.5     0 /var/lib/systemimager/images/service-rhel6.5
compute-rhel6.5     1 /var/lib/systemimager/images/compute-rhel6.5
```

8. Use the `cinstallman` command, in the following format, to clone the current operating system image:

```
cinstallman --create-image --clone --source current_image --image new_image
```

The variables in this command are as follows:

- For *current\_image*, type the name of the operating system image you want to use that is on the system right now. Choose one that appears in the output from the `cinstallman --show-images` command in the **Image Name** list. For example, `compute-rhel6.5`.
- For *new\_image*, type a new name for the operating system image that is to include the MPSS file RPMs. SGI recommends that you include information in the new name that can enable you to identify this image as the operating system image that includes MPSS software. For example, `compute-rhel6.5-mic-6720-19` identifies the new image as a RHEL image that contains a revision of the MPSS MIC software.

For example:

```
# cinstallman --create-image --clone --source compute-rhel6.5 --image compute-rhel6.5-mic-6720-19
```

9. Type the following command to display the images and confirm that the new image appears in the list:

```
# cinstallman --show-images
Image Name          BT Path
lead-rhel6.5        0 /var/lib/systemimager/images/lead-rhel6.5
service-rhel6.5     0 /var/lib/systemimager/images/service-rhel6.5
compute-rhel6.5     1 /var/lib/systemimager/images/compute-rhel6.5
```



```
compute-rhel6.5-mic-6720-19 1 /var/lib/systemimager/images/compute-rhel6.5-mic-6720-19
```

10. Use the `cinstallman` command, in the following format, to install the `sgi-mic-compute` package and the MPSS RPMs:

```
cinstallman --yum-image --image image_name install sgi-mic-compute
```

For *image\_name*, specify the *new\_image* that you created in the following step:

Procedure 3-15, step 8 on page 70

For example:

```
# cinstallman --yum-image --image compute-rhel6.5-mic-6720-19 install sgi-mic-compute
```

11. (Conditional) Enable the SLES operating system to load the MPSS package.

Perform this step if you are installing the MPSS packages on a SLES platform.

Complete the following steps:

- Open file `/etc/modprobe.d/unsupported-modules` with a text editor.
- Add the following line at the end of the file:

```
allow_unsupported_modules 1
```

- Save and close the file.

12. Type the following command to display all the images:

```
# cimage --list-images
image: compute-rhel6.5
      kernel: 2.6.32-358.el6.x86_64
image: compute-rhel6.5-mic-6720-19
      kernel: 2.6.32-358.el6.x86_64
```

The preceding output shows the newly installed image, `compute-rhel6.5-mic-6720-19`.

13. Use the `cimage` command, in the following format, to set the default image for the compute nodes:

```
cimage --set-default --file_system compute-rhel6.5-mic-6720-19 kernel
```

The variables in this command are as follows:

- For *file\_system*, type either `nfs` or `tmpfs`, according to your site practice.

- For *kernel*, type the kernel associated with the new image, as shown in the output from the `cimage --list-images` command in the following step:

Procedure 3-15, step 12 on page 71

For example:

```
# cimage --set-default --compute-rhel6.5-mic-6720-19 2.6.32-358.el6.x86_64
```

14. Type the following command to confirm that the new image is the default image:

```
# cimage --show-default
compute-rhel6.5-stout7-mic-6720-15 2.6.32-358.el6.x86_64 tmpfs
```

15. Proceed to one of the following:

- If your SGI ICE X system contains MIC devices on the service nodes, proceed to "Creating Images for the Service Nodes That Include MIC Devices" on page 72.
- If your SGI ICE X system does not contain MIC devices on the service nodes, proceed to "Configuring the Switches" on page 76.

## Creating Images for the Service Nodes That Include MIC Devices

The following procedure explains how to create service node images that include MIC device software.

**Procedure 3-16** To create service node images for service nodes with MIC devices

1. On the SAC, use the `mkdir(1)` command, in the following format, to create a directory for the RPM repository:

```
mkdir -p /tftpboot/intel/mpss_repository_directory
```

For *mpss\_repository\_directory*, type a name for the directory that is to contain the MPSS repository. For convenience, make sure to include an identifier for the MPSS release level you downloaded.

For example:

```
# mkdir -p /tftpboot/intel/mpss_u3-2.1.6720-19
```

2. Use operating system commands to copy the RPM files you downloaded to the `/tftpboot/intel/mpss_repository_directory` directory on the SAC node.

For example, use `cp(1)`, `ftp(1)`, `rsync(1)`, `scp(1)`, or another method.

- Use the `crepo` command, in the following format, to specify a custom repository for the MPSS RPMs:

```
crepo --add rpm_repo_directory -custom rpm_repo_name
```

The variables in this command are as follows:

- For `rpm_repo_directory`, specify the full path to the directory that contains the RPM files.
- For `rpm_repo_name`, create a name for the image. You can specify the same name for both `rpm_repo_directory` and `rpm_repo_name`. After the image is built, the `cinstallman --show-images` command returns this label in the **Image Name** column of its output.

For example:

```
# crepo --add /tftpboot/intel/mpss_u3-2.1.6720-19 --custom mpss_u3-2.1.6720-19
```

- Type the following command to confirm that the MPSS image is in the correct repository:

```
# crepo --show
```

- Use the `crepo` command, in the following format, to add the custom repository to the generated RPM list in `/etc/opt/sgi/rpmlists`:

```
crepo --select rpm_repo_name
```

For `rpm_repo_name`, create a name for the image. You can specify the same name for both `rpm_repo_directory` and `rpm_repo_name`. After the image is built, the `cinstallman --show-images` command returns this label in the **Image Name** column of its output. This is the same `rpm_repo_name` that you specified in the following step:

Procedure 3-16, step 3 on page 73

For example, the following command adds the custom repository and displays the content of the repository:

```
# crepo --select mpss_u3-2.1.6720-19
Selecting: mpss_u3-2.1.6720-19
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-lead-rhel6.5.rpmlist
```

Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist

6. Type the following command to confirm that you selected the new repository that contains the MPSS RPMs:

```
# crepo --show
* mpss_u3-2.1.6720-19 : /tftpboot/intel/mpss_u3-2.1.6720-19
* Red-Hat-Enterprise-Linux-6.5 : /tftpboot/distro/rhel6.5
* SGI-MPI-1.7-rhel6 : /tftpboot/sgi/SGI-MPI-1.7-rhel6
* SGI-Management-Center-1.9-rhel6 : /tftpboot/sgi/SGI-Management-Center-1.9-rhel6
* SGI-Foundation-Software-2.9-rhel6 : /tftpboot/sgi/SGI-Foundation-Software-2.9-rhel6
* SGI-Tempo-2.9-rhel6 : /tftpboot/sgi/SGI-Tempo-2.9-rhel6
```

The asterisk character (\*) in column 1 indicates an image that is selected.

7. Type the following command to display the images that are available for installation on the service nodes:

```
# cinstallman --show-images
Image Name          BT Path
lead-rhel6.5        0 /var/lib/systemimager/images/lead-rhel6.5
service-rhel6.5     0 /var/lib/systemimager/images/service-rhel6.5
compute-rhel6.5     1 /var/lib/systemimager/images/compute-rhel6.5
```

8. Use the `cinstallman` command, in the following format, to clone the current operating system image:

```
cinstallman --create-image --clone --source current_image --image new_image
```

The variables in this command are as follows:

- For *current\_image*, type the name of the operating system image you want to use. Choose one that appears in the output from the `cinstallman --show-images` command in the **Image Name** list. For example, `service-rhel6.5`.
- For *new\_image*, type a new name for the operating system image that is to include the MPSS file RPMs. SGI recommends that you include information in the new name that can enable you to identify this image as the operating system image that includes MPSS software. For example, `service-rhel6.5-mic-6720-19` identifies the new image as a RHEL image that contains a revision of the MPSS MIC software.

For example:

```
# cinstallman --create-image --clone --source service-rhel6.5 --image service-rhel6.5-mic-6720-19
```

9. Type the following command to display the images and confirm that the new image appears in the list:

```
# cinstallman --show-images
Image Name          BT Path
lead-rhel6.5        0 /var/lib/systemimager/images/lead-rhel6.5
service-rhel6.5     0 /var/lib/systemimager/images/service-rhel6.5
compute-rhel6.5     1 /var/lib/systemimager/images/compute-rhel6.5
service-rhel6.5-mic-6720-19 1 /var/lib/systemimager/images/service-rhel6.5-mic-6720-19
```

10. Use the `cinstallman` command, in the following format, to install the `sgi-mic-service` package and the MPSS RPMs:

```
cinstallman --yum-image --image image_name install sgi-mic-service
```

For *image\_name*, specify the *new\_image* that you created in the following step:

Procedure 3-16, step 8 on page 74

For example:

```
# cinstallman --yum-image --image service-rhel6.5-mic-6720-19 install sgi-mic-service
```

11. (Conditional) Enable the SLES operating system to load the MPSS package. Perform this step if you are installing the MPSS packages on a SLES platform. Complete the following steps:

- Open file `/etc/modprobe.d/unsupported-modules` with a text editor.
- Add the following line at the end of the file:

```
allow_unsupported_modules 1
```

- Save and close the file.

12. Type the following command to display all the images:

```
# cimage --list-images
image: compute-rhel6.5
      kernel: 2.6.32-358.el6.x86_64
image: compute-rhel6.5-mic-6720-19
```

```
kernel: 2.6.32-358.el6.x86_64
image: service-rhel6.5-mic-6720-19
kernel: 2.6.32-358.el6.x86_64
```

The preceding output shows the newly installed image, `service-rhel6.5-mic-6720-19`.

13. Use the `cimage` command, in the following format, to set the default image for the service nodes:

```
cimage --set-default --file_system service-rhel6.5-mic-6720-19 kernel
```

The variables in this command are as follows:

- For `file_system`, type either `nfs` or `tmpfs`, according to your site practice.
- For `kernel`, type the kernel associated with the new image, as shown in the output from the `cimage --list-images` command in the following step:

Procedure 3-16, step 12 on page 75

For example:

```
# cimage --set-default --service-rhel6.5-mic-6720-19 2.6.32-358.el6.x86_64
```

14. Type the following command to confirm that the new image is the default image:

```
# cimage --show-default
service-rhel6.5-stout7-mic-6720-15 2.6.32-358.el6.x86_64 tmpfs
```

15. Proceed to the following:

"Configuring the Switches" on page 76

## Configuring the Switches

The `discover` command initializes and configures the system components for the SGI ICE X system. The switch configuration procedures explain how to use the `discover` command to configure the SGI ICE system's management switches first. After you configure the management switches, if you have MCells, you configure the cooling equipment on the MCell network's switch ports.

The SGI ICE X configuration requires that the same IP address be assigned to the system's head gateway and to the first management switch, usually `mgmtswitch0`.

The procedures in this topic assume that you want to use the default IP address, which is 172.23.0.254, for both components, but the procedures include example commands that show how to configure an alternate IP address.

Switch configuration can proceed much more quickly if you have a media access control (MAC) file. For new SGI ICE X systems, you can obtain a MAC file from your SGI representative.

The MAC file shows the MAC addresses of the components in your environment. Switch discovery and configuration can complete more quickly if you obtain this file. Without this file, you need to power cycle each switch manually.

The following is an example MAC file:

```
r1lead 00:30:48:9e:f2:59 00:30:48:F2:7E:A2
r2lead 00:25:90:01:4e:3c 00:25:90:01:6c:cc
service0 00:25:90:00:3b:8f 00:25:90:01:6e:9e
mgmtsw0 00:26:f3:c3:7a:40 00:26:f3:c3:7a:40
```

The content of a MAC file is as follows:

Column	Content
1	The component's system ID.
2	For nodes, column 2 contains the MAC address of baseboard management controller (BMC) on the component.  For switches, column 2 contains the MAC address of the first network interface card (NIC), eth0. Switches do not have a BMC.
3	The MAC address of the first network interface card (NIC), eth0. For switches, columns two and three are identical in the MAC file.

The following list explains the procedures you need to follow to configure the switches:

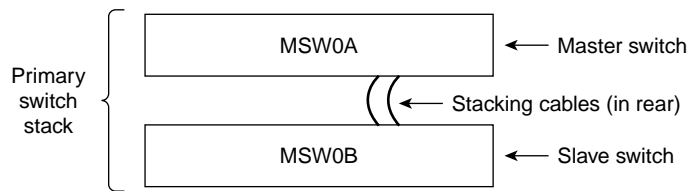
Procedure	Circumstance
"Verifying the Switch Cabling" on page 78	All switch configuration circumstances. Perform this procedure regardless of the presence of a MAC file or MCells.

"Configuring Management Switches With a MAC File" on page 81	If you have a MAC file.
"Configuring Management Switches Without a MAC File" on page 85	If you do not have a MAC file.
"(Conditional) Configuring the Cooling Racks and Cooling Distribution Units (CDUs) on the MCell Network's Switch Ports" on page 88	(Conditional) If you have MCells. This extra procedure configures the MCell switches separately from the rest of the ICE X switches.

### Verifying the Switch Cabling

The following figures show example switch cabling. Depending on the switch configuration procedure you use, you might have to plug cables out from and into switch ports during the configuration process. Regardless of the procedure you use, the cables on your switches eventually become cabled as shown in the following figures.

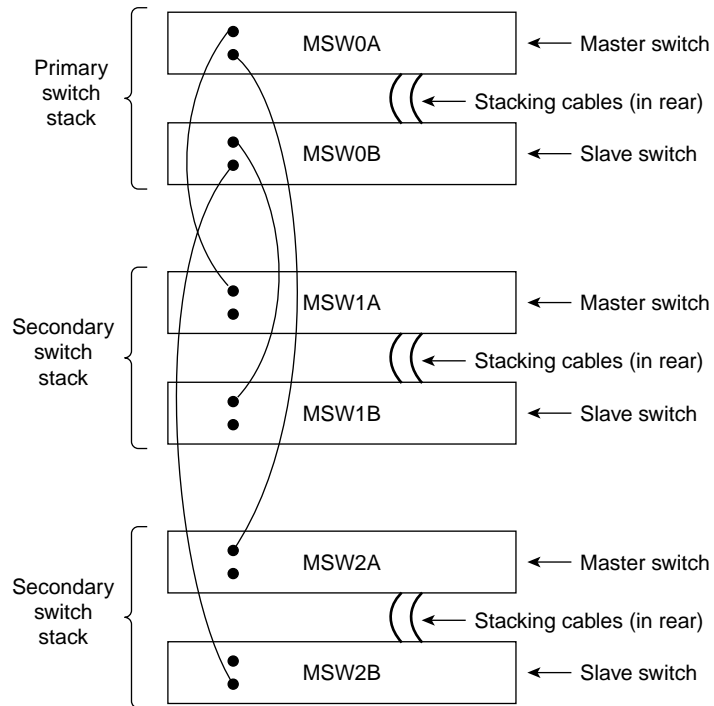
Figure 3-10 on page 78 shows a switch stack with two switches. In this switch stack, the two switches constitute the primary switch stack. One is the master switch and the other is the slave switch.



**Figure 3-10** Primary Switch Stack With Two Switches

Figure 3-11 on page 79 shows a switch stack with multiple switches. The first two switches constitute the primary switch stack, and the other switches constitute the secondary switch stack.





**Figure 3-11** Switch Stack With Multiple Switches

The following procedure explains how to inspect your switches and prepare for the configuration procedure.

**Procedure 3-17** To verify your switches

1. Visually inspect your system.

Note the types of switches you have and their identifiers. Switches are configured in stacks of two switches per stack. In each stack, the top switch is typically the master switch, and the bottom switch is typically the slave switch. At a minimum, you have one primary switch stack. The system admin controller (SAC) connects to the master switch in the primary switch stack. You might have additional secondary switch stacks.

Within each stack, each switch is labeled  $MSW_{xx}$ . In the primary switch stack, the master switch is labeled  $MSW_{0A}$ , and the slave switch is labeled  $MSW_{0B}$ . In the first

secondary switch stack, the master switch is labeled MSW1A, and the slave switch is labeled MSW1B. The A and B on the switch labels identify the master switch and slave switch in the stack. The `switchconfig set` command operates on a switch stack (not just one switch), so you need to note only the characters on the label that precede the A and B when you provide information to the `switchconfig set` command. Use the following table to determine the value you need to use for *switch* on the `switchconfig` commands:

<b>switch</b>	<b>Label</b>
mgmtsw0	MSW0A or MSW0B
mgmtsw1	MSW1A or MSW1B
mgmtsw2	MSW2A or MSW2B
mgmtsw3	MSW3A or MSW3B
mgmtsw4	MSW4A or MSW4B
mgmtsw5	MSW5A or MSW5B
mgmtsw6	MSW6A or MSW6B
mgmtsw7	MSW7A or MSW7B
mgmtsw8	MSW8A or MSW8B
mgmtsw9	MSW9A or MSW9B

2. Make sure that only the SAC is plugged in and that all rack circuit breakers are powered off.

If you have a MAC file, connect all other nodes and switches to a power source, but do not power them on. That is, make sure that all chassis management controllers (CMCs), all rack leader controllers (RLCs), all service nodes, all switches, and so on, are not powered on.

If you do not have a MAC file, unplug all other nodes and switches. The SAC should be the only component that is plugged in.

3. (Conditional) Remove the cascade cables that connect the slaves switches together.

Perform this step only if you have two or more switch stacks.

Locate the cascading cables that connect the switch stacks to each other, and unplug the cascading cable end from the lower switch in the neighboring switch stack. On the front of the switch stack, unplug the cascading cables that plug into

the ports on the slave switches. When you unplug one end of each cable, you prevent a switching loop.

---

**Note:** Do not unplug the stacking cables in the rear of the switch stack. The installation procedure instructs you to plug or unplug only the cascading cables in the front of the switch stack.

---

For example, if you have one secondary switch stack, locate the cascading cable that runs from the slave switch on the primary switch to the slave switch on the secondary switch. Unplug that cable from the port on the slave switch.

4. Proceed as follows:

- If you have a MAC file, proceed to the following:  
"Configuring Management Switches With a MAC File" on page 81
- If you do not have a MAC file, proceed to the following:  
"Configuring Management Switches Without a MAC File" on page 85

## Configuring Management Switches With a MAC File

The following procedure explains how to configure your switches when you have each switch's MAC information in a MAC file.

**Procedure 3-18** To configure switches — with a MAC file

1. Through an `ssh` connection, log in as root to the SAC, and write the MAC file to a location on your SAC.

For example, write it to `/var/tmp/mac_file`.

2. Plug in all the switches.
3. (Conditional) Specify a site-specific IP address for the head gateway.

Perform this step if you need to set a site-specific, nondefault IP address for the primary switch stack (head gateway).

By default, SGI sets the IP address for the primary switch stack (head gateway) to 172.23.0.254. Type the following command to set a site-specific IP address:

```
# cadadmin -set-head-gateway IP_addr
```

For *IP\_addr*, specify your site-specific IP address.

4. Use the `discover` command, in the following format, to configure the primary switch stack (the switches labeled `MSW0A` and `MSW0B`), which is attached to the SAC:

```
# discover --mgmtswitch 0,gateway=yes --macfile path
```

For *path*, type the full path to the location of the MAC file.

This step assigns an IP address to the primary switch stack. The primary switch stack becomes the head gateway.

5. (Conditional) Type additional `discover` commands for each secondary switch stack.

Complete this step if you have more than one switch.

The formats for these additional commands are as follows:

```
# discover --mgmtswitch num --macfile path
```

The arguments are as follows:

<b>Argument</b>	<b>Specification</b>
-----------------	----------------------

<i>num</i>	The identifier for the switch stack.
------------	--------------------------------------

<i>path</i>	The full path to the location of the MAC file.
-------------	--

For example:

```
# discover --mgmtswitch 1 --macfile /var/tmp/mac_file
```

6. (Conditional) On the front of the switch stack, plug the cascading cables into the ports on the slave switches.

Perform this step only if you have two or more switch stacks.

This step is the opposite of the following step, in which you unplugged the cascading cables from the ports on the slave switches:

Procedure 3-17, step 3 on page 80

7. Type the following command to retrieve information about the switches that you discovered, and examine the output for errors:

```
# cnodes --mgmtswitch
```

8. Use the `ssh(1)` command to open a console window to the SAC.
9. Type the following command, and monitor the power-on process in the SAC's console window:

```
# tail -f /var/log/cmcdetectd.log
```

10. Flip the power breakers on the SGI ICE X's CMCs, one rack at a time.

The `cmcdetectd` daemon runs on the SAC. It configures the top level switches so that the CMCs are on the appropriate rack VLAN. After you power on rack one, the `cmcdetectd` daemon detects the rack and adds the rack to the switch. After the CMCs for rack one appear on the switch, power on the CMCs for rack two.

11. Use the `switchconfig` command, in the following format, to list the CMCs associated with each switch:

```
switchconfig list -s switch
```

For *switch*, specify the system ID for a switch, for example `mgmtsw0`, `mgmtsw1`, and so on.

Issue one of these commands for each switch in your configuration and examine the output. If the output includes all the CMCs in your SGI ICE X system, you can proceed with your configuration. If the output contains errors or does not include all expected CMCs, contact your SGI representative for troubleshooting information.

12. Save the configuration to the nonvolatile memory (flash) on the switches.

---

**Note:** This step is very important. In the event of a power outage or other interruption, the switch stack boots with the saved configuration.

---

Type the `switchconfig` command in the following format:

```
# switchconfig save_running_config -s mgmtsw0[,mgmtsw1,mgmtsw2,...]
```

Include the parameters `mgmtsw1`, `mgmtsw2`, and so on, only if there are switches in addition to the primary switch (`mgmtsw0`).

13. Use the `switchconfig` command, in the following format, to back up the switch configuration to a file on the SAC:

```
switchconfig pull_switch_config -s switch_ID -f file [--debug]
```

The arguments are as follows:

Argument	Specification
<i>switch_ID</i>	The switch system ID. For example, this could be <code>mgmtsw0</code> , <code>mgmtsw1</code> , <code>mgmtsw2</code> , and so on. The output from the <code>cnodes</code> command includes this information.
<i>file</i>	The name of the file to receive the switch configuration information. The command writes the file to the <code>/tftpboot/<i>file</i>.cfg</code> . If your <i>file</i> specification ends in <code>.cfg</code> , the command does not append another <code>.cfg</code> string to the file name.

The `--debug` parameter is optional. When specified, the command writes debugging information to `/var/log/switchconfig`.

For example, the following command writes the configuration file for `mgmtsw0` to file `/tftpboot/mgmtsw0_startup1.cfg` on the SAC:

```
switchconfig pull_switch_config -s mgmtsw0 -f mgmtsw0_startup1 --debug
```

Issue one of these commands per switch.

In the future, if you need to replace the switch, you can save configuration time if you push out this configuration file from the SAC to the new switch.

14. After all management switches have been configured, proceed as follows::

- If you have MCells, proceed to the following:

"(Conditional) Configuring the Cooling Racks and Cooling Distribution Units (CDUs) on the MCell Network's Switch Ports" on page 88

- If you do not have MCells, proceed to the following:

"Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the `discover` Command" on page 91

## Configuring Management Switches Without a MAC File

The following procedure explains how to configure your switches when you do not have the switch information in a MAC file.

**Procedure 3-19** To configure switches — without a MAC file

1. (Conditional) Specify a site-specific IP address for the head gateway.

Perform this step if you need to set a site-specific, nondefault IP address for the primary switch stack (head gateway).

By default, SGI sets the IP address for the primary switch stack (head gateway) to 172.23.0.254. Type the following command to set a site-specific IP address:

```
# cadmin -set-head-gateway IP_addr
```

For *IP\_addr*, specify your site-specific IP address.

2. Through an `ssh` connection, log in to the SAC as the root user, and type the following command:

```
# discover --mgmtswitch 0,gateway=yes
```

3. When prompted, connect the primary switch stack to a power source.

To complete this step, plug in the master switch and then the slave switch so that the entire primary switch stack is powered up. In this way, the master switch boots just a few seconds before the slave switch.

The `discover` command configures the MAC address of the switch after you connect the primary switch stack to a power source.

4. (Conditional) Configure each secondary switch stack.

Perform this step if you have secondary switch stacks.

Complete the following steps for each secondary switch stack:

- Plug in the switch stack when prompted by the system.
- Type a `discover` command for the switch in the following format:

```
# discover --mgmtswitch num
```

For *num*, type the identifier for the switch.

For example:

```
# discover --mgmtswitch 1
```

5. (Conditional) On the front of the switch stack, plug the cascading cables into the ports on the slave switches.

Perform this step only if you have secondary switch stacks.

This step is the opposite of the following step, in which you unplugged the cascading cables from the ports on the slave switches:

Procedure 3-17, step 3 on page 80

6. Type the following command to retrieve information about the switches that you configured, and examine the output for errors:

```
# cnodes --mgmtswitch
```

7. Use the `ssh(1)` command to open a console window to the SAC.
8. Type the following command, and monitor the power-on process in the SAC's console window:

```
# tail -f /var/log/cmcdetectd.log
```

9. Flip the power breakers on the SGI ICE X's CMCs, one rack at a time.

The `cmcdetectd` daemon runs on the SAC. It configures the top level switches so that the CMCs are on the appropriate rack VLAN. After you power on rack one, the `cmcdetectd` daemon detects the rack and adds the rack to the switch. After the CMCs for rack one appear on the switch, power on the CMCs for rack two.

10. Use the `switchconfig` command, in the following format, to list the CMCs associated with each switch:

```
switchconfig list -s switch
```

For *switch*, specify the system ID for a switch, for example `mgmtsw0`, `mgmtsw1`, and so on.

Issue one of these commands for each switch in your configuration and examine the output. If the output includes all the CMCs in your SGI ICE X system, you can proceed with your configuration. If the output contains errors or does not



include all expected CMCs, contact your SGI representative for troubleshooting information.

11. Save the configuration to the nonvolatile memory (flash) on the switches.

---

**Note:** This step is very important. In the event of a power outage or other interruption, the switch stack boots with the saved configuration.

---

Type the `switchconfig` command in the following format:

```
# switchconfig save_running_config -s mgmtsw0[,mgmtsw1,mgmtsw2,...]
```

Include the parameters `mgmtsw1`, `mgmtsw2`, and so on, only if there are switches in addition to the primary switch (`mgmtsw0`).

12. For each switch you configured, use the `switchconfig` command, in the following format, to back up the switch configuration to a file on the SAC:

```
switchconfig pull_switch_config -s switch_ID -f file [--debug]
```

The arguments are as follows:

Argument	Specification
<i>switch_ID</i>	The switch system ID. For example, this could be <code>mgmtsw0</code> , <code>mgmtsw1</code> , <code>mgmtsw2</code> , and so on. The output from the <code>cnodes</code> command includes this information.
<i>file</i>	The name of the file to receive the switch configuration information. The command writes the file to the <code>/tftpboot/<i>file</i>.cfg</code> . If your <i>file</i> specification ends in <code>.cfg</code> , the command does not append another <code>.cfg</code> string to the file name.

The `--debug` parameter is optional. When specified, the command writes debugging information to `/var/log/switchconfig`.

For example, the following command writes the configuration file for `mgmtsw0` to file `/tftpboot/mgmtsw0_startup1.cfg` on the SAC:

```
switchconfig pull_switch_config -s mgmtsw0 -f mgmtsw0_startup1 --debug
```

Issue one of these commands per switch.

In the future, if you need to replace a switch, you can save configuration time if you push out this configuration file from the SAC to the new switch.

13. Use the `discover` command, in the following format, to save the MAC addresses to a MAC file:

```
# discover --show-macfile > path
```

For *path*, type the full path to the location of the MAC file. For example, `/var/tmp/mac_file`.

You can use this MAC file if you ever have to configure the switches again.

14. After all management switches have been configured, proceed as follows:
  - If you have MCells, proceed to the following:

"(Conditional) Configuring the Cooling Racks and Cooling Distribution Units (CDUs) on the MCell Network's Switch Ports" on page 88
  - If you do not have MCells, proceed to the following:

"Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the `discover` Command" on page 91

### **(Conditional) Configuring the Cooling Racks and Cooling Distribution Units (CDUs) on the MCell Network's Switch Ports**

Perform the procedure in this topic if you have an SGI ICE X system that includes MCells.

An SGI ICE X system contains CDUs and cooling rack controllers (CRCs). The CDUs and CRCs have statically assigned IP addresses. These IP addresses are critical to associating the individual rack units (IRUs) with specific CDUs or CRCs. For information about these IP addresses, see the following:

Appendix C, "MCell Network IP Addresses" on page 161

The following procedure explains how to configure the switches attached to the MCells.

**Procedure 3-20** To configure MCell switches

1. Gather information about the MCell switches in your SGI ICE X system.

Visually inspect your system. Note the switches identifiers, and note the port identifiers.
2. Log in as the root user to the system admin controller (SAC).

3. Type following command to retrieve information about the virtual local area networks (VLANs) that are configured at this time:

```
# cattr list -g mcell_vlan
global
  mcell_vlan          : 3
```

The preceding output shows that the MCell VLAN is VLAN 3.

4. Use the `switchconfig set` command, in the following format, to configure the ports on which the CDUs and the CRCs are connected to the MCells:

```
switchconfig set -v num=vlan_num -b=none -d vlan_num -p ports -s switch
```

Type an individual `switchconfig set` command for each switch on the SGI ICE X system network.

The arguments are as follows:

Argument	Specification
<i>vlan_num</i>	The VLAN number of the MCell network. For <i>vlan_num</i> , use the output from the <code>cattr list</code> command as shown earlier in this procedure. The default is 3, and SGI recommends that you do not change this value. This argument appears in two places in the <code>switchconfig</code> command.
<i>ports</i>	Specify the target ports. The command configures both the target ports and the corresponding redundant ports.
<i>switch</i>	The ID number of the management switch to which the CDU or CDC is attached. For example: <code>mgmtsw0</code> .

To determine this value, you need to visually inspect the switch, as follows:

- Locate each CDU or CDC. The following are example labels for CDUs: DU01, DU02, and so on.
- Follow the cable that connects each CDU or CDC to a switch. The following is an example label for a cable that connects each CDU to a switch: DU01-LAN1 | 101MSW0A-36.

- Note the label on the switch. Make sure that the labels on the cables correspond to the labels on the switch ports.

**Example 1.** The following command configures VLAN 3 on management switch 0 for target ports 1/31, 1/32, and 1/33 and for redundant ports 2/31, 2/32, and 2/33:

```
switchconfig set -v num=3 -b=none -d 3 -p 1/31,1/32,1/33 -s mgmtsw0
```

**Example 2.** The following command configures VLAN 3 on management switch 0 for target ports 2/31, 2/32, and 2/33 and for redundant ports 1/31, 1/32, and 1/33:

```
switchconfig set -v num=3 -b=none -d 3 -p 2/31,2/32,2/33 -s mgmtsw0
```

---

**Note:** If you make a mistake in your configuration, you can disable the ports from the VLANs you configured. The following example command removes the configuration of VLAN 3 from the target ports and the redundant ports:

```
switchconfig unset -p 1/31,1/32,1/33 -s mgmtsw0 -v num=3
```

- 
5. Repeat the following step for each CDU and each CRC attached to your system:

Procedure 3-20, step 4 on page 89

If you encounter errors, issue a `switchconfig set` command again.

6. Save the configuration to the nonvolatile memory (flash) on the switches.

---

**Note:** This step is very important. In the event of a power outage or other interruption, the switch stack boots with the saved configuration.

---

Type the `switchconfig` command in the following format:

```
switchconfig save_running_config -s mgmtsw0[,mgmtsw1,mgmtsw2,...]
```

Include the parameters `mgmtsw1`, `mgmtsw2`, and so on, only if there are switches in addition to the primary switch (`mgmtsw0`).

7. Type the following command to back up the switch configuration to a file on the SAC:

```
switchconfig pull_switch_config -s switch_ID -f file
```

The arguments are as follows:

Argument	Specification
<i>switch_ID</i>	The switch system ID. For example, this could be <code>mgmtsw0</code> , <code>mgmtsw1</code> , <code>mgmtsw2</code> , and so on. The output from the <code>cnodes</code> command includes this information.
<i>file</i>	The name of the file to receive the switch configuration information. The command writes the file to the <code>/tftpboot/<i>file</i>.cfg</code> . If your <i>file</i> specification ends in <code>.cfg</code> , the command does not append another <code>.cfg</code> string to the file name.

For example, the following command writes the configuration file for `mgmtsw0` to file `/tftpboot/mgmtsw0_startup1.cfg` on the SAC:

```
switchconfig pull_switch_config -s mgmtsw0 -f mgmtsw0_startup1 [--debug]
```

The `--debug` parameter is optional. When specified, the command writes debugging information to `/var/log/switchconfig`.

In the future, if you need to replace the switch, you can save configuration time if you push out this configuration file from the SAC to the new switch.

8. After all switches have been discovered, proceed to the following:

"Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the `discover` Command" on page 91

## Configuring the Rack Leader Controllers (RLCs) and Service Nodes with the `discover` Command

The `discover` command finds and configures rack leader controllers (RLCs), service nodes, and external switches. If you have a media access control file (MAC file), this procedure can complete more quickly. The procedure includes configuration steps that explain how to complete the procedure both with and without a MAC file.

The following procedure explains how to configure the RLCs and service nodes.

**Procedure 3-21** To configure the RLCs and service nodes

1. Visually inspect your SGI ICE X system and note the labels on the RLCs and service nodes.

RLCs are numbered starting with 1.

Service nodes are numbered starting with 0.

2. Check the power cords on all RLCs and services nodes, as follows:
  - If you have a MAC file to specify as input to the `discover` command, make sure all RLCs and service nodes are plugged in.

Do not power-on the nodes at this time. When the node is plugged in and connected to a power source, the baseboard management controller (BMC) is started, and that is all that is required at this time.

- If you do not have a MAC file, make sure that all RLCs and service nodes are unplugged from their power sources.
3. Through an `ssh` connection, log into the system admin controller (SAC) as the root user.
  4. Type the following command to retrieve the option code that is in use:

```
# cadmin --show-dhcp-option
```

The RLCs and service nodes determine the integrated Ethernet devices by accepting DHCP leases that belong only to the SGI ICE X system. SGI ICE X systems use DHCP option code 149 by default. In rare situations, a house network DHCP server could be configured to use this option code. In this case, nodes that are connected to the house network can mistake a house DHCP server as belonging to the SGI ICE X DHCP server, which can lead to an installation failure. Change this option code only if absolutely necessary.

To change the `dhcp` option code number used for this operation, type a command such as the following:

```
# cadmin --set-dhcp-option 150
```

This command sets the DHCP option code to 150.

5. (Conditional) Plug in all the racks and service nodes.

Perform this step if you plan to specify a MAC file as input to the `discover` command.

A MAC file contains the MAC addresses for RLCs and service nodes. If you use a MAC file, the configuration process can complete more quickly. Contact your SGI

representative to find out if a MAC file is available. For more information about MAC files, see "Configuring the Switches" on page 76.

6. Use one or more `discover` commands, in the following format, to configure the RLCs and service nodes on your SGI ICE X system:

```
discover --rack[set] count[,mic=mic_num] --service[set] count[,mic=mic_num] [--macfile macfile]
```

The arguments are as follows:

Argument	Specification
----------	---------------

<i>count</i>	<p>Specify the ID number for the rack (or racks) or service node (or service nodes) that you want to configure.</p> <p>If you want to configure one rack or one service node, specify <code>--rack</code> or <code>--service</code> and the system ID number that corresponds to that rack or service node.</p> <p>If you want to configure a range of racks or a range of service nodes, specify <code>--rackset</code> or <code>--serviceset</code>, the starting system ID number, a comma (,), and the ending system ID number.</p> <p>Examples:</p> <pre>--rack 2 --rackset 1,3 --service0 --service1</pre>
<i>mic_num</i>	<p>Specify the number of Intel® Many Integrated Core (MIC) devices that reside on each blade or service node. By default, the <code>discover</code> command assumes zero (0). If you have MIC devices on your compute nodes, specify the number you have, which can be 1 or 2. For service nodes, <i>mic_num</i> can be 1, 2, 3, or 4.</p> <p>Specify the <code>,mic=<i>mic_num</i></code> parameter only if your SGI ICE X system includes MIC devices.</p>
<i>macfile</i>	<p>Specify the full path to the media access control (MAC) input file. Specify this option if you have a MAC file.</p>

To retrieve more information about the `discover` command, type `discover --h`.

Example 1. The following command uses a MAC file to configure rack 1 and service node 0:

```
# discover --rack 1 --service 0 --macfile mymacfile
```

Example 2. If you have one rack and one service node, type the following command:

```
# discover --rack 1 --service 0
```

Example 3. If you have five racks and three service nodes, type the following command:

```
# discover --rackset 1,5 --serviceset 1,3
```

Example 4. If you have one rack, one service node, two MIC devices attached to each blade on the rack, and two MIC devices attached to the service node, type the following command:

```
# discover --rack 1,mic=2 --service 0,mic=2,image=mic_serv_image
```

In example 4, for *mic\_serv\_image*, specify the label (name) of the service node image that you created in "Creating Images for the Service Nodes That Include MIC Devices" on page 72.

7. (Conditional) When prompted to do so by the system, plug in each individual rack or service node.

Perform this step if you did not use a MAC file as input to the `discover` command.

The system prompt for this action is as follows:

```
At this time, please turn on the power to this service node.  
Do not turn the system on.
```

The blue light on each component turns on when configuration is complete.

You can use the `console(1)` command if you want to watch the installation progress. The sessions are also logged.

8. Type the following commands to update the configuration files:

```
# update-configs  
# update-switch-routing
```



9. Type the following commands to save the configuration to the nonvolatile memory (NVM) on the switches:

```
# switchconfig save_running_config -s mgmtsw0[,mgmtsw1,mgmtsw2,...]
```

Include the parameters `mgmtsw1`, `mgmtsw2`, and so on, only if there are switches in addition to management switch 0.

10. (Conditional) Confirm the status of the MIC devices.

Perform this step if your SGI ICE X system has MIC devices.

Complete the following steps:

- Type the following command to make sure that the MIC devices came online:

```
# cexec --all 'micctrl --status'
```

The following example output shows that the MIC devices came online correctly:

```
***** rack_1 *****
***** rack_1 *****
----- r1i0n6-----
mic0: online (mode: linux image: /lib/firmware/mic/uos.img)
mic1: online (mode: linux image: /lib/firmware/mic/uos.img)
----- r1i0n7-----
mic0: online (mode: linux image: /lib/firmware/mic/uos.img)
mic1: online (mode: linux image: /lib/firmware/mic/uos.img)
----- r1i0n8-----
mic0: online (mode: linux image: /lib/firmware/mic/uos.img)
mic1: online (mode: linux image: /lib/firmware/mic/uos.img)
```

- Type the following command to verify the IP addresses and MTU size of the MIC devices:

```
# cexec --all 'micctrl --config | grep -E \"MIC IP|Bits|Mtusize\"'
```

The following example output shows that all the MIC devices have addresses on the 10.157.1.0/24 and the 10.158.1.0/24 networks and that the MIC devices use the correct MTU size of 9000:

```
***** rack_1 *****
***** rack_1 *****
----- r1i0n4-----
```

```

MIC IP:    10.157.1.6
Net Bits:  24
MtuSize:   9000
MIC IP:    10.158.1.6
Net Bits:  24
MtuSize:   9000
----- rli0n5-----
MIC IP:    10.157.1.7
Net Bits:  24
MtuSize:   9000
MIC IP:    10.158.1.7
Net Bits:  24
MtuSize:   9000
----- rli0n6-----
MIC IP:    10.157.1.8
Net Bits:  24
MtuSize:   9000
MIC IP:    10.158.1.8
Net Bits:  24
MtuSize:   9000
----- rli0n7-----
MIC IP:    10.157.1.9
Net Bits:  24
MtuSize:   9000
MIC IP:    10.158.1.9
Net Bits:  24
MtuSize:   9000
----- rli0n8-----
MIC IP:    10.157.1.10
Net Bits:  24
MtuSize:   9000
MIC IP:    10.158.1.10
Net Bits:  24
MtuSize:   9000
```

11. Proceed as follows:

- If you want to configure a backup domain name service (DNS) server, proceed to the following:

"(Optional) Configuring a Backup Domain Name Service (DNS) Server" on page 97

- To configure the InfiniBand subnetworks, proceed to the following:  
"Configuring the InfiniBand Subnetworks" on page 98

## (Optional) Configuring a Backup Domain Name Service (DNS) Server

Typically, the DNS on the system admin controller (SAC) provides name services for the SGI ICE X system. When you configure a backup DNS, however, the compute nodes can use a service node as a secondary DNS server if the SAC is down, being serviced, or is otherwise not available. You can configure a backup DNS only after you run the `discover` command to configure the cluster. This is an optional feature.

The following procedure explains how to configure a service node to act as a DNS.

**Procedure 3-22** To enable a backup DNS

1. Through an `ssh` connection, log into the system admin controller (SAC) as the root user.
2. Type the following command to retrieve a list of available service nodes:

```
# cnodes --service
```

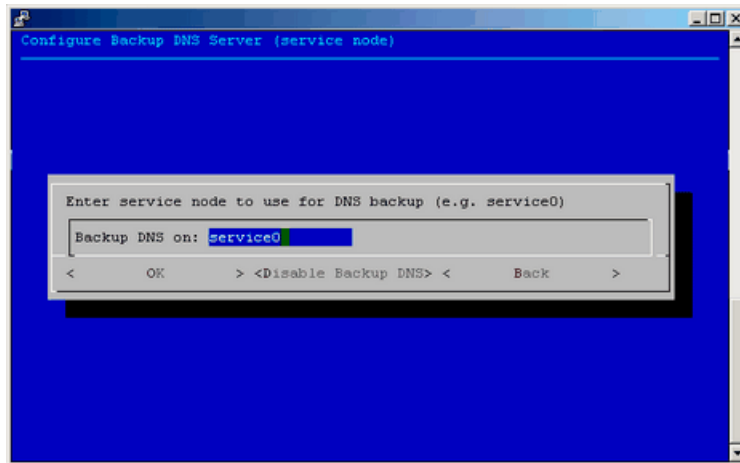
The service node you want to use as a backup DNS must be configured in the system already. That is, you must have run the `discover` command to configure the service nodes.

3. Type the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

4. On the **Main Menu** screen, select **B Configure Backup DNS Server (optional)**, and select **OK**.
5. On screen that appears, type the identifier for the service node that you want to designate as the backup DNS, and select **OK**.

Figure 3-12 on page 98 shows how to specify `service0` as the backup DNS.



**Figure 3-12 Configure Backup DNS Server (service node) screen**

To disable this feature, select **Disable Backup DNS** from the same menu and select **Yes** to confirm your choice.

## Configuring the InfiniBand Subnetworks

The InfiniBand network on SGI ICE X systems uses Open Fabrics Enterprise Distribution (OFED) software. For information about OFED, see the following website:

<http://www.openfabrics.org>

For more information about the InfiniBand fabric implementation on SGI ICE X systems, see the *SGI ICE X Administration Guide*.

Each SGI ICE X system has two InfiniBand fabric network cards, `ib0` and `ib1`. Each subnetwork has a subnet manager, which runs on a rack leader controller (RLC). The following procedure explains how to specify the RLC that you want to configure as the master and the RLC that you want to configure as the stand-by.

**Procedure 3-23** To configure the InfiniBand network

1. Through an `ssh` connection, log into the system admin controller (SAC) as the root user.

2. Type the following command to disable InfiniBand switch monitoring:

```
% catrr set disableIbSwitchMonitoring true
```

The system sometimes issues InfiniBand switch monitoring errors before the InfiniBand network has been fully configured. The preceding command disables InfiniBand switch monitoring.

3. Use one of the following methods to access the InfiniBand network configuration tool:

- Type the following command to start the cluster configuration tool:

```
# /opt/sgi/sbin/configure-cluster
```

After the cluster configuration tool starts, select **F Configure InfiniBand Fabric**, and select **OK**.

- Type the following command to start the InfiniBand management tool:

```
# tempo-configure-fabric
```

Both of the preceding methods lead you to the same InfiniBand configuration page. On the InfiniBand configuration pages, **Quit** takes you to the previous screen.

4. Select **A Configure InfiniBand ib0**, and select **Select**.
5. On the **Configure InfiniBand** screen, select **A Configure Topology**, and select **Select**.
6. On the **Topology** screen, select the topology your system uses, and select **Select**.

The menu selections are as follows:

- **H HYPERCUBE**
- **E EHYPERCUBE** (Enhanced Hypercube)
- **F FAT TREE**
- **G BFTREE**

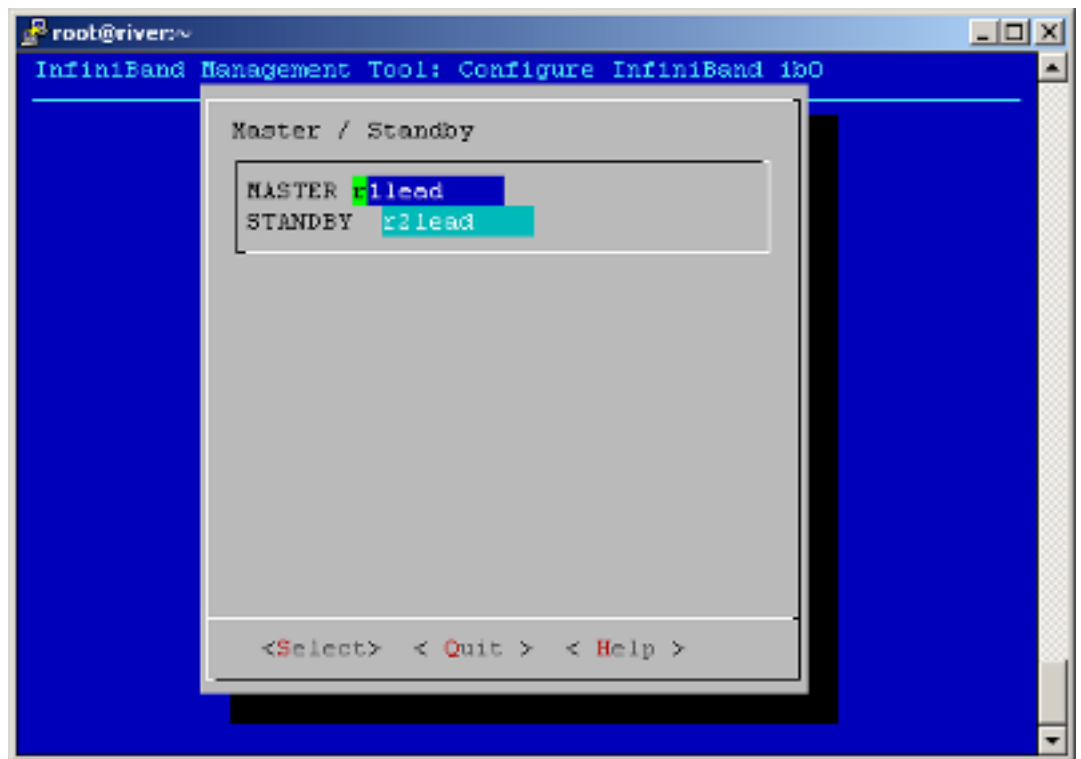
7. On the **Configure InfiniBand** screen, select **B Master / Standby**, and select **Select**.

8. On the **Master / Standby** screen, type the RLC identifiers for the master (primary) and the standby (backup, secondary) subnetwork, and select **Select**.

If you have only one rack leader controller (RLC), type `r1lead` in the **MASTER** field, and leave the **STANDBY** field blank.

If you have more than one RLC, specify different RLCs in the **MASTER** and **STANDBY** fields.

For example, Figure 3-13 on page 100 shows a completed screen.



**Figure 3-13** Completed InfiniBand (ib0) Master / Standby Screen

9. On the **Configure InfiniBand** screen, select **Commit**.

Wait for the confirmatory messages to appear in the window before you continue.

The next few steps repeat the preceding steps, but this time you configure the `ib1` interface.

10. On the InfiniBand Management Tool main menu screen, select **B Configure InfiniBand ib1**, and select **Select**.
11. On the **Configure InfiniBand** screen, select **A Configure Topology**, and select **Select**.
12. On the **Topology** screen, select the topology your system uses, and select **Select**.  
Select the topology that exists on your system. The menu selections are as follows:
  - **H HYPERCUBE**
  - **E EHYPERCUBE** (Enhanced Hypercube)
  - **F FAT TREE**
  - **G BFTREE**
13. On the **Configure InfiniBand** screen, select **B Master / Standby**, and select **Select**.
14. On the **Master / Standby** screen, type the RLC identifiers for the master (primary) and the standby (backup, secondary) subnetwork, and select **Select**.  
  
If you have only one rack leader controller (RLC), type `r1lead` in the **MASTER** field, and leave the **STANDBY** field blank.  
  
If you have two RLCs, you can flip the specifications for `ib1`. For example, assume that for `ib0`, you specified **MASTER** as `r1lead` and **STANDBY** as `r2lead`. For `ib1`, you can specify **MASTER** as `r2lead` and **STANDBY** as `r1lead`.  
  
If you have three or more RLCs, specify different RLCs in the **MASTER** and **STANDBY** fields.
15. On the **Configure InfiniBand** screen, select **Commit**.  
Wait for the confirmatory messages to appear in the window before you continue.
16. On the InfiniBand Management Tool main menu screen, select **C Administer Infiniband ib0**, and select **Select**.
17. On the **Administer InfiniBand** screen, select **Start**, and select **Select**.
18. On the **Start SM master\_ib0 on ib0 succeeded!** screen, select **OK**.

19. Select **Quit** to return to the InfiniBand Management Tool main menu screen.  
The next few steps repeat the preceding steps, but this time you start the `ib1` interface.
20. On the InfiniBand Management Tool main menu screen, select **D Administer Infiniband ib1**, and select **Select**.
21. On the **Administer InfiniBand** screen, select **Start**, and select **Select**.
22. On the **Start SM master\_ib1 on ib1 succeeded!** screen, select **OK**.
23. On the **Administer InfiniBand** screen, select **Status**, and select **Select**.

The **Status** option returns information similar to the following:

```
Master SM
Host = r1lead
Guid = 0x0002c9030006938b
Fabric = ib0
Topology = hypercube
Routing Engine = dor
OpenSM = running
```

24. Wait for the status messages to stop, and press `Enter`.
25. Select **Quit** on the menus that follow to exit the configuration tool.
26. Type the following command to retrieve the list of rack leader controller (RLC) IDs:

```
SAC:~ # cnodes --leader
r1lead
r2lead
```

In the next few steps, you verify that the InfiniBand network is working.

27. Through an `ssh(1)` connection, log into a leader node.

For example:

```
SAC:~ # ssh r2lead
```

28. Type the following command to retrieve the IDs of the compute nodes (blades):

```
r2lead:~ # cnodes --compute
r2i0n0
```



```
r2i0n1  
r2i0n2  
r2i0n3
```

29. Type a `ping(8)` command to make sure that the RLC can reach its compute nodes.

For example:

```
r2lead:~ # ping -c1 r2i0n0
```

If the `ping(8)` is successful, the InfiniBand network is configured properly.

30. Type the following command to reenale InfiniBand switch monitoring:

```
% catrr unset disableIbSwitchMonitoring
```

31. (Optional) Configure additional features.

The SGI ICE X system supports several optional features, for example, networking features such as network address translation. For information about how to configure optional features, see the following:

Chapter 4, "Configuring Additional Features" on page 105



## Configuring Additional Features

This chapter includes the following topics:

- "Enabling Hardware Event Tracker (HET) Notifications" on page 105
- "CPU Frequency Scaling" on page 109
- "Installing MPI on a Running SGI ICE X System" on page 114

### Enabling Hardware Event Tracker (HET) Notifications

The following topics contain information about HET:

- "About HET" on page 105
- "Customizing HET Notifications" on page 106
- "HET Examples" on page 108

---

**Note:** This documentation includes information about all SGI systems for the sake of completeness.

---

### About HET

All of your SGI system's IPMI controllers send SNMP traps to the management node, which is the system admin controller (SAC) on an SGI ICE X system or the system management node (SMN) on an SGI UV system. The SGI Foundation Software's HET tools process these system alerts and send an email notification after critical hardware events occur.

On SGI ICE X and SGI UV platforms, the HET tools are configured by default. You do not need to perform any additional system configuration, but SGI recommends that you customize the email address to which the HET tools send critical event notifications. On SGI Rackable systems, you need to configure the system manually as noted in the `het(8)` man page. The `het(8)` man page also contains information about HET defaults and internal processes.

HET accumulates information about system events in the following log file:

```
/var/log/het/het_trap_processor.log
```

As an event-driven system monitoring tool, HET listens for system events. When HET receives information about an event, it converts the message from coded numbers into a readable form, as follows:

- When a noncritical event occurs, HET simply logs the event. As an option, you can configure an email address to receive noncritical event notifications.
- When a critical event occurs, HET logs the event and send an email message. SGI recommends that you edit file `/etc/sysconfig/het` and specify an email address specific to your site. By default, HET sends event information to `root@localhost`. For more information about how to customize HET notifications, see "Customizing HET Notifications" on page 106.

The firmware for each baseboard management controller (BMC) and the firmware for each cooling node on an SGI ICE X system includes threshold values for each component. If a system condition becomes too low or too high for its threshold, the BMC sends a critical event alert. The following are examples of critical system events that cause an alert:

- Ambient air temperature outside of recommended range
- Voltage sensor unable to attain a critical low voltage
- Power supply failure
- Loss of redundant power supply
- Fan speed unable to attain a critical threshold or a loss of fan redundancy
- Board processor modules that exceed a critical temperature threshold
- Memory uncorrectable errors

## Customizing HET Notifications

You can customize the email addresses to which event information about `NON-RECOVERABLE` events is sent. As an option, you can specify a site-specific email address for less-severe events, or all HET events, too.

The HET log file, `/var/log/het/het_trap_processor.log`, contains information about all HET events. You can consult this file periodically to monitor noncritical events.

The following procedure explains how to configure an email address or email alias to receive HET notifications.

**Procedure 4-1** To customize HET notifications

1. Log in as root and open the following file with a text editor:

```
/etc/sysconfig/het
```

On an SGI ICE X system, log into the system admin controller (SAC) node.

On an SGI UV system, log into the system management node (SMN). HET requires an SMN. If your system does not include an SMN, you cannot enable HET.

2. Search the file for the following string:

```
HET_MAIL_TROUBLE_TO
```

3. Change the default recipient, `root`, to be the email address of a person or the email alias of a group who can attend to the system when `NON-RECOVERABLE` events occur.

4. (Optional) Configure an email recipient for notifications about `CRITICAL` events.

Search the file for the following string:

```
HET_MAIL_NEWS_TO
```

Specify an email address or alias to receive `CRITICAL` event notifications.

5. Save and close file `/etc/sysconfig/het`.
6. (Optional) Configure an email recipient for all HET events.

Complete the following steps:

- Open file `/etc/het.action.d/het_mail` with a text editor.
- Search for the following lines in `/etc/het.action.d/het_mail`:

```
# NOTE: Adjust if needed
# Default is an empty mailing list audience for
# non (NON-RECOVERABLE or CRITICAL) events.
```

```
to= " "
```

- Edit the `to= " "` line to specify an email address or an email alias between the quotation marks.
- Save and close the file.

## HET Examples

The following is an example of a HET log file that contains critical information:

```
dump      2013-10-23.07.13.21 [het_process_thread:2] # begin -----
dump      2013-10-23.07.13.21 [het_process_thread:2] agentAddr      172.24.0.2
dump      2013-10-23.07.13.21 [het_process_thread:2] het_type        ipmi
dump      2013-10-23.07.13.21 [het_process_thread:2] guid            r1lead
dump      2013-10-23.07.13.21 [het_process_thread:2] sn              X1-----
dump      2013-10-23.07.13.21 [het_process_thread:2] alertSeverity   NONE
dump      2013-10-23.07.13.21 [het_process_thread:2] event           uncorrectableECC
dump      2013-10-23.07.13.21 [het_process_thread:2] sensorName      None-memory
dump      2013-10-23.07.13.21 [het_process_thread:2] sensorNumber    0x00
dump      2013-10-23.07.13.21 [het_process_thread:2] sensorTypeName  memory
dump      2013-10-23.07.13.21 [het_process_thread:2] eventClassName  discrete
dump      2013-10-23.07.13.21 [het_process_thread:2] event1          0x51
dump      2013-10-23.07.13.21 [het_process_thread:2] event2          0xff
dump      2013-10-23.07.13.21 [het_process_thread:2] event3          0x51
dump      2013-10-23.07.13.21 [het_process_thread:2] flap_count      1
dump      2013-10-23.07.13.21 [het_process_thread:2] # end -----
```

The corresponding email message that HET sends is as follows:

```
X-Original-To: root
Delivered-To: root@saturn9-1.americas.sgi.com
Date: Wed, 18 Dec 2013 14:36:52 -0600
From: HET.ALERT.donotreply@saturn9-1.americas.sgi.com
To: root@saturn9-1.americas.sgi.com
Subject: HET ALERT from cb9 - NON-RECOVERABLE
User-Agent: Heirloom mailx 12.2 01/07/07
```

The following HET(Hardware Environment Tracking) event has been recorded:  
HET ALERT from cb9 - NON-RECOVERABLE

## Event Details:

EVENT	uncorrectableECC
HET	r1i0n4
LOCATION	r1i0n4
SENSOR	None-memory
SENSORNUMBER	0x00
SENSORTHRESHOLD	81
SENSORTYPE	memory
SENSORVALUE	255
SEVERITY	NON-RECOVERABLE
SN	X1-----
TYPE	ipmi

## CPU Frequency Scaling

CPU frequency scaling allows the operating system to automatically and dynamically scale the processor frequency. CPU frequency scaling needs to be enabled in a compute node image if you want to take advantage of the Intel Turbo Boost technology that is built into each processor.

The Intel Turbo Boost Technology allows processor cores to run faster than the base operating frequency as long as they are operating below the limits set for power, current, and temperature. The CPU frequency scaling setting also affects power consumption and enables you to manage power consumption. For example, you can theoretically cut power consumption in half if you clock the computer's processors from 2 GHz down to 1 GHz.

The following procedures pertain to CPU frequency:

- "Enabling or Disabling CPU Frequency Scaling" on page 109
- "(Optional) Changing the Governor Setting and Configuring Turbo Mode" on page 111

### Enabling or Disabling CPU Frequency Scaling

The procedure in this topic explains how to enable or disable CPU frequency scaling. CPU frequency scaling is disabled by default on SGI ICE X systems.

The following procedure explains how to change your CPU frequency scaling setting.

**Procedure 4-2** To control CPU frequency scaling

1. Log into the system admin controller (SAC) as `root`.
2. Use the `cimage --list-images` command to retrieve a list of the compute node images you can edit:

For example:

```
# cimage --list-images
image: compute-sles11sp3.mpt
      kernel: 3.0.76-0.11-default
      kernel: 3.0.76-0.11-trace
image: compute-sles11sp3
      kernel: 3.0.76-0.11-default
```

The previous example shows the names of two images:  
`compute-sles11sp3.mpt` and `compute-sles11sp3`.

3. Type the following command to change to the directory that contains the image you want to edit:

```
# chroot /var/lib/systemimager/images/image_name
```

For *image\_name*, specify one of the compute node image names that the `cimage` command returned. For example, using the output from the preceding step, specify either `compute-sles11sp3.mpt` or `compute-sles11sp3`.

4. Use a text editor to open file `/etc/modprobe.d/acpi-cpufreq.conf`.
5. Note the following line in this file:

```
install acpi-cpufreq /bin/true
```

To enable CPU frequency scaling, insert a pound (#) character as the first character in this line, which makes the line appear as follows:

```
#install acpi-cpufreq /bin/true
```

To disable CPU frequency scaling, make sure that the `install acpi-cpufreq /bin/true` line does not contain a # character in column 1, which makes the line appear as follows:

```
install acpi-cpufreq /bin/true
```



6. Save and close file `/etc/modprobe.d/acpi-cpufreq.conf`.
7. Push the changes out to the compute nodes.  
Perform the procedure in the following topic:  
"Pushing Changes to the Compute Nodes" on page 14
8. (Optional) Change the CPU frequency governor setting and configure turbo mode.  
The default governor setting and the default turbo mode setting are appropriate for most SGI ICE X systems. If you want to change these settings, proceed to the following:  
"(Optional) Changing the Governor Setting and Configuring Turbo Mode" on page 111

### (Optional) Changing the Governor Setting and Configuring Turbo Mode

Use the procedure in this topic to change the governor setting and, optionally, to configure turbo mode. When you enable turbo mode, you enable the CPU frequency to exceed its nominal level for short periods of time, depending on the processor, temperature, current, power, and other factors. For general information about turbo mode, see the following website:

<https://www-ssl.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html>

The following procedure explains how to set the CPU frequency governor appropriately and how to configure turbo mode.

**Procedure 4-3** To change the governor setting and configure turbo mode

1. Make sure that CPU frequency is enabled.  
For information, see "Enabling or Disabling CPU Frequency Scaling" on page 109.
2. Examine the following list and choose a power governor setting:

<i><b>governor Setting</b></i>	<b>Effect</b>
ondemand	Dynamically switches between the available CPUs if at 95% of CPU load. Default.
performance	Runs the CPUs at the maximum frequency.

conservative	Dynamically switches between the available CPUs if at 75% of CPU load.
powersave	Runs the CPUs at the minimum frequency.
userspace	Runs the CPUs at user-specified frequencies.

3. Use the `cimage --list-images` command to retrieve a list of the compute node images you can edit:

For example:

```
# cimage --list-images
image: compute-sles11sp3.mpt
      kernel: 3.0.76-0.11-default
      kernel: 3.0.76-0.11-trace
image: compute-sles11sp3
      kernel: 3.0.76-0.11-default
```

The previous example shows the names of two images: `compute-sles11sp3.mpt` and `compute-sles11sp3`.

4. Type the following command to change to the directory that contains the image you want to edit:

```
# chroot /var/lib/systemimager/images/image_name
```

For *image\_name*, specify one of the compute node image names that the `cimage` command returned. For example, using the output from the preceding step, specify either `compute-sles11sp3.mpt` or `compute-sles11sp3`.

5. Use one of the following platform-specific methods to change the setting:
  - On RHEL platforms, complete the following steps:
    1. Open file `/etc/sysconfig/cpuspeed`.
    2. Search for the `GOVERNOR=` string.
    3. Edit the setting, adding the *governor* setting you chose in the previous step.
    4. Save and close the file.
    5. Type the following command:

```
# service cpuspeed restart
```

- On SLES platforms, complete the following steps:

1. Type the following command:

```
# cpupower frequency-set -g governor
```

For *governor*, specify the setting you chose in the previous step.

2. Type the following command:

```
# cpupower frequency-info
```

3. Verify that the *governor* setting you specified appears in the command in the output in the `current policy` field.

4. Use a text editor to edit the `/etc/init.d/after.local` file and add the following line:

```
cpupower frequency-set -g governor
```

The preceding line ensures that after each boot, the system sets the *governor* setting you specified.

6. Push the changes out to the compute nodes.

Perform the procedure in the following topic:

"Pushing Changes to the Compute Nodes" on page 14

7. Use the `cat(1)` command to retrieve the list of available frequencies. For example:

```
# cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_available_frequencies
3301000 3300000 3200000 3100000 3000000 2900000 2800000 2700000 2600000
2500000 2400000 2300000 2200000 2100000 2000000 1900000 1800000 1700000
1600000 1500000 1400000 1300000 1200000
```

The preceding output shows the available frequencies, listed in order from the highest, 3301000 KHz, to the lowest, 1200000 KHz.

On SGI systems, the second frequency listed is always the processor's nominal frequency. This is a 3.3 GHz processor, so 3300000 KHz is the nominal frequency.

You can also obtain the nominal frequency by typing the following command and examining the information in the model name field:

```
# cat /proc/cpuinfo
```

8. Use the `cpupower` command to set the frequency to the nominal frequency of 3.3 GHz plus 1 MHz.

That is, specify a frequency of 3301 MHz. For example:

```
# cpupower frequency-set -u 3301MHz
```

Later, if you want to disable turbo mode, type the following command to set the maximum frequency back to the nominal frequency:

```
# cpupower frequency-set -u 3300MHz
```

## Installing MPI on a Running SGI ICE X System

This topic explains how to install MPI on an SGI ICE X system. The instructions in this section update existing images instead of creating new ones. It should be noted that integrating MPI before cluster deployment is easier.

The SGI® MPI and SGI® Accelerate™ software packages have embedded in them suggested package lists for each node type. The `crepo` command, used in the following example, makes use of these lists and recomputes the lists when new media is added and then selected.

File names in this example are just illustrations.

Register SGI MPI and SGI Accelerate with SGI Tempo, as follows:

```
# crepo --add accelerate-1.0-cd1-media-rhel6-x86_64.iso
# crepo --add mpi-1.0-cd1-media-rhel6-x86_64.iso
```

Update the `crepo` selected repositories so that all repositories associated with the software distribution (distro) you are installing are present. For example, if you want MPI to work on RHEL 6, you might do something like this:

Show what is currently selected (the asterisks to the left):

```
# crepo --show
* SGI-Management-Center-1.7-rhel6 : /tftpboot/sgi/SGI-Management-Center-1.7-rhel6
* SGI-Foundation-Software-2.10-rhel6 : /tftpboot/sgi/SGI-Foundation-Software-2.10-rhel6
* SGI-XFS-XVM-2.5-for-RHEL-rhel6 : /tftpboot/sgi/SGI-XFS-XVM-2.5-for-RHEL-rhel6
* SGI-Accelerate-1.8-rhel6 : /tftpboot/sgi/SGI-Accelerate-1.8-rhel6
* SGI-Tempo-2.9.0-rhel6 : /tftpboot/sgi/SGI-Tempo-2.9.0-rhel6
* SGI-MPI-1.8-rhel6 : /tftpboot/sgi/SGI-MPI-1.8-rhel6
```

```
* Red-Hat-Enterprise-Linux-6.5 : /tftpboot/distro/rhel6.5
```

#### Unselect unrelated repositories:

```
# crepo --unselect SGI-Tempo-2.5-rhel6
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
# crepo --unselect SGI-Foundation-Software-2.5-rhel6
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
# crepo --unselect Red-Hat-Enterprise-Linux-6.5
Removing: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Removing: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
```

#### Select RHEL 6 related repositories:

```
# crepo --select Red-Hat-Enterprise-Linux-6.5
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-lead-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
# crepo --select SGI-Foundation-Software-2.5-rhel6
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-lead-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
# crepo --select SGI-XFS-XVM-2.5-for-RHEL-rhel6
Updating: /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-lead-rhel6.5.rpmlist
Updating: /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmlist
```

After performing the preceding steps, the proper repositories are registered and selected so you can operate on them by default. Since you are using an already deployed system, you need to update existing images and potentially existing service nodes themselves. This example uses SGI suggested/ generated rpmlists. If you have custom rpmlists, you need to manually reconcile the two lists for each node type. The list fragments in `/var/opt/sgi/sgi-repodata/` may help you.

---

**Note:** The commands in this topic include a continuation character (`\`).

---

For a service node image, type the following:

```
# cinstallman --refresh-image --image service-rhel6.5 \  
--rpmfile /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmfile
```

For a compute node image, type the following:

```
# cinstallman --refresh-image --image compute-rhel6.5 \  
--rpmfile /etc/opt/sgi/rpmlists/generated-compute-rhel6.5.rpmfile
```

Finally, you need to push the updated compute image to the rack leader controllers (RLCs).

---

**Note:** If the compute nodes are booted on the image and are using NFS for roots, you need to shut the compute nodes down before being able to run this command.

---

```
# cimage --push-rack compute-rhel6.5 r***
```

To make sure the compute nodes you are operating on have the associated compute image you just updated, perform a command similar to the following”

```
# cimage --set compute-rhel6.5 2.6.32-71.el6.x86_64 ***
```

You can find the available images and kernels using the `cimage --list-images` command.

If you have booted service/login nodes, you likely want to refresh those running nodes also. (You could also reinstall them, as well). Here is a refresh example:

```
# cinstallman --refresh-node --node service0 \  
--rpmfile /etc/opt/sgi/rpmlists/generated-service-rhel6.5.rpmfile
```

Now reset or bring up the nodes (depends on the state you left them). If you want to bring up all nodes, this command will not disrupt nodes already operating:

```
# cpower --system --up
```

## Troubleshooting Configuration Changes

If a configuration change does not affect the SGI ICE X system in the intended manner, try one of the following approaches:

- Edit the node image on the system admin controller (SAC). For example, you can reconfigure the service node image on the SAC and reimage the service nodes with that new image.
- Edit the node customization scripts. For example, the compute node update scripts reside on the SAC in the `/opt/sgi/share/per-host-customization/global` directory.





## Troubleshooting

This chapter covers the following topics:

- "About Troubleshooting" on page 119
- "Using the `switchconfig` Command" on page 120
- "Repairing a Broken DCM Enablement" on page 124
- "Compute Nodes Are Taking Too Long To Boot" on page 125
- "Verify the Bonding Mode on the Rack Leader Controller (RLC)" on page 126
- "`cimage --push-rack` Pushes Too Many (or Too Few) Expansions" on page 129
- "Cannot ping the CMCs from the Rack Leader Controller (RLC)" on page 130
- "Restarting the `blademon` Daemon" on page 132
- "Log Files" on page 132
- "CMC `slot_map` / `blademon` Debugging Hints" on page 133
- "Resolving CMC Slot Map Ordering Issues" on page 134
- "In `tmpfs` Mode, File Has Date in the Future Warnings" on page 135
- "Ensuring Hardware Clock Has the Correct Time" on page 136
- "Troubleshooting a Rack Leader Controller (RLC) With Misconfigured Switch Information" on page 137
- "Switch Wiring Rules" on page 139
- "System Admin Controller (SAC) `eth2` Link in the Bond is Down" on page 140
- "Installing SGI Tempo Versions Older than 2.9.0" on page 140

### About Troubleshooting

This chapter provides answers to some common problems users encounter when installing or upgrading an SGI ICE X system. It includes diagnosis and troubleshooting information.

## Using the `switchconfig` Command

The `switchconfig` command displays switch settings and enables you to configure switches. The command's format is as follows:

```
switchconfig subcommand [--help | -h] [--debug] [--log file | -l file]
[--switch host [, host] [, ...]] [--switches system_ID] [--vlan number] | -v number]
[--ip address | -i address] [--netmask mask | -n mask] [-c old_password] [-n new_password]
```

For *subcommand*, specify one of the following:

**Table 5-1** `switchconfig` Subcommands

Group	Subcommand	Purpose
Informational	<code>list</code>	Displays current settings.
Configuration	<code>set</code>	Assigns virtual local area networks (VLANs) and other settings to one or more MAC addresses.
	<code>unset</code>	Returns ports to default settings.
IP Management	<code>show_ip</code>	Returns the IP address that is configured for the switch.
	<code>set_ip</code>	Adds an IP address for the VLAN on the switch. Used to route traffic from MIC devices.
	<code>unset_ip</code>	Removes the IP address for the VLAN on the switch.
OSPF Management	<code>show_ospf</code>	Returns the open shortest path first (OSPF) router protocol used to route the MIC traffic between the switches.
	<code>set_ospf</code>	Sets the OSPF router protocol used to route traffic between the switches.
	<code>unset_ospf</code>	Disables OSPF and all the network statements that are associated with OSPF.

Group	Subcommand	Purpose
MTU Management	show_mtu	Displays the assigned maximum transportation unit (MTU) value for all ports in the switch stack.
	set_mtu	Assigns all ports in the switch stacks to the MTU value.
	unset_mtu	Removes the MTU value from all ports in the switch stack.
Gateway Management	show_default_gateway	Displays the default gateway assigned to the switch network.
	set_default_gateway	Sets the default gateway for the switch network.
	unset_default_gateway	Clears the default gateway that is assigned to the switch network.
Propagate Configuration	pull_switch_config	Copies the start-up switch's configuration file from the TFTP server to the switch(es) you specify on the command and loads the switch.
	push_switch_config	Copies the switch configuration file from the switch you specify and saves it to the TFTP server.
Miscellaneous	save_running_config	Saves the current switch configuration to the switch's nonvolatile memory.
	reset_factory_defaults	Reverts to the default, factory configuration and reboots.

Group	Subcommand	Purpose
	sanity_check	Runs a sanity check on the switch configuration. For example, it checks the switch ports for trunking. It returns configuration information regarding misconfigured elements and points out anomalies.
	change_password	Changes the administrator password on the switch. By default, the administrator login is <code>admin</code> , and the password is <code>admin</code> . This subcommand is the only subcommand that uses the <code>-c old_password</code> and <code>-n new_password</code> parameters

The rest of the `switchconfig` command arguments are as follows:

Argument	Specification
<i>file</i>	The full path to the file to which <code>switchconfig</code> can write log output.
<i>host</i>	The hostname or IP address of the management switch.
<i>system_ID</i>	The system ID of the switch stack you want to reference. Specify the switch's IP address or hostname for <i>system_ID</i> . If you have more than one switch, specify a comma-separated list of <i>system_ID</i> values. For example, <code>mgmtsw0</code> is the hostname and <i>system_ID</i> for the first management switch.
<i>number</i>	The VLAN number.
<i>address</i>	The IP address needed for the <i>command</i> . This is typically the IP address of a switch that you want the command to operate on, the IP address you want to assign, and so on.
<i>mask</i>	The netmask you want to use.
<i>old_password</i>	The old password that you want to change. By default, this password is the same as the SAC system password.

*new\_password*                      The new password that you want to use on the switches.

Example 1. The following `switchconfig` command returns help text for the entire `switchconfig` command:

```
# switchconfig --help
```

Example 2. The following `switchconfig` command returns help text for only the `set_ip` parameter:

```
# switchconfig set_ip --help
```

The preceding command shows how to display help output for `set_ip`, which is only one of the `switchconfig` subcommands. You can display output for any of the other `switchconfig` subcommands in the same way.

Example 3. The following `switchconfig` command shows how to set the IP address to 100.100.100.100 on for VLAN 101 on switch `mgmtsw0`:

```
# switchconfig set_ip --switch mgmtsw0 --vlan 101 --ip 100.100.100.100
```

Example 4. The following `switchconfig` command adds an IP address for the VLAN on the switches that route traffic for the MIC devices on the compute nodes:

```
# switchconfig set_ip --switch mgmtsw0 --vlan 101 --ip 10.159.2.54 --netmask 255.255.255.0
```

The following command is equivalent:

```
# switchconfig set_ip --s mgmtsw0 --v 101 --i 10.159.2.54 --n 255.255.255.0
```

Example 5. The following two `switchconfig` commands return the switch IP addresses used for routing.

Command one:

```
# switchconfig show_ip -s mgmtsw0 -v 1
VLAN 1 is Administrative Up - Link Up
Address is B4-0E-DC-39-C4-83
Index: 1001, MTU: 1500
Address Mode is DHCP
IP Address: 172.23.0.254 Mask: 255.255.0.0
Proxy ARP is disabled
```

Command two:

```
# switchconfig show_ip -s mgmtsw0 -v 101
VLAN 101 is Administrative Up - Link Up
  Address is B4-0E-DC-39-C4-83
  Index: 1101, MTU: 1500
  Address Mode is User specified
  IP Address: 10.159.1.254 Mask: 255.255.255.0
  IP Address: 10.157.1.254 Mask: 255.255.255.0 Secondary
  IP Address: 10.158.1.254 Mask: 255.255.255.0 Secondary
  IP Address: 10.160.1.254 Mask: 255.255.255.0 Secondary
  Proxy ARP is disabled
```

Example 6. All of the switches for an SGI ICE X system must have the same password. The following command changes the switch password for a system with two switches:

```
# switchconfig change_password -c admin -n mynewpassword --switches mgmtsw0,mgmtsw1
```

## Repairing a Broken DCM Enablement

The following procedure explains a way to troubleshoot a broken DCM enablement.

**Procedure 5-1** To troubleshoot a DCM enablement

1. On the SAC, type the `lk_verify` command to ensure that the SGI ICE X license and the SMC Power Opt license is installed correctly.
2. Open file `/opt/sgi/sgimc/etc/dcm.properties` and add the following line to enable debugging:

```
dcm.debug.log = 1
```

3. Open file `/opt/sgi/sgimc/etc/system-clustermanager.profile` and add the following line to enable debugging:

```
system.logging:com.lnxi.debug
logging.level:DEBUG
logging.subkeys:DCM
```

4. Type the following command to restart the SMC daemon:

```
# service mgr restart
```

5. Use a text editor to open file `/opt/sgi/smc/log/SGIMC-server.log` and examine the file contents.

If the service started successfully, it contains the following entries:

```
Starting communication plugin com.sgi.clusterman.comms.dcm.client.DcmClient...
Starting service plugin com.sgi.clusterman.server.dcm.DcmStateService...
[...]
DCM: Creating SEI proxy with timeout: 30000
com.sgi.clusterman.server.dcm.DcmStateService registered as ServiceProvider DcmStateService
DCM: Event Listener setup complete.
```

## Compute Nodes Are Taking Too Long To Boot

If compute nodes are taking a long time to boot, perform the following:

- See "Verify the Bonding Mode on the Rack Leader Controller (RLC)" on page 126 to verify that the compute nodes have the proper bonding setup.
- Verify the rack leader controller (RLC) has a MegaRAID controller. 144 nodes will not boot well with 106x controllers, for example. You can verify this with `lspci` command.

To verify the MegaRAID battery is working and charged, perform the following:

```
# /opt/MegaRAID/MegaCli/MegaCli64 -ShowSummary -a0
```

You should see 'Status : Healthy' under 'BBU' (BBU = Battery Backup Unit).

---

**Note:** If this is the first time the node has booted up, it takes several hours for the BBU to be charged.

---

- Verify cache is set to write-back, as follows:

```
# /opt/MegaRAID/MegaCli/MegaCli64 -LDGetProp -Cache -LALL -a0
```

---

**Note:** Never force write-back on if bad BBU (`-CachedBadBBU`) as data loss happens with an orderly shutdown that includes a power off.

---

When you see the output: `Cache Policy:WriteBack`, write-back is enabled.

To enable the write-back policy, perform the following:

```
# /opt/MegaRAID/MegaCli/MegaCli64 -LDSetProp -NoCachedBadBBU -LALL -a0
```

## Verify the Bonding Mode on the Rack Leader Controller (RLC)

The redundant management network (RMN) is the default configuration for SGI ICE X systems. To verify the bonding mode, perform the following from the RLC:

```
rllead:~ # cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v3.5.0 (November 4, 2008)
```

```
Bonding Mode: IEEE 802.3ad Dynamic link aggregation
Transmit Hash Policy: layer2+3 (2)
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 0
Down Delay (ms): 0
```

```
802.3ad info
LACP rate: slow
Aggregator selection policy (ad_select): stable
Active Aggregator Info:
    Aggregator ID: 1
    Number of ports: 2
    Actor Key: 17
    Partner Key: 4
    Partner Mac Address: b4:0e:dc:37:4f:a7
```

```
Slave Interface: eth0
MII Status: up
Link Failure Count: 1
Permanent HW addr: 00:25:90:38:e5:22
Aggregator ID: 1
```

```
Slave Interface: eth1
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:25:90:38:e5:23
Aggregator ID: 1
```



If you see Bonding Mode: IEEE 802.3ad Dynamic link aggregation, RMN is on.

If you see Bonding Mode: fault-tolerance (active-backup), it means that the bonding mode and potentially redundant management networking is disabled.

Use the `configure-cluster` GUI **Configure Redundant Management Network** option to turn on the redundant management network (RMN) system for nodes being discovered going forward.

Set the redundant management networking mode on, as follows:

```
# cadmin --enable-redundant-mgmt-network --node r1lead
```

Set the bonding mode per node, as follows:

```
# cadmin --set-mgmt-bonding --node r1lead 802.3ad
```

You need to reboot the system.

The `/proc/net/bonding/bond0` file, should show the bonding mode with link aggregation configured, as follows:

```
Bonding Mode: IEEE 802.3ad Dynamic link aggregation
```

The number of ports should be the following:

```
Number of ports: 2
```

2 is the correct value for an RMN configuration. If the number is 1, it mean the trunk has not formed. The likely causes for this are, as follows:

- The Ethernet cable is not connected to top level switch. From the RLC, use the `/sbin/ethtool` on `eth0` and `eth1` to verify the link is present, as follows:

```
r1lead:~ # /sbin/ethtool eth0
Settings for eth0:
    Supported ports: [ TP ]
    Supported link modes:   10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
                           1000baseT/Full
    Supports auto-negotiation: Yes
    Advertised link modes:  10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
```

```

                                1000baseT/Full
Advertised auto-negotiation: Yes
Speed: 1000Mb/s
Duplex: Full
Port: Twisted Pair
PHYAD: 1
Transceiver: internal
Auto-negotiation: on
Supports Wake-on: umbg
Wake-on: g
Current message level: 0x00000003 (3)
Link detected: yes

```

- The Ethernet cable is connected, but linking is wrong. When the `/sbin/ethtool` command output shows the link speed as 100 Mb/s due to a bad cable the trunk leg is rejected.
- The top level Ethernet switch misconfigured: Perhaps the `switchconfig` tool did not get this port configured properly. You can either log in to the switch to try to diagnose, or try the following procedure:

1. Find the MAC address of the `r1lead` bond interface, as follows:

```

r1lead:~ # ifconfig bond0
bond0    Link encap:Ethernet  HWaddr 00:25:90:38:E5:22
         inet addr:172.23.0.7  Bcast:172.23.255.255  Mask:255.255.0.0
         inet6 addr: fe80::225:90ff:fe38:e522/64 Scope:Link
         UP BROADCAST RUNNING MASTER MULTICAST  MTU:1500  Metric:1
         RX packets:286749167 errors:0 dropped:0 overruns:0 frame:0
         TX packets:328574062 errors:0 dropped:0 overruns:0 carrier:0
         collisions:0 txqueuelen:0
         RX bytes:38868281915 (37067.6 Mb)  TX bytes:153036792319 (145947.2 Mb)

```

2. From the system admin controller (SAC), run the `switchconfig list --switches mgmtsw0` command to list the MAC addresses trunks from the switches, as follows:

```

sys-admin:~ # switchconfig list --switches mgmtsw0
Current MAC/port configuration:

Switch Identifier: mgmtsw0   IP Address: 172.23.0.6

          MAC          Port  Trunk  default-VLAN  allowed-VLANs

```

```

-----
00-25-90-3F-16-C4    1/6                1                1(u)
00-30-48-F7-84-65  1/48                1                1(u)
00-25-90-38-E5-22  1/5                1                1 1(u), 101(t)
00-25-90-38-E5-23  1/5                1                1 1(u), 101(t)
00-25-90-38-E5-22  1/5                1               101 1(u), 101(t)
00-25-90-38-85-BC  1/7                2                1                1(u)
00-25-90-38-85-BD  1/7                2                1                1(u)

```

...

If the RLC `r1lead` bond interface MAC address shows up in the `Port` column and not the `Trunk` column, the switch is not configured correctly.

3. To properly configure the switch, from the SAC node, perform a command similar to the following:

```
# switchconfig set -s mgmtsw0 -v num=1 -v num=101,tag=tagged -b lacp -d 1 -m 00:25:90:38:E5:30
```

This replaces 101 with the proper VLAN number. 101 for rack group 1, 102 for rack group 2, and so on.

4. ssh onto the `r1lead` and verify that the RLC shows `Number of ports: 2`.

## **cimage --push-rack Pushes Too Many (or Too Few) Expansions**

When you perform `cimage --push-rack` (or when `blademon` calls `discovery-rack`), it creates read/write expansions for each compute node.

Use the `configure-cluster` GUI **Configure Default Max Rack IRU Setting** option to set the default number of individual rack units (IRUs), supported by a rack leader controller (RLC). Set this value to the number of CMCs that will be served by each RLC. The default is 8. When you change it, it only impacts node discoveries in the future.

You can change the setting per-node with the `cadadmin` command, as follows:

```
sys-admin:~ # cadadmin --set-max-rack-irus --node r1lead 8
```

## Cannot ping the CMCs from the Rack Leader Controller (RLC)

If this is an RLC with a brand new, never-before-discovered top level switch (or set of switches), the `cmcdetectd` daemon will see CMCs asking for IP addresses on the HEAD network. It configures the top level switch(es) so that the CMCs are on the appropriate rack VLAN. Make sure `cmcdetectd` is running, restart if needed.

You can diagnose this some by running the `tcpdump` command looking for DHCP requests. The requests should be seen on the RLC and not the system admin controller (SAC). For example, type the following command from `r1lead`:

```
# /usr/sbin/tcpdump -i bond0 -s600 -nn -vv -e -t -l -p broadcast and src port 68 and dst port 67
tcpdump: listening on bond0, link-type EN10MB (Ethernet), capture size 600 bytes
00:25:90:3f:16:c4 > ff:ff:ff:ff:ff:ff, ethertype IPv4 (0x0800), length 590: (tos 0x0, ttl 64, id 0, offset 0, \
flags [none], proto UDP (17), length 576) 0.0.0.0.68 > 255.255.255.255.67:
[udp sum ok] BOOTP/DHCP, Request from 00:25:90:3f:16:c4, length 548, xid 0x8b8d332a, Flags [none] (0x0000)
  Client-IP 172.24.0.2
  Client-Ethernet-Address 00:25:90:3f:16:c4
  Vendor-rfc1048 Extensions
    Magic Cookie 0x63825363
    DHCP-Message Option 53, length 1: Request
  ...
```

If the switch was previously discovered but you are reinstalling the system or discovering a new root slot, `cmcdetectd` will not detect any CMC DHCP requests on HEAD. In this case, you need to be sure to run `configure-cluster` and set **Configure Switch Management Network** to yes. Note that changing `configure-cluster` only takes effect for nodes discovered in the future. If you have an existing RLC already discovered, you will need to run a command like the following:

```
# cadmin --enable-switch-mgmt-network --node r1lead
```

After rebooting the RLC, make sure that the `ifconfig` command shows `vlan101` as an interface and not `vlan1` or `vlan2` interfaces, as follows:

```
r1lead:~ # ifconfig
...
vlan101  Link encap:Ethernet  HWaddr 00:25:90:38:E5:22
          inet addr:192.168.160.1  Bcast:192.168.160.255  Mask:255.255.255.0
          inet6 addr: fe80::225:90ff:fe38:e522/64 Scope:Link
          UP BROADCAST RUNNING MASTER MULTICAST  MTU:1500  Metric:1
          RX packets:290550897 errors:0 dropped:0 overruns:0 frame:0
```

```
TX packets:268387414 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:0
RX bytes:30869741447 (29439.6 Mb) TX bytes:120262245830 (114691.0 Mb)
```

Confirm `dhcpd` is running on the RLC. If `dhcpd` is not running, CMCs will not get their IP addresses. Check for errors starting `dhcpd`. If `blademon` failed to create the `ice.conf dhcpd` configuration file (`/etc/dhcpd.conf.d`), see "Restarting the `blademon` Daemon" on page 132.

Verify proper CMC configuration. The CMC is configured for its rack number and slot number. If they are not configured correctly, multiple CMCs can be configured the same way resulting in problems. This can also result in the `ice.conf dhcp` configuration file being corrupted. You may need a USB serial cable to fix the CMCs if this is the case.

One troubleshooting approach is to run `tcpdump` on the RLC, as follows:

```
usr/sbin/tcpdump -i bond0 -s600 -nn -vv -e -t -l -p broadcast and src port 68 and dst port 67
```

Watch the DHCP requests over several minutes. If you see the same Client Identifier being requested by more than one MAC address, you are in a situation where the CMCs are not configured correctly.

Verify that the RLC is properly configured in the switch (see "Troubleshooting a Rack Leader Controller (RLC) With Misconfigured Switch Information" on page 137).

Confirm the wiring rules. See "Switch Wiring Rules" on page 139.

If you moved some CMCs from one RLC number to another and you already adjusted the rack and slot number in the CMC, The switch likely does not know about the changes. The CMCs are likely "lost" in the wrong VLAN, potentially a VLAN that is no longer in use. For example, if you had the CMCs served by the `r3lead` RLC but decided to decommission `r3lead` and move the CMCs to `r1lead` instead this situation could arise. In this case, the switch must be reconfigured. Use the `switchconfig` command to configure the ports connected to those CMCs for head. The system admin controller (SAC) `cmcdetectd` daemon will move them to the correct ultimate location.

You need to know the MACs of the CMC embedded Linux for this, so perhaps record this when you change the slot/rack number in the CMC. **Hint:** `dbdump` may still have the information depending on how you removed the RLC.

An example command is, as follows:

```
# switchconfig set -v num=1 -b manual -d 1 -m 08:00:69:16:51:49 --switches mgmtsw0
```

If you have more than one management switch, then list them in a comma-separated-list for `--switches`.

In a non-redundant-management configuration (switches not stacked), if the `dhcpd` daemon shows DHCP requests from the CMC but the CMC remains unpingable, it could be that both CMC-0 and CMC-1 are connected and linked. This breaks the wiring rules. When we are **not** wired for redundant management networking, only CMC-0 should be connected.

When **not** wired for redundant management networking (when switches are not stacked), do not connect CMC-1.

## Restarting the `blademon` Daemon

From the rack leader controller (RLC), perform the following steps:

1. Stop the daemon:

```
rllead:~ # service blademon stop
```

2. Remove `etc/dhcpd.conf.d/ice.conf` or `/etc/dhcp/dhcpd.conf.d/ice.conf`:

```
rm ice.conf dhcpd.conf
```

3. Remove `slot_map`:

```
rllead:~ # rm /var/opt/sgi/lib/blademon/slot_map
```

4. Start the daemon:

```
rllead:~ # service blademon start
```

## Log Files

All of the log files reside in the `/var/log` directory. In addition to the `messages` log file and in some cases `dhcpd` file on the rack leader controller (RLC), here are some interesting `/var/log` directory log files:

- `/var/log/discover-rack`

On the system admin controller (SAC), the `discover-rack` call is facilitated by `blademon` when new nodes are found. This log will often show problems with discovering nodes.

- `var/log/blademon`

On the RLCs, this shows the `blademon` daemon actions. This includes showing when blade changes are found and it also shows its call to `discover-rack`, and so on. If there are CMC communication issues, they will often be noticed in this log.

- `/var/log/cmcdetected.log`

On the SAC, `cmcdetected` logs its actions as it configures the switches for CMCs in the system. Watch for progress or errors here.

- `/var/log/switchconfig.log`

On the SAC, there is a `switchconfig` command line tool. This tool is largely used by the `discover` command as nodes are discovered. Its actions are logged in to this log file. If RLC VLANs are not functioning properly, check the `switchconfig` log file.

## CMC `slot_map` / `blademon` Debugging Hints

This section describes what to do when the `blademon` daemon cannot find a system blade, as follows:

- Can you ping the CMCs? See "Cannot ping the CMCs from the Rack Leader Controller (RLC)" on page 130.
- If the CMCs are pingable, verify that they have a valid slot map. If the slot map returned by the CMC is missing entries, then `blademon` cannot function properly. It operates on information passed to it by the CMC. Some commands to run from the rack leader controller (RLC) are, as follows:

- Dump the slot map from each CMC to your screen:

```
rlllead:~ # /opt/sgi/lib/dump-cmc-slot-tables
```

- Query an individual slot map:

```
rlllead:~ # echo STATUS | netcat r1i0c 4502
```

---

**Note:** In some software distributions, `netcat` is `nc`.

---

If the CMCs are pingable and the CMCs have valid slot maps, then you can focus on how `blademon` is functioning.

You can turn on debug mode in the `blademon` daemon by sending it a `SIGUSR1` signal from the RLC, as follows:

```
# kill -USR1 pid
```

To turn debug mode off, send it another `SIGUSR1` signal. You should see a message in the `blademon` log about debug mode being enabled or disabled.

The `blademon` daemon maintains the slot map at `/var/opt/sgi/lib/blademon/slot_map` on the RLCs. This appears as `/var/opt/sgi/lib/blademon/slot_map.rack_number` on the system admin controller (SAC).

For a `blademon --help` statement, `ssh` onto the `r1lead` RLC, as follows:

```
[root@admin ~]# ssh r1lead
Last login: Tue Jan 17 13:21:34 2012 from admin
[root@r1lead ~]#
[root@r1lead ~]# /opt/sgi/lib/blademon --help
Usage: blademon [OPTION] ...
```

Discover CMCs and blades managed by CMCs.

Note: This daemon normally takes no arguments.

```
--help      Print this usage and exit.
--debug     Enable debug mode (also can be enabled by setting CM_DEBUG)
--fakecmc   Development only: Discover fake CMCs instead of real ones
--scan-once Initialize, scan for blades, set blades up. Do not daemonize.
            Do not keep looping - do one pass and exit.
```

## Resolving CMC Slot Map Ordering Issues

If there are `ssh(1)` key failures or if the compute node hosts seem to be BMCs, it is possible that there are problems with the CMC slot map might be corrupted.



The CMC maintains a cache file that records which MACs are BMC MACs and which are host MACs. It uses this information, combined with switch port location information in the embedded Broadcom switch, to generate the slot map used by the `blademon` daemon.

In certain situations, such as, a CMC reflash, may remove the cache file but leave CMC power active. In this situation, the CMC does not know which MACs on a given embedded switch port are host and which are BMC and gets the order randomly incorrect. It then caches the incorrect order. To fix this for each CMC, turn the power off with `pfctl`, zero out the MAC cache file, and reset each CMC. Then have `blademon` start over from scratch (see "Restarting the `blademon` Daemon" on page 132). Perform the following steps:

1. `ssh` as root to the rack leader controller (RLC), as follows:

```
sys-admin:~ # ssh rlead
Last login: Thu Jan 26 13:57:53 2012 from admin
rlead:~ #
```

2. Disable the `blademon` daemon, as follows:

```
rlead:~ # service blademon off
```

3. Turn off IRU power for each CMC using the `cpower` command, as follows:

```
# PDSH_SSH_ARGS_APPEND="-F /root/.ssh/cmc_config" pdsh -g cmc pfctl off
```

4. Zero out the slot map cache file, as follows:

```
# PDSH_SSH_ARGS_APPEND="-F /root/.ssh/cmc_config" pdsh -g cmc cp /dev/null /work/net/broadcom_mac_addr_cache
```

5. Reboot the CMC, as follows:

```
# PDSH_SSH_ARGS_APPEND="-F /root/.ssh/cmc_config" pdsh -g cmc reboot
```

6. Restart `blademon` from scratch, see "Restarting the `blademon` Daemon" on page 132.

## In `tmpfs` Mode, File Has Date in the Future Warnings

If you boot a compute node with `tmpfs`, part of the process transfers a root tarball using multicast. This tarball is then expanded. If you see hundreds of "file X has a time in the future" messages, it likely means your hardware clock is not set

to system time properly (see "Ensuring Hardware Clock Has the Correct Time" on page 136).

## Ensuring Hardware Clock Has the Correct Time

Some software distributions do not synchronize the system time to the hardware clock as expected. As a result, the hardware clock may not get synchronized with the system time as it should. At shut down, the system time is copied to the hardware clock, but sometimes this does not happen.

To set all the compute node hardware clocks up properly, perform the following:

- Make sure the system admin controller (SAC) and rack leader controller (RLC) have the correct time
- Make sure the SAC and RLCs are synchronized with `ntp`. A SAC can show a message like the following:

```
ntp[20489]: synchronized to 128.162.244.1, stratum 2
```

- An RLC might show a message like the following:

```
20 Jan 22:54:14 ntp[16831]: synchronized to 172.23.0.1, stratum 3
```

- Make sure the compute nodes have the correct time. They use `ntp` broadcast packets but still will display this:

```
20 Jan 23:05:16 ntp[4925]: synchronized to 192.168.159.1, stratum 4
```

You can also use a command like the following and view the output:

```
sys-admin:~ # pdsh -g leader pdsh -g compute date
```

- Issue the following command to set the hardware clock to the system clock, as follows:

```
sys-admin:~ # pdsh -g leader pdsh -g hwclock --systohc
```

- You can run the `hwclock` without options to confirm the current hardware clock time, as follows:

```
sys-admin:~ # hwclock  
Thu 26 Jan 2012 10:57:27 PM CST -0.750431 seconds
```

## Troubleshooting a Rack Leader Controller (RLC) With Misconfigured Switch Information

Normally, as you discover RLCs, `switchconfig` is called automatically and the switch ports associated with the RLC are configured in the special way needed for RLCs, as follows:

- Default VLAN 1
- Accept rack VLAN packets tagged (rack 1 vlan is vlan101)
- Link Aggregation is the bonding mode between the two ports associated with the RLC

If an RLC is moved in the switch or if `switchconfig` failed during discovery for some reason, you can run `switchconfig` by hand to configure the switch, as follows:

1. Certain switch wires rules must be followed in switch configuration, see "Switch Wiring Rules" on page 139.
2. Make sure all management switches are reachable from the system admin controller (SAC).
3. Find the MAC addresses associated with the RLC interfaces. You can do this by running the following command on the RLC in question:

```
r1lead:~ # cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v3.5.0 (November 4, 2008)

Bonding Mode: IEEE 802.3ad Dynamic link aggregation
Transmit Hash Policy: layer2+3 (2)
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 0
Down Delay (ms): 0

802.3ad info
LACP rate: slow
Aggregator selection policy (ad_select): stable
Active Aggregator Info:
    Aggregator ID: 1
    Number of ports: 2
    Actor Key: 17
    Partner Key: 4
```

Partner Mac Address: b4:0e:dc:37:4f:a7

Slave Interface: eth0  
MII Status: up  
Link Failure Count: 0  
Permanent HW addr: 00:25:90:38:e5:22  
Aggregator ID: 1

Slave Interface: eth1  
MII Status: up  
Link Failure Count: 0  
Permanent HW addr: 00:25:90:38:e5:23  
Aggregator ID: 1



**Caution:** Because bonded interfaces are in play, you cannot get both MAC addresses from using the `ifconfig` command. The `ifconfig` command will show the same MAC address for `eth0` and `eth1` if redundant management networking is enabled.

---

4. Determine which management switches are present, as follows:

```
r1lead:~ # cnodes -mgmtsw  
mgmtsw0
```

5. When you have the list of management switches and the MAC addresses of the RLCs, run a command similar to the following:

```
# switchconfig set --vlan num=1 --vlan num=101,tag=tagged --bonding=802.3ad --default-vlan 1 /  
--macs 00:e0:ed:0a:f2:0d,00:e0:ed:0a:f2:0e --switches mgmtsw0,mgmtsw
```

This replaces the MACs and management switches with the proper ones. It replaces the 101 with the VLAN for the rack, normally "100 + rack number" so rack 1 is 101, rack 2 102.

## Switch Wiring Rules

This section is mainly of interest to SGI ICE X system configurations that have a redundant management network setup (stacked pairs of switches) or larger systems that have switch stacks cascaded from the top level switch.

When discovering cascaded switches, it is impossible to know the connected switch ports of all trunks in advance. So when discovering cascaded switches, you can only start with one cable for discovering, then add the second one later on.

When trunks are configured, it is often hard to find the MAC address of both legs of the trunk. This is because the trunked connection just uses one MAC for the connection. Therefore, you need to rely on rules that infer the second port's connection based on the first port.

Some simple wiring rules are, as follows:

- In a redundant management network (RMN) configuration, when connecting system admin controllers (SACs), rack leader controllers (RLCs), service nodes, and CMCs, you must always use the same port number for the same node in both switches in the stack. In other words:
  - If you connect `r1lead eth0` to switch A, port 43, then you must connect `r1leadeth1` to switch B, port 43.
  - Likewise, if you connect CMC `r1i0c CMC-0` port to switch A, port 2, then `r1i0c CMC-1` port must go to switch B port 2.
- When adding cascaded switch stacks, all switch stacks must cascade from the primary switch stack. In other words, there is always only, at most, one switch hop.
- When discovering cascaded switches pairs in an RMN setup, observe the following:
  - If you are connecting switch stack 1, switch A, port 48 to switch stack 2, then you must connect the second trunked connection to stack 2, switch B, port 48.
  - Until the cascaded switch stack is discovered, you must leave one trunk leg unplugged temporarily to prevent looping.
  - The `discover` command will tell you when it is safe to plug in the second leg of the trunk. This avoids circuit loops.

## System Admin Controller (SAC) eth2 Link in the Bond is Down

A problem occasionally occurs, especially in SGI XE270 SACs, where the active-backup or 802.3ad bonded bond0 interface contains an Ethernet eth2 interface that is down/not linked. To verify this, perform the following:

- Check the Ethernet port of the add-in card and confirm that it is lit.
- Confirm that the add-in card connection to the management switches is using port 0 with port 1 not connected (so not miswired).
- If you look at `/proc/net/bonding/bond0` file, you can confirm that eth2 is the link that is down.
- Use the `/sbin/ethtool eth2` command and confirm that the `Link detected:` is no.
- Run the commands `ifconfig up eth3` and then run the `/sbin/ethtool eth3` command to determine if the link detected is yes.

In this scenario, it is likely that the eth2/eth3 interfaces have been swapped. Another clue is that if eth2 (look at `/proc/net/bonding/bond0` since the bond enforces the same MAC address for all bonded members) has a MAC address that is larger than the MAC address of eth3 (as seen by `ifconfig eth3`).

To correct this situation, edit the `/etc/udev/rules.d/70-persistent-net.rules` file and swap the MACs associated with eth2 and eth3 in the file.

When you reboot the system, the SAC comes back up with eth2 and eth3 properly ordered.

## Installing SGI Tempo Versions Older than 2.9.0

After you upgrade or install SGI Tempo 2.9.0 on any slot, the boot manager is changed to GRUB version 2. At this point, you can no longer install SGI Tempo versions earlier than 2.9.0 on any slot. The procedure in this topic explains how to install an earlier SGI Tempo version. For information about booting an SGI Tempo 2.9.0 system, see the following:

"Disk Partition Layout and Boot Order Information" on page 23

The following procedure explains how to install a version of SGI Tempo that is earlier than 2.9.0. After you run this procedure, the boot system is again GRUB version 1. Subsequently, if you install SGI Tempo 2.9.0, the boot system changes again to GRUB version 2.

**Procedure 5-2** To install an SGI Tempo version that is earlier than SGI Tempo 2.9.0

1. Log into the slot upon which you installed SGI Tempo 2.9.0 as the root user.
2. Type the following command to revert to the GRUB version 1 boot manager on the SAC:

```
# /opt/sgi/lib/revert-admin-to-legacy-grub
```

3. Use the SGI DVD to install the older version of SGI Tempo that you want to use.





---

## YaST2 Navigation

The following list shows SLES YaST2 navigation key sequences:

<b>Key</b>	<b>Action</b>
Tab	
Alt + Tab	
Esc + Tab	
Shift + Tab	
	Moves you from label to label or from list to list.
Ctrl + L	Refreshes the screen.
Enter	Starts a module from a selected category, runs an action, or activates a menu item.
Up arrow	Changes the category. Selects the next category up.
Down arrow	Changes the category. Selects the next category down.
Right arrow	Starts a module from the selected category.
Shift + right arrow	
Ctrl + A	
	Scrolls horizontally to the right. Useful in screens if use of the <code>left arrow</code> key would otherwise change the active pane or current selection list.
Alt + <i>letter</i>	
Esc + <i>letter</i>	
	Selects the label or action that begins with the <i>letter</i> you select. Labels and selected fields in the display contain a highlighted <i>letter</i> .
Exit	Quits the YaST2 interface.



## Subnetwork Information

This chapter contains the following topics:

- "About the SGI ICE X Subnetworks" on page 145
- "VLAN Ethernet Network Configurations" on page 145
- "SGI ICE X IP Address Ranges and VLANs for Management and Application Software" on page 150
- "SGI ICE X Subnetworks for Systems That Include MIC Devices"
- "Component Naming Conventions" on page 155
- "System Control Configuration" on page 157

### About the SGI ICE X Subnetworks

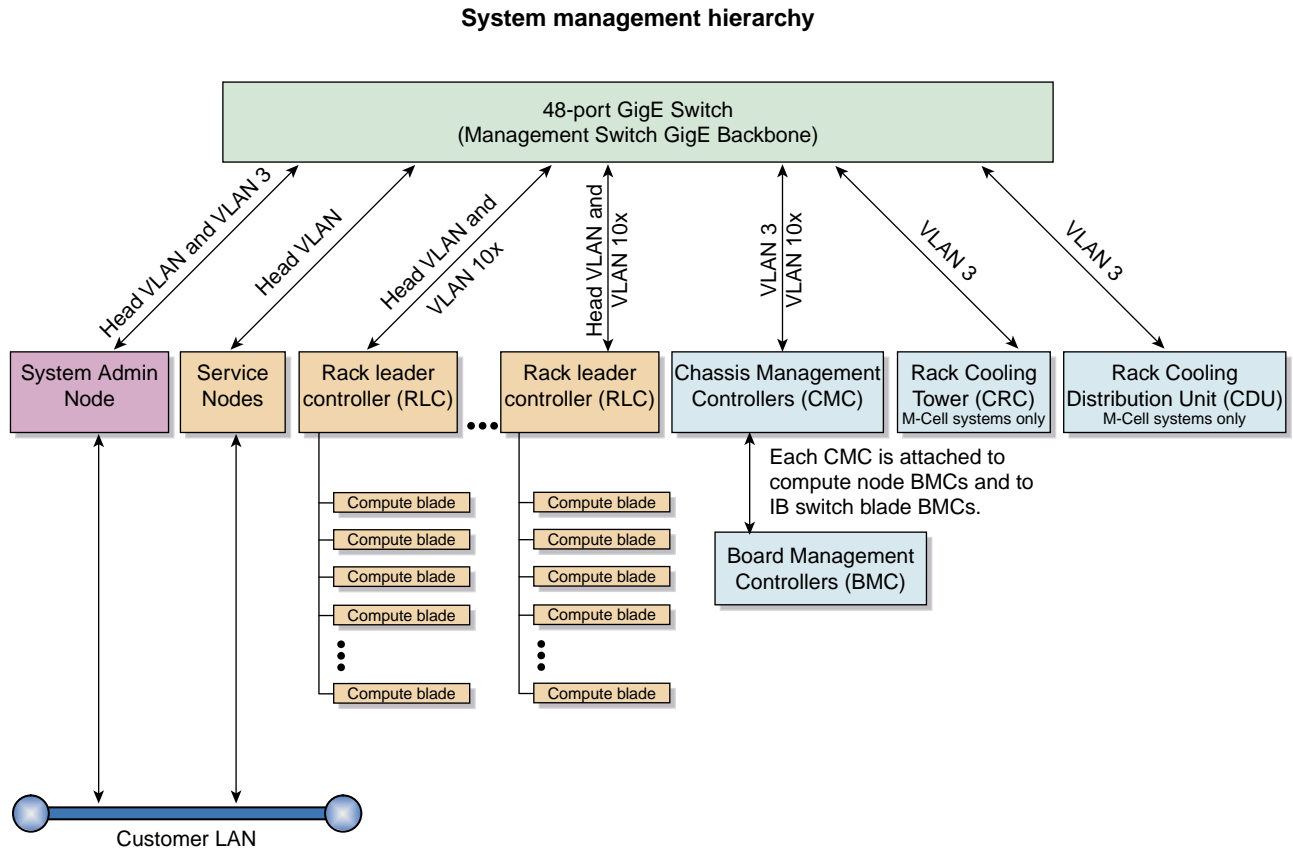
SGI ICE X hardware components are attached to one or more VLANs. This appendix section contains networking reference material that can be useful if you want to reconfigure or debug an SGI ICE X network.

This appendix section includes VLAN information for SGI ICE X systems that include Intel Xeon Phi Many Integrated Core (MIC) devices. The VLAN information for the MIC devices pertains only to SGI ICE X systems that include MIC devices on the compute blades and service nodes. Within this appendix section, the subnetworks are named as you see them in the cluster configuration tool's **Subnet Network Address** menu.

### VLAN Ethernet Network Configurations

An SGI ICE X system includes at least two VLANs: one head VLAN and one or more rack VLANs. There is one rack VLAN for each rack in the system. The VLAN configuration file is `/opt/sgi/lib/discover-ice-backend`.

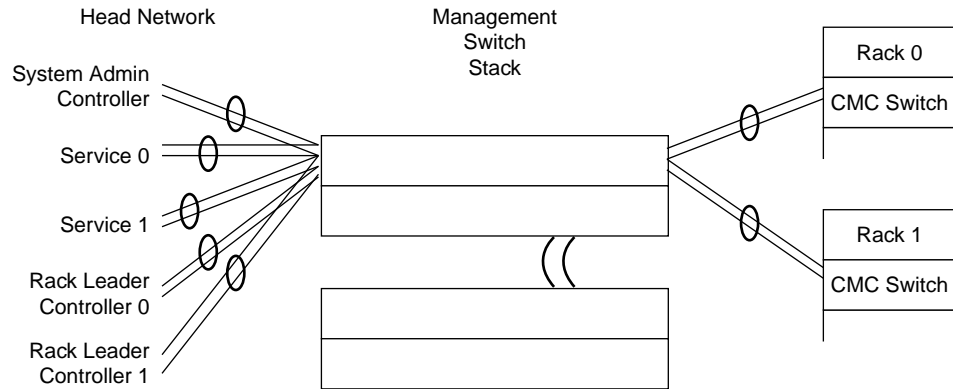
Figure B-1 on page 146 shows an SGI ICE X system with three VLANs.



**Figure B-1** VLAN Overview

If your SGI ICE X system is equipped with MIC devices, each compute blade includes one or two IP addresses for each device. Only one cable connects each compute blade to the network, but each MIC device requires its own, unique IP address on your network.

Figure B-2 on page 147 shows how the components in an SGI ICE X system connect to the master switch stack.



**Figure B-2** SGI ICE X Management Switch Connections

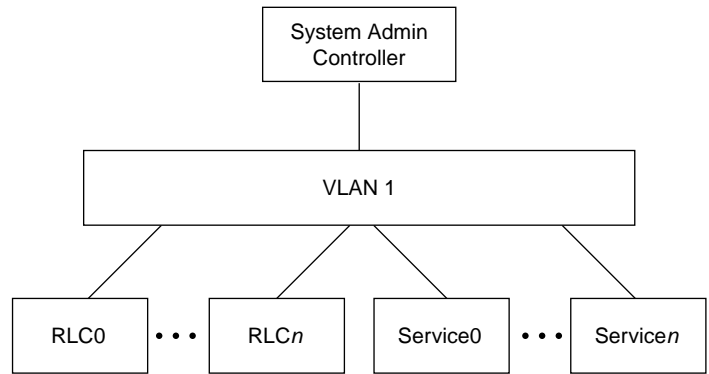
The following topics contain more information about the head and rack VLANs:

- "Head VLAN" on page 147
- "Rack VLANs" on page 148
- "MCell Cooling VLAN" on page 149

## Head VLAN

The head VLAN is VLAN 1. The head VLAN includes the system admin controller (SAC), all rack leader controllers (RLCs), and all service nodes. The head VLAN is always configured as untagged. Any untagged packets coming into a SAC, RLC, or service node port are associated with the head VLAN. In the cluster configuration tool's menus, the head VLAN appears as Head.

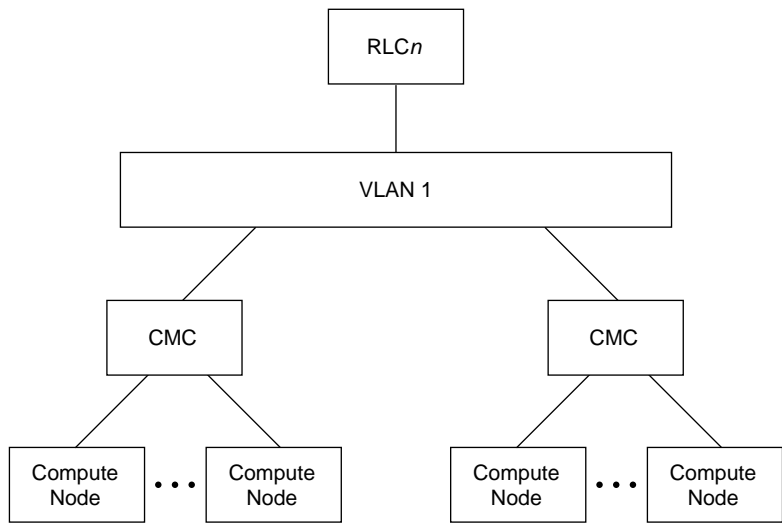
Figure B-3 on page 148 shows the components that participate in the head network.



**Figure B-3** HEAD VLAN Ethernet Connections

### Rack VLANs

There is one rack VLAN for each rack in the system. The VLANs are numbered incrementally. The VLAN for rack 1 is 101. The VLAN for rack 2 is 102. The VLAN for rack 99 is 199. At the maximum, the VLAN number for rack 1000 is 1100. Figure B-4 on page 148 shows the rack VLAN for an example rack.



**Figure B-4** RACKx VLAN Ethernet Connections

The following system components reside on a rack VLAN:

- Rack leader controllers (RLCs). Each RLC resides in both the head VLAN and on its own rack VLAN. The dual residence enables the RLCs to communicate with both the components in the head network and with the compute nodes in their rack.

The RLC's management-related IP address subnetworks are as follows:

- One IP address on the head VLAN.
- The following two IP addresses on the rack VLAN:
  - The BMC network IP address
  - The GBE network IP address

The RLC Ethernet interface is configured with VLAN tagging. VLAN tagging ties these networks to the rack VLAN.

- Chassis management controllers (CMCs). Each CMC internal switch cascades down from the top level switch. The switch ports for all CMCs in a rack are configured to be on the rack VLAN. Each CMC connects to a port on a top level switch. The port is configured so that all traffic coming in and going to that port travels to the rack VLAN by default. The CMC gets its rack VLAN IP address using DHCP.
- Compute nodes. The backplane connects the compute nodes to the cascaded CMC switch. The compute node's BMC has a shared Ethernet connection with the host interface. Both the compute node and BMC traffic are on the rack VLAN.

The compute nodes and baseboard management controllers (BMCs) reside on the same rack VLAN. The BMCs have a subnetwork that is separate from the host interfaces.

## MCell Cooling VLAN

Figure B-5 on page 150 shows VLAN 3, which is the VLAN that the MCell cooling system uses on SGI ICE X systems. This VLAN resides on your system only if you have MCells.

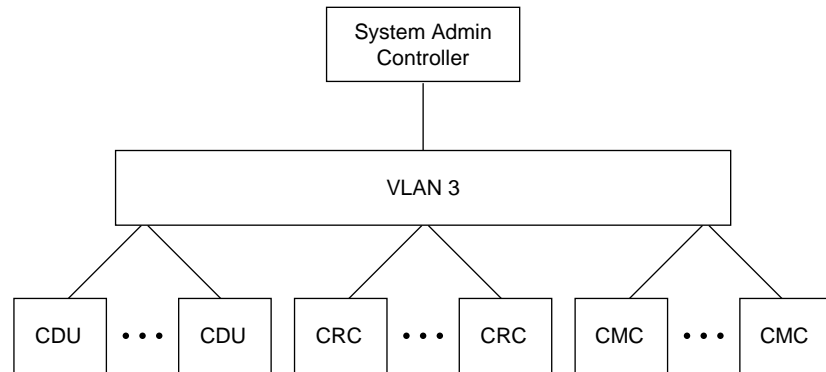


Figure B-5 MCell Cooling VLAN

## SGI ICE X IP Address Ranges and VLANs for Management and Application Software

Table B-1 on page 150 shows the system-wide SGI ICE X IP address ranges that the cluster management software uses. The following notes pertain to this table:

- The `Head_bmc` network is a separate IP subnetwork.
- The `Mic0`, `Mic1`, `Mic2`, and `Mic3` subnetworks apply only to the service nodes on SGI ICE X systems that include Intel Many Integrated Core Architecture (Intel MIC Architecture) based products. For these processors, the IP address ranges for the `Mic0`, `Mic1`, `Mic2`, and `Mic3` subnetworks intentionally overlap with the IP address range for the head network. If you change the default IP addressing scheme, make sure your new IP addresses also overlap.

Table B-1 System-wide IP Address Ranges for the Head Network



VLAN	Subnetwork Name	IP Range	Nodes
1	Head	172.23.0.0/16	SAC Service nodes RLCs
1	Head_bmc	172.24.0.0/16	SAC BMC Service BMCs RLC BMCs
1	Mic0	172.23.160.0	Service nodes with MIC devices
1	Mic1	172.23.176.0	Service nodes with MIC devices
1	Mic2	172.23.192.0	Service nodes with MIC devices
1	Mic3	172.23.208.0	Service nodes with MIC devices

Table B-2 on page 151 shows the per-rack IP address ranges that the cluster management software uses in the rack VLANs.

**Table B-2** Per-rack IP Address Ranges for Cluster Management Software

Rack VLAN Name	Subnetwork Name	IP Range	Components
101	Gbe	10.159.1.0/24	Rack 1's RLC Rack 1's CMCs Rack 1's compute nodes
101	Bmc	10.160.1.0/24	Rack 1's RLC's BMC Rack 1's CMCs' BMC Rack 1's compute nodes' BMCs
	Mic0-gbe	10.157.1.0/24	

B: Subnetwork Information

---

Rack VLAN Name	Subnetwork Name	IP Range	Components
	Mic1-gbe	10.158.1.0/24	
102	Gbe	10.159.2.0/24	Rack 2's RLC Rack 2's CMCs Rack 2's compute nodes
102	Bmc	10.160.2.0/24	Rack 2's RLC's BMC Rack 2's CMCs' BMC Rack 2's compute nodes' BMCs
	Mic0-gbe	10.157.2.0/24	
	Mic1-gbe	10.158.2.0/24	
X	Gbe	10.159.X.0/24	Rack X's RLC Rack X's CMCs Rack X's compute nodes
X	Bmc	10.160.X.0/24	Rack X's RLC's BMC Rack X's CMCs' BMCs Rack X's compute nodes' BMCs
	Mic0-gbe	10.157.X.0/24	
	Mic1-gbe	10.158.X.0/24	

Table B-3 on page 153 shows the system-wide IP address ranges for cluster application software. Only the RLCs that provide InfiniBand subnetwork services need to connect.

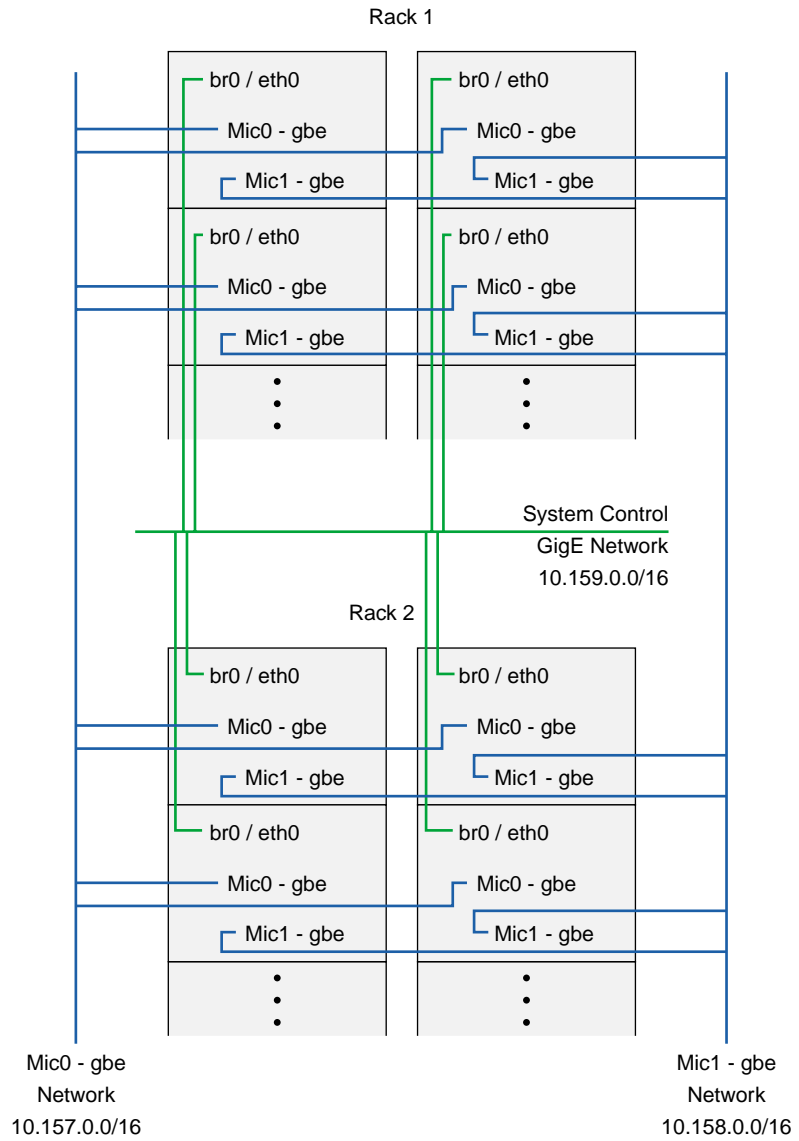
**Table B-3** Application Software System-wide IP Address Ranges

VLAN Name	Subnetwork		Nodes
	Name	IP Range	
IB0	ib-0	10.148.0.0/16	Service nodes Some RLCs Compute nodes
IB1	ib-0	10.149.0.0/16	Service nodes Some RLCs Compute blades

## SGI ICE X Subnetworks for Systems That Include MIC Devices

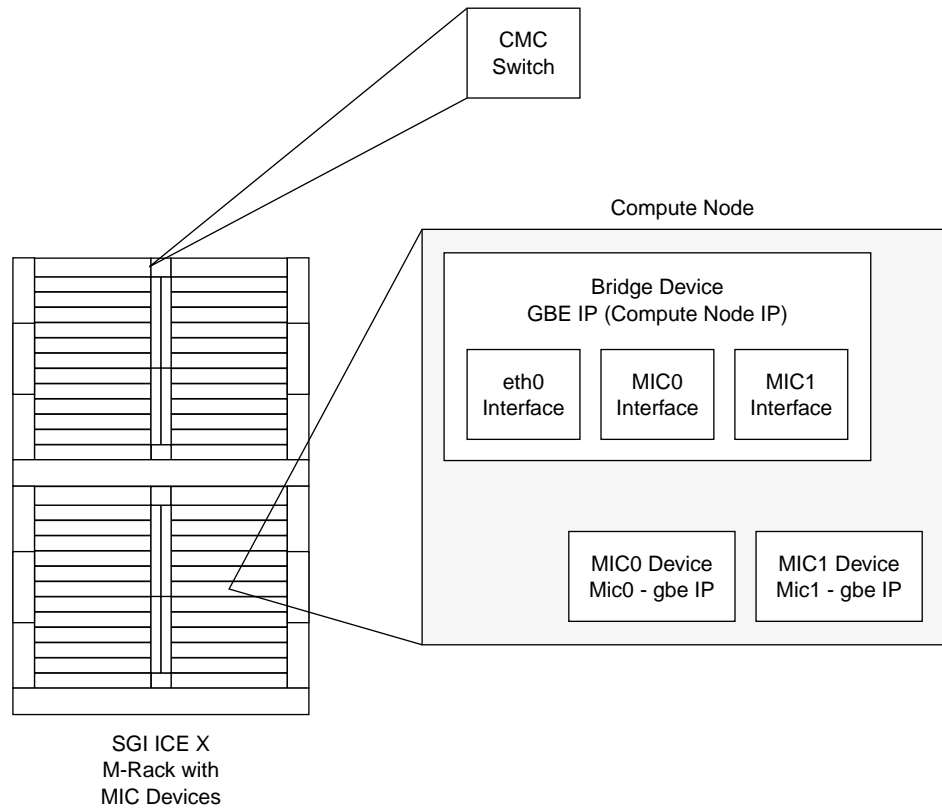
The Intel Xeon Phi Coprocessor is the brand name for all Intel Many Integrated Core Architecture (Intel MIC Architecture) based devices. If the compute nodes of your SGI ICE X system include MIC devices, then your system includes the additional networks upon which these MIC devices reside. The following topics explain these networks and describe how to troubleshoot an SGI ICE X system that includes MIC devices.

Figure B-6 on page 154 shows the various subnetworks that exist on an SGI ICE X system M-Rack that contains compute nodes with MIC devices.



**Figure B-6** SGI ICE X M-Rack Network Connections

Figure B-7 on page 155 shows the network interfaces that reside on the compute nodes of an SGI ICE X system with MIC devices.



**Figure B-7** SGI ICE X M-Rack with MIC Devices

## Component Naming Conventions

The SGI ICE X commands enable you to run commands and perform some procedures on only one component or on a range of similar components. Addressing methods differ depending on the component, the VLAN (or VLANs) in which the component resides, and whether or not the component has an IP address that is externally available.

The topics that follow use the following terms:

- **Component.** The name of the component that you typically use in speech or in writing. For example: SAC, RLC, and so on.
- **Node name.** The system-wide unique identifier for the component.

Table B-4 on page 156 explains how to specify components when you run administrative and user commands. *x* is always an integer number.

**Table B-4** Naming Conventions

Component	Node name	Examples
SAC	Site-defined hostname	icex1 mysiteicex sleet
RLC	rxlead	r0lead, the RLC on the first rack r1lead, the RLC on the second rack
RLC BMC	rxlead-bmc	r1lead-bmc, the BMC on the second rack
Service node	servicex servicex-mic0 servicex-mic1 servicex-mic2 servicex-mic3	service0, the first service node service3, the fourth service node
Service node BMC	servicex-bmc	service1-bmc, the BMC on the second service node
InfiniBand switch	ibswitchx-bmc	ibswitch1-bmc, the BMC on the second InfiniBand switch
Compute node	rxixnx rxixnx-eth rxixnx-mic0 rxixnx-mic1	r1i3n10, the compute node on the second rack, in the fourth IRU, in position 10

Component	Node name	Examples
Compute node BMC	<code>rxixnx-bmc</code>	<code>r1i3n10-bmc</code> , the BMC on the second rack, in the fourth IRU, in position 10
CMC	<code>rxixc</code>	<code>r1i1c</code> , the CMC for the second rack, in the second IRU

## System Control Configuration

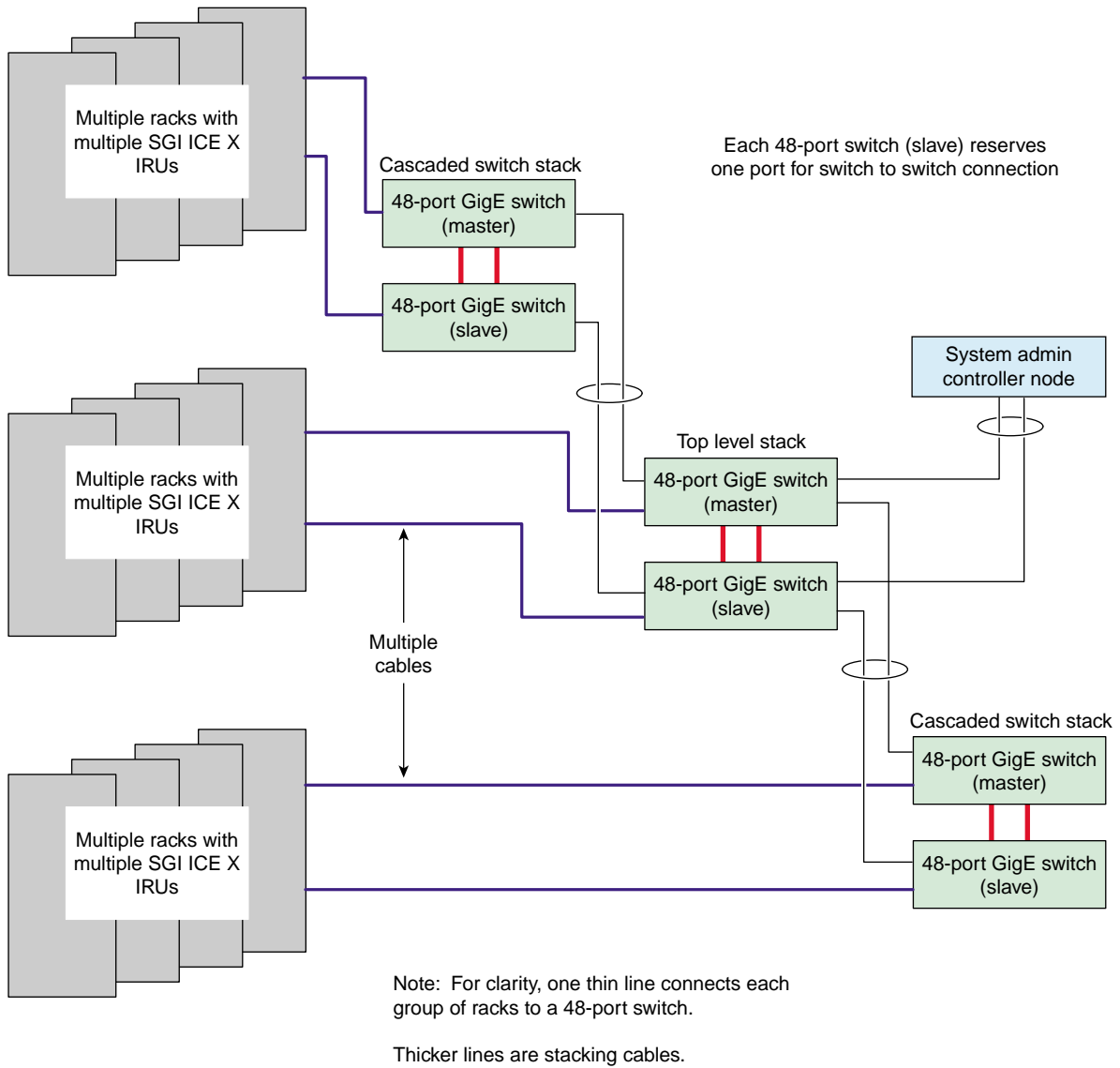
The system control network for an SGI ICE X system can be configured in one of the following ways:

- A redundant management network configuration. This is the default. In a redundant management network configuration, the number of GigE switches in the system control network is doubled. A redundant management network also includes the following:
  - The GigE switches are stacked (using stacking cables).
  - The links from the CMCs are doubled.
  - Links from the SAC, RLCs, and most service nodes are doubled. BMC connections are not doubled. Certain failures can cause temporary inaccessibility to the BMCs, but the host interfaces remain accessible.

Figure B-8 on page 158 shows the switches in a redundant management network configuration.

- A nonredundant management network configuration. In the nonredundant configuration, a single GigE fabric has a single connection to the SAC, RLCs, and CMCs. Figure B-9 on page 159 shows the switches in a nonredundant management network configuration.

Figure B-8 on page 158 shows a redundant management network cascaded switch configuration.



**Figure B-8** Redundant Cascaded Switch Configuration



Figure B-9 on page 159 shows a nonredundant management network cascaded switch configuration.

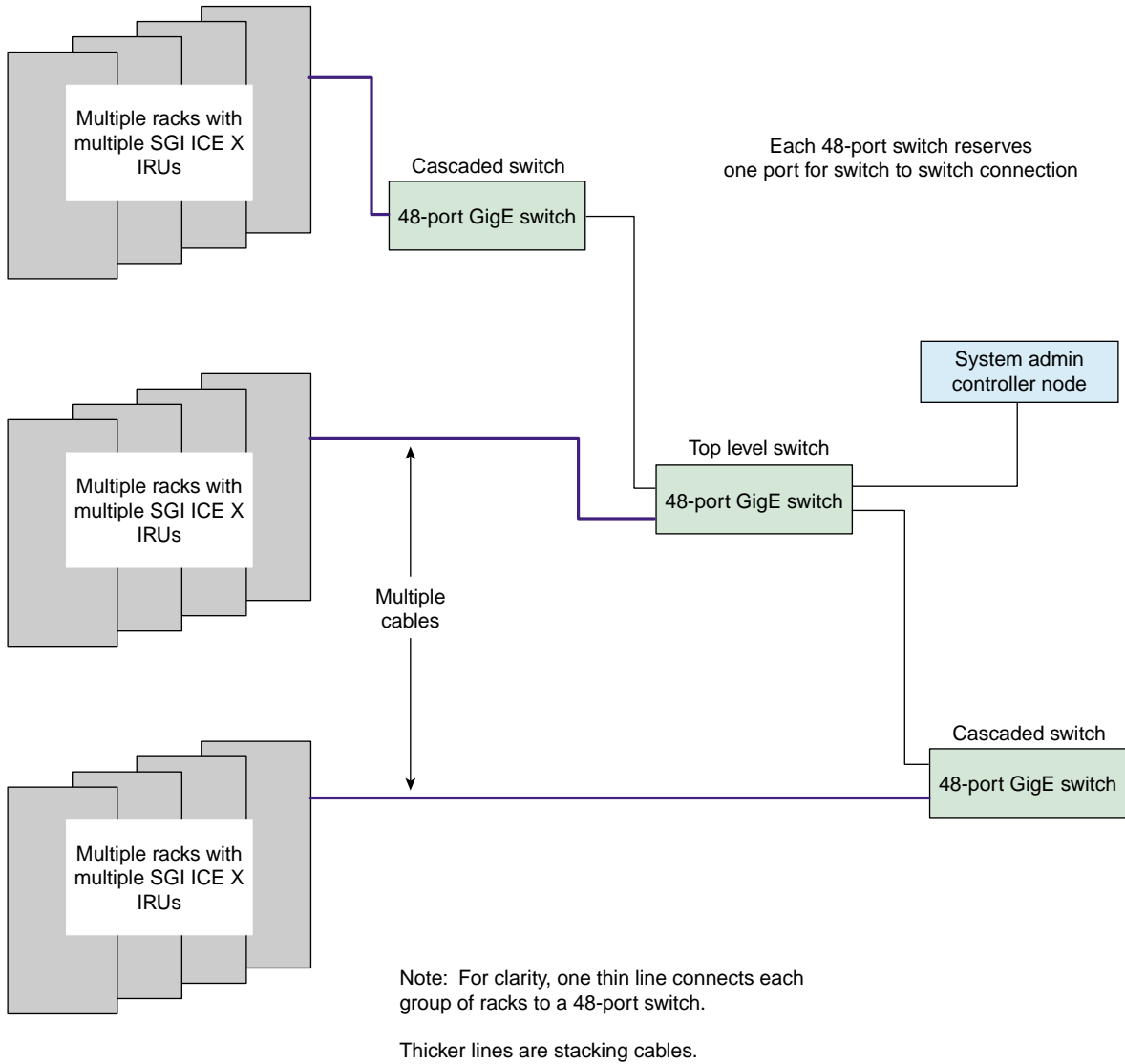


Figure B-9 Nonredundant Cascaded Switch Network Configuration

For diagrams that show both redundant and nonredundant management network wiring, see the chassis manager interconnect diagrams in the *SGI ICE X System Hardware User Guide*.

---

## MCell Network IP Addresses

If you need to troubleshoot the MCell cooling equipment on an SGI ICE X system, you need to know the IP addresses of the cooling rack controllers (CRCs) and cooling distribution units (CDUs) so that you can type a `ping(8)` command to the component. Each piece of equipment bears a label with its equipment number.

For CRCs, the IP address is `172.26.128.number`.

For CDUs, the IP address is `172.26.144.number`.

Table C-1 on page 161 shows the *number* at the end of the IP address for CRCs and CDUs.

**Table C-1** MCell Network Associations

Rack	Cooling Rack Controllers (CRCs)	Cooling Distribution Unit (CDUs)
1	1	1
2	1	1
3	2	1
4	2	1
5	3	2
6	3	2
7	4	2
8	4	2
9	5	3
10	5	3
11	6	3
12	6	3
13	7	4
14	7	4

---

Rack	Cooling Rack Controllers (CRCs)	Cooling Distribution Unit (CDUs)
15	8	4
16	8	4
17	9	5
18	9	5
19	10	5
20	10	5
21	11	6
22	11	6
23	12	6
24	12	6
25	13	7
26	13	7
27	14	7
28	14	7
29	15	8
30	15	8
31	16	8
32	16	8
33	17	9
34	17	9
35	18	9
36	18	9
37	19	10
38	19	10
39	20	10
40	20	10

---

Rack	Cooling Rack Controllers (CRCs)	Cooling Distribution Unit (CDUs)
41	21	11
42	21	11
43	22	11
44	22	11
45	23	12
46	23	12
47	24	12
48	24	12
49	25	13
50	25	13
51	26	13
52	26	13
53	27	14
54	27	14
55	28	14
56	28	14
57	29	15
58	29	15
59	30	15
60	30	15
61	31	16
62	31	16
63	32	16
64	32	16
65	33	17
66	33	17

---

Rack	Cooling Rack Controllers (CRCs)	Cooling Distribution Unit (CDUs)
67	34	17
68	34	17
69	35	18
70	35	18
71	36	18
72	36	18
73	37	19
74	37	19
75	38	19
76	38	19
77	39	20
78	39	20
79	40	20
80	40	20
81	41	21
82	41	21
83	42	21
84	42	21
85	43	22
86	43	22
87	44	22
88	44	22
89	45	23
90	45	23
91	46	23
92	46	23

---

Rack	Cooling Rack Controllers (CRCs)	Cooling Distribution Unit (CDUs)
93	47	24
94	47	24
95	48	24
96	48	24
97	49	25
98	49	25
99	50	25
100	50	25

---





## Legacy Partition Layout Information

This appendix section includes the following topics:

- "About the Legacy Partition Layout" on page 167
- "Legacy Partition Layout for a One-slot SGI ICE X System" on page 167
- "Legacy Partition Layout for a Two-slot SGI ICE X System" on page 168
- "Legacy Partition Layout for a Five-slot SGI ICE X System" on page 169

### About the Legacy Partition Layout

As of the Tempo 2.9.0 release, the partition layout scheme has changed for SGI ICE X systems. If you upgraded a system to the Tempo 2.9.0 or later release, it is possible for your system to have the following mixture of slots:

- Slots with the legacy, Tempo pre-2.9.0 partition scheme
- Slots with the new Tempo 2.9.0 partitioning scheme

For information about Tempo and partitioning, see the following:

"Disk Partition Layout and Boot Order Information" on page 23

### Legacy Partition Layout for a One-slot SGI ICE X System

Table D-1 on page 168 shows the legacy partition layout for a one-slot SGI ICE X system. This layout yields one boot partition.

**Table D-1** Partition Layout for a Single-boot SGI ICE X System

Partition	File System Type	File System Label	Notes
1	ext3	sgiboot	/boot partition.
2	-	N/A	Extended partition. Makes logicals out of the rest of the disk.
5	swap	sgiswap	Swap partition.
6	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot	/ partition.

## Legacy Partition Layout for a Two-slot SGI ICE X System

Table D-2 on page 168 shows the legacy partition layout for a two-slot SGI ICE X system. This layout yields two boot partitions.

**Table D-2** Partition Layout for a Dual-boot SGI ICE X System (Default Layout)

Partition	File System Type	File System Label	Notes
1	swap	sgiswap	Partition layout for multiple slots.
2	ext3	sgidata	SGI data partition.
3	-	N/A	Extended partition. Makes logicals out of the rest of the disk.
5	ext3	sgiboot	Slot 1 /boot partition.
6	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot	Slot 1 / partition.

Partition	File System Type	File System Label	Notes
7	ext3	sgiboot	Slot 2 /boot partition.
8	ext3 on SAC, service nodes. XFS on RLCs.	sgiroot	Slot 2 / partition.

## Legacy Partition Layout for a Five-slot SGI ICE X System

Table D-3 on page 169 shows the legacy partition layout for a five-slot SGI ICE X system. This layout yields five boot partitions.

**Table D-3** Legacy Partition Layout for a Quintuple-boot SGI ICE X System

Partition	File System Type	File System Label	Notes
1	swap	sgiswap	Partition layout for multiple slots.
2	ext3	sgidata	SGI data partition.
3	-	N/A	Extended partition. Makes logicals out of the rest of the disk.
5	ext3	sgiboot	Slot 1 /boot partition.
6	ext3 or XFS	sgiroot	Slot 1 / partition.
7	ext3	sgiboot	Slot 2 /boot partition.
8	ext3 or XFS	sgiroot	Slot 2 / partition.
9	ext3	sgiboot	Slot 3 /boot partition.
10	ext3 or XFS	sgiroot	Slot 3 / partition.
11	ext3	sgiboot	Slot 4 /boot partition.
12	ext3 or XFS	sgiroot	Slot 4 / partition.
13	ext3	sgiboot	Slot 5 /boot partition.
14	ext3 or XFS	sgiroot	Slot 5 / partition.



---

# Index

## C

Configure backup DNS server, 97

## D

DHCP option code, 92

dhcp options  
changing, 92

disabling InfiniBand switch monitoring, 98

## I

InfiniBand configuration, 98

disabling InfiniBand switch monitoring, 98

initial configuration of a RHEL 6 SAC, 35

Intel Turbo Boost Technology feature, 109

## N

network interface naming conventions, 145

## networks

network interface naming conventions, 145

## O

overview, 1

## S

system management software, 1

system overview, 1

## T

### troubleshooting

frequently asked questions, 119

initial system setup, 119