



SC072-PDS User's Guide

For Software Version 3.1-WS



Trademarks

Cray is a registered trademark of Cray, Inc.

Intel is the registered trademark of Intel Corporation.

Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. The registered trademark Linux is used pursuant to a sublicense from the Linux Mark Institute, the exclusive licensee of Linus Torvalds, owner of the mark in the U.S. and other countries.

Red Hat® Enterprise Linux® is a registered trademark of Red Hat, Inc. in the United States and other countries.

Lustre is the registered trademark of Cluster File Systems, Inc.

MIPS and MIPS64 are registered trademarks of MIPS Technologies, Inc.

NIST is a registered trademark of the National Institute of Standards and Technology, U.S. Department of Commerce.

OpenMP is a trademark of Silicon Graphics, Inc.

PCI, PCI Express, and PCIe are registered trademarks, and EXPRESSMODULE is a trademark of PCI-SIG.

Perl is the registered trademark of The Perl Foundation

SiCortex is a registered trademark, and the SiCortex logo, SC5832, SC1458, SC648, SC072-PDS, and PathScale are trademarks of SiCortex, Incorporated.

TAU Performance System is a trademark of the joint developers: University of Oregon Performance Research Lab; Los Alamos National Laboratory Advanced Computing Laboratory; and The Research Centre Jülich, ZAM, Germany.

Vampir is a registered trademark of Wolfgang E. Nagel.

All other brand and product names are trademarks or service marks of their respective owners.

Copyrights

Copyright© 2008 SiCortex Incorporated. All rights reserved.

Disclaimer

The content of this document is furnished for informational use only, is subject to change without notice, and should not be construed as a commitment by SiCortex, Inc.

Document Number 2910-02 Rev. 03
Published January 20, 2009

Contacting SiCortex and Getting Support

SiCortex is on-line at <http://www.sicortex.com>. Our Web pages provide information on the company and products, including access to technical information and documentation, product overviews, and product announcements.

You can search the SiCortex Knowledge Base or participate in forum discussions online at <http://www.sicortex.com/support> after registering.

Customers with service contracts can reach SiCortex Technical Support by e-mailing questions to support@sicortex.com. All customers can participate in the SiCortex support forums.

Purpose and Audience of this Book

This manual describes how to configure and manage the SC072-PDS system using the Red Hat desktop GUI. It also describes how to run applications on the system. For instructions on using the command line tools, see the Red Hat documentation.

For features and functions that must be configured by modifying configuration files supplied by SiCortex, you will be directed to the appropriate set of instructions in the *SiCortex[®] System Administration Guide*.

Accessing User Documentation

PDFs of the SC072-PDS user documentation are installed on the SSP in `/opt/sicortex/doc/[hardware|software]`. You can access the documentation two ways:

- Locally on the workstation, navigate to the document you want to view, and click the `pdf` file to open it in the PDF viewer.
- Remotely, log on to the workstation, navigate to the document you want to view, copy (`scp`) it to your local workstation, and open it in a PDF viewer.

From here, you can print the document.

Conventions of Notation

Bold	Denotes a selection to make in a GUI program. For example, File → Process → Startup directs the user to select File located on the application's toolbar, then Process , and then Startup .
monospaced font	Denotes code examples wherever they occur and command sequences and their arguments, which are entered at the system prompt.
<i>monospaced italics</i>	Denotes an argument in a command for which you substitute the actual value.
<i>Italics</i>	Denotes a term or a cross reference in general text.
	Denotes a caution or warning, such as a dependency that must be satisfied before continuing a process.
	Denotes a tip, hint, or reminder.



Table of Contents

Contacting SiCortex and Getting Support.....	iii
Purpose and Audience of this Book	iii
Accessing User Documentation	iii
Chapter 1 — Introducing the SC072-PDS	9
Hardware	9
Software.....	11
Chapter 2 — Setting up the System	13
Powering on the System.....	13
Booting the System’s Nodes	14
Logging in to the Head Node as root	14
Chapter 3 — Managing the System	15
About the Root File System	15
Root File System for the Workstation.....	15
Root File System for the Nodes	15
Editing the Node Rootfs.....	16
Alias head for the Head Node	16
Changing the Root Password	16
Changing Date and Time Parameters	17
Setting up Networking.....	19
Connecting the SSP to the LAN	19
Setting up Networking for the Nodes	20
Using the SSP as the Router to the LAN	21
Using the Head Node as the Router to the LAN.....	22
Using Node 9, 10, or 11 as the Router to the LAN.....	24

Setting up Extra Routes.....	24
Specifying the DNS Servers to Use	25
Setting up as a Standalone System	26
Managing User Access to the System	26
Configuring User Authentication for a Site Server.....	27
Configuring Standalone User Authentication	28
Configuring Network File Systems	30
Mounting Home Directories.....	30
Setting up a Lustre File System on Direct-Attached Storage.....	31
SiCortex System Services.....	32
Services Started on the Nodes.....	32
Chapter 4 — Using the System	33
Logging on to the System.....	33
Logging in Locally Through the Workstation.....	33
Logging in Remotely via the Workstation to the Head Node.....	34
Logging in Remotely to the Head Node.....	35
Managing Jobs and Resources on the System	36
About SLURM	36
Running Jobs	36
Managing Jobs.....	36
Compiling and Running Applications on the Nodes	37
Running Applications.....	37
Chapter 5 — Monitoring the System.....	39
Monitoring the System	39
Nagios Alerts.....	40
Policyd Log File Entries.....	40
Viewing System Logs.....	41
Chapter 6 — Updating System Software	45
Checking Your Current Software Version	45
Installing a New Release	45
Install DVDs.....	45
Installing from DVD.....	46
Chapter 7 — Troubleshooting	49
Diagnosing Failed Nodes and Links.....	49
Declaring Nodes or Links Disabled.....	51
Restoring Disabled Components	53
Cancelling Jobs.....	53
Cancelling Jobs When the System Is Working	54
Cancelling Jobs When the System Is Having Problems	55
Appendix A — System Administration Defaults.....	57
Services	60
System Log	60

Users and Groups	60
Appendix B — The sicortex-system.conf File	61
Index	i

Chapter 1 Introducing the SC072-PDS

The SC072-PDS (Personal Development System) is a 72-processor cluster with a built-in x86 workstation, that fits on or beside your desk.

The workstation runs Red Hat® Enterprise Linux®. You can use the Red Hat Linux desktop GUI to configure and control the cluster. From the workstation, you can configure access to the nodes and monitor their operational status, and you can also monitor the environmental state of the entire cluster.

Hardware

Figure 1 provides a graphical overview of the system's internal components and connections. For installation instructions, see the *SC072-PDS Quick Start Guide*.

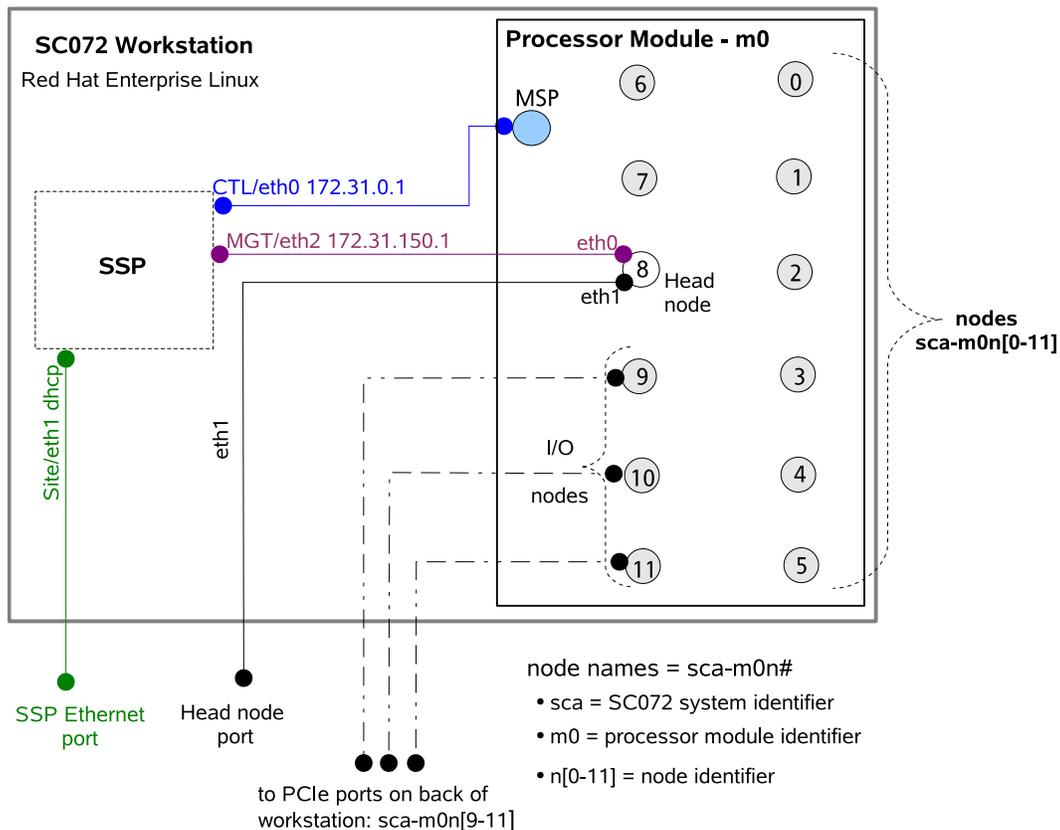


Figure 1. SC072-PDS internal components and their connections.

- SSP (System Service Processor)

Provides the management interface to the system. It provides the mechanism to monitor and control system power and temperature, bootup and shutdown, and error logging and analysis.
- MSP (Module Service Processor)

Under the control of the SSP, the MSP controls, monitors, and reports the environmental state of the system and handles critical communications between the SSP and the nodes.
- Head node

The head node is the entry point to all nodes in the system. On the SC072-PDS, node `sca-m0n8` is the head node.
- Nodes

The system comprises twelve fully interconnected 6-way SMP nodes and either 4 or 8 GB of memory per node, for a total of 48 or 96 GB of memory. Each node consists of six 64-bit MIPs processors, caches and memory system controllers, and the fabric switch circuitry necessary to support interprocessor communications.

12 nodes x 6 processors each = 72 processors
- 2 Ethernet ports

Two Gbit Ethernet ports on the back of the workstation provide network connection to the system.

The SSP Enet port connects to the workstation, and provides the pathway used by the system administrator.

The head port connects directly to the head node: `sca-m0n8` and provides the pathway used by regular users.
- 3 PCIExpress® slots

Three PCIe slots on the back of the workstation provide optional connection to three I/O nodes: `sca-m0n9`, `sca-m0n10`, and `sca-m0n11`. For the location and assignment of these ports, see the *SC072-PDS Quick Start Guide*.
- Room for extra disk drives

The workstation has space for up to six additional industry-standard disk drives. For details on installing the workstation, see the *SC072-PDS Quick Start Guide*.

Software

Red Hat on the SSP The SSP ships pre-installed with Red Hat® Linux version RHEL 5.2.

Gentoo on the nodes The nodes run SiCortex Linux, which is based on Gentoo Linux.

Root file system The root file system resides on the SSP, and is read/write. The system serves the root file system to the nodes via NFS. For more information on the root file system and how it is served to the nodes, see *About the Root File System* on [page 15](#).

SLURM The SC072-PDS uses the Simple Linux Utility for Resource Management (SLURM) for managing resources and running and monitoring jobs. For an introduction to SLURM, see:

<https://computing.llnl.gov/linux/slurm/quickstart.html>

Chapter 2 Setting up the System

In this section:

- [Powering on the System](#)
- [Booting the System's Nodes](#)
- [Logging in to the Head Node as root](#)

Powering on the System

These instructions assume that you have attached a monitor, keyboard, and mouse and connected power and Ethernet to the workstation as described in the *SC072-PDS Quick Start Guide*.

1. Toggle on the Power switch on the back of the cabinet, and wait five seconds for the hardware components to power up.



Do not insert the DVDs in the system. The SC072-PDS ships with the software already installed. The DVDs are provided in case your system becomes damaged at a later date. See *Installing from DVD on page 46*.

2. Turn on the monitor.
3. Press the Power Start button on the front of the cabinet, and then press the Reset button next to it.

If you don't hear a beep within about seven seconds, press the Reset button again.

4. Wait for the *login* screen.

It takes Red Hat about four minutes to boot.

5. Log in as `root` using the default `root` password:

```
Username: root
Password: sicortex
```

You are now logged on as `root` to the SC072-PDS workstation.

- ☀ We strongly recommend that you change the `root` password on the SSP. For details, see *Changing the Root Password* on [page 15](#).

Booting the System's Nodes

You boot the nodes, then you configure the machine and boot the nodes again so your changes take effect. If you prefer, you can do the configuration before you boot the nodes the first time.

1. Open a terminal. Right-click the Red Hat desktop then click **Open Terminal**.

```
[root@<hostname>~] #
```

2. To boot the system nodes, type `scboot` then press **Enter**.

After a few minutes, the boot software starts a counter that counts in seconds to track the bootstrapping process. Boot up should finish before the counter reaches 300 seconds.

For example:

```
secs      kernel  fabric  initfs  slurm
175       12      12     12      12
```

- ☀ If any node fails to boot or errors occur, press **Ctrl C**, then reissue the command, or try `scboot --start_msp=force`.

Logging in to the Head Node as `root`

1. Log in to the head node; type:

```
[root@<hostname>] # ssh head
```

2. The first time you log on after booting, the system issues a warning about the authenticity of host `head`, and asks if you want to continue. Type `yes`.

3. Run `sinfo` to check that all nodes are up and in the `IDLE` state.

```
[root@sca-m0n8] # sinfo
```

```
PARTITION  AVAIL  TIMELIMIT  NODES  STATE  NODELIST
sca        up      infinite   12     idle  sca-m0n[0-11]
```

Chapter 3 Managing the System

In this section:

- [About the Root File System](#)
- [Changing the Root Password](#)
- [Changing Date and Time Parameters](#)
- [Setting up Networking](#)
- [Setting up as a Standalone System](#)
- [Managing User Access to the System](#)
- [Configuring Network File Systems](#)
- [SiCortex System Services](#)

About the Root File System

The SC072-PDS has one root file system for the workstation and another root file system for the twelve nodes. Both root file systems reside on the internal disk of the workstation (the SSP). This section provides a brief description of the root file systems.

Root File System for the Workstation

The root file system (*rootfs*) for the workstation resides in the standard Linux directories on the workstation (ie. the SSP). It creates a standard Red Hat Linux operating environment.

 You should not modify the root file system for the workstation.

Root File System for the Nodes

Many of the changes you make to configure and customize your system are changes you make to files in the root file system for the nodes.

The rootfs for the nodes resides on the workstation in this directory:

```
/opt/sicortex/rootfs/default/
```

As part of the boot process when you run `scboot`, all of the nodes mount the node rootfs that resides in the node rootfs directory on the workstation. Therefore:

- There is only one copy of the node rootfs.
- The node rootfs is accessible read/write from the workstation and from every node.

Editing the Node Rootfs

Since the node rootfs resides on the workstation and is mounted by all nodes, you can edit any file in the node rootfs from the workstation or from any node.

The rootfs for the nodes resides on the workstation in this directory:

```
/opt/sicortex/rootfs/default/
```

- ☀ If multiple users are logged onto your system, there is a risk that two users could experience contention, while trying to update the same file in the node rootfs at the same time.

Alias head for the Head Node

When you boot the SC072-PDS, the alias **head** for the head node is automatically pre-configured by the install process.

Throughout this document, procedures that access the head node refer to it as head. You can use the alias **head** to log into the head node from the SC072-PDS workstation, or from another node on the system.

This alias is available only on the SC072-PDS system itself.

Changing the Root Password

As shipped, the default root password is `sicortex` on both the workstation and the nodes. We strongly recommend that you change it on the workstation and nodes.

- On the workstation** Use the Red Hat desktop GUI tool (**System** → **Administration** → **Root Password**) to change the `root` password on the workstation.

- ☀ The workstation password does not automatically propagate to the nodes.

On the nodes You can change the root password on the nodes to match the workstation root password, or create a unique node password.

1. Log on to **head** and run `/usr/bin/passwd`, and specify the new root password.
2. On the workstation, run `scboot` to boot the nodes. Booting the nodes causes the new node password to take effect.

Changing Date and Time Parameters

You need to set the date and time parameters only if:

- They do not match those at your location, or
- You connect to the site LAN and want to enable Network Time Protocol (NTP) to synchronize the system clock.

Date and time defaults The system ships with these default settings for date and time:

- Time Zone - Eastern Time (UTC)
- Date - factory-set to current date
- Time - factory-set to current time
- Network Time Protocol (NTP) - Disabled

Date and Time tool Use the Red Hat **System** → **Administration** → **Date & Time** tool to change the Date and Time Parameters. The tool has three tabs: **Date & Time**, **Network Time Protocol**, and **Time Zone**.

Setting the date and time Set the workstation date and time only if you do not plan to use NTP. Bypass the *Date & Time* tab if you will be connecting the system to the network and enabling NTP.

Set the workstation date and time before you boot the nodes. When the nodes boot, they take their initial date and time setting from the date on the workstation. If you change the date on the workstation, you must `scboot` to propagate the new date/time to the nodes.

 Clicking **OK** exits the *Date & Time* tool and immediately applies whatever changes you have made to the workstation, but not to the nodes.

Enabling NTP Enabling Network Time Protocol (NTP) disables the *Date & Time* settings. To use NTP, the SSP Ethernet port must be connected to your site's LAN. See the *SC072-PDS Quick Start Guide* for details.

When NTP is enabled, the system clock synchronizes to some external time server, usually the public NTP pool, a local clock, or other time source.

To enable NTP, use the click **System** → **Administration** → **Date & Time** tool, and go to the **Network Time Protocol** tab.

 Do not use the SSP as an NTP server.

Setting the Time Zone Setting the time zone on the workstation does not set the time zone on the nodes.

1. On the *Time Zone* tab, select the time zone and check **System clock uses UTC** to enable the system clock to automatically switch between Standard and Daylight Savings time.
2. From the workstation or any node, run these commands to edit the node root file system to replace the `localtime` file:

```
# rm /opt/sicortex/rootfs/default/etc/localtime
# ln -s /usr/share/zoneinfo/<TIMEZONE> \
/opt/sicortex/rootfs/default/etc/localtime
```

where `<TIMEZONE>` is one of the time zone variables used by the CoporateTime Server open-standards-based calendaring software. The following table shows the time zone variables representing the standard time zones for the Continental United States.

Value	Time Zone
EST5EDT	Eastern Standard Time
CST6CDT	Central Standard Time
MST7MDT	Mountain Standard Time
PST8PDT	Pacific Standard Times

For a complete list of the time zone variables, see:

<http://www.cs.berkeley.edu/CT/ag4.0/appendid.htm#1018805>.

Setting up Networking

The basic steps to set up networking are:

- Connect the workstation (SSP) to the LAN and configure if needed.
- Set up networking for the nodes.
- Specify the DNS servers to use.

Connecting the SSP to the LAN

Connecting SSP Ethernet port

1. To connect the workstation to the LAN, plug an Ethernet cable into the SSP Ethernet port on the back of the SC072-PDS.

For details, see the *SC072-PDS Quick Start Guide*.

Using DHCP

By default, the system software sets up networking for the SSP to use dynamic IP addresses and DHCP. If you are using DHCP, you don't need to configure anything further on the SSP.

 Skip to *Setting up Networking for the Nodes* on [page 20](#).

Using Static IP Addresses

Execute this procedure only if you are using static IP addresses.

To configure the SSP to use static IP addresses:

1. In the Red Hat desktop menu, choose **System** → **Administration** → **Network** to display the **Network Configuration** tool on the **Devices** tab.
2. In the **Devices** tab, select and edit **eth1**. **eth1** is the workstation (SSP) interface you will configure to connect to the site LAN.

 **Edit *only* eth1. Never change the settings for eth0 or eth2.**

Why: **eth0** and **eth2** are preconfigured for the system's internal CTL and MSP networks. The system uses these internal networks to communicate with the nodes. Changing **eth0** or **eth2** would cause serious problems in the system's internal communications.

3. Select **Statically set IP address**, then supply the IP address and any other information required by your site.
 4. Click the **DNS** tab to set up the DNS parameters to match the configuration on your site DNS server.
 - hostname and domain
 - DNS name servers
 - DNS search path
-  Do not change the settings of any other SiCortex components that appear on the **Devices**, **Hardware**, and **Host** tabs.

Setting up Networking for the Nodes

To give the nodes on the SC072-PDS access to the network, you need to connect and configure a router. The choices are:

- the workstation (via the SSP's built-in Ethernet port)
- the **head** node (via the head node's built-in Ethernet port)
- node 9, 10, or 11 (by plugging in a PCIExpress card)

One Available Ethernet Connection

If your site provides only one Ethernet connection to connect the SC072-PDS to the LAN, configure the workstation (SSP) as the router to give both the workstation and the nodes access to the network.

See *Using the SSP as the Router to the LAN* on [page 21](#).

-  If you connect and configure only the **head** node as the router, the nodes will have access to the network but the workstation will not have network access.

Two Available Ethernet Connections

If your site provides two Ethernet connections, it's more efficient to use the **head** node as the router for the rest of the nodes.

See *Using the Head Node as the Router to the LAN* on [page 22](#).

Available Broadband Connection

If your site offers you a broadband connection, you can plug a PCIExpress® card into one of the three interfaces that connect to Nodes 9, 10,

and 11. You can then use the corresponding node as the router to the LAN.

See *Using Node 9, 10, or 11 as the Router to the LAN* on [page 24](#).

Using the SSP as the Router to the LAN

You can enable remote access to the SC072-PDS via the workstation. You set up the SSP on the workstation as the router to the LAN, and configure the nodes to use the SSP as their router.

To do this, your next steps are:

- Ensure the SSP's Ethernet interface is connected to the LAN (already done in *Connecting the SSP to the LAN* on [page 19](#).)
- Configure the nodes to use the SSP as their router to access the LAN (below).
- Configure the SSP to serve as a router (see below).
- Configure the next-hop router on the LAN to recognize the SSP as a router.

Configuring the Nodes to Use the SSP as Their Router

1. On the workstation, save a backup copy of the `/etc/sicortex-system.conf` file and rename it `/etc/sicortex-system.conf.backup`.

See *Appendix B, The `sicortex-system.conf` File* on [page 61](#) for the contents of the default `sicortex-system.conf` file.

2. Edit the `/etc/sicortex-system.conf` file to uncomment the lines shown below, remove any leading spaces, and modify these parameters as shown:

```
sca.cluster.external-network = default
sca.cluster.default-router = ssp
sca.node.sca-m0n8.interfaces = eth0
sca.node.sca-m0n8.eth0.address = dhcp
sca.node.sca-m0n8.router = yes
```

This group of settings does the following things:

- Selects the default external network.
- Makes the SSP the default router.

Using the Head Node as the Router to the LAN

- Tells the node software to route via the interface between `eth2` on the SSP and `eth0` on the `head` node.
- Selects DHCP.
- Specifies that the head node will act as a router for the rest of the nodes.

Note: The `eth0` cited above is the `eth0` interface on the `head` node, `sca-m0n8`. This `eth0` interface is connected at the factory to the `eth2` interface on the SSP. See Figure 1, “SC072-PDS internal components and their connections,” on page 10.

Configuring the SSP to Serve as a Router

To configure the SSP as a router, and set up the appropriate `iptables` rules:

1. Run the following commands. These commands modify the firewall configuration on the SSP to enable routing on the workstation and to define firewall rules that allow packet routing:

```
iptables -t nat -A POSTROUTING -o eth1 -j MASQUERADE
iptables -I RH-Firewall-1-INPUT 1 -i eth2 -o eth1 -j ACCEPT
/etc/init.d/iptables save

echo 1 >/proc/sys/net/ipv4/ip_forward
```

2. Edit the file `/etc/sysctl.conf` to change the following value (from 0 to 1):

```
net.ipv4.ip_forward = 1
```

3. Run `schboot` to reboot the nodes, to make the changes in both the above sections take effect.
4. Skip over the next section and go to:

Setting up Extra Routes on [page 24](#)

or *Managing User Access to the System* on [page 26](#)

Using the Head Node as the Router to the LAN

You can enable remote access to the SC072-PDS via the `head` node. This gives all nodes on the SC072-PDS access to the LAN without going through the SSP. To do this, you must:

- Physically connect the `head` node's Ethernet interface to the LAN.
- Configure the `head` node's Ethernet connection to the LAN.
- Configure the nodes to use the `head` node as their router to access the LAN
- Configure the next-hop router on the LAN to recognize the `head` node.

Connecting the head node's Ethernet interface

To connect the `head` node to the LAN:

1. Plug an Ethernet cable into the `head` Enet port on the back of the SC072-PDS.

For details, see the *SC072-PDS Quick Start Guide*.

Enabling the head node's network interface

Edit the `/etc/sicortex-system.conf` file and specify the address, net-mask, and gateway value for each interface. For instructions, see *Appendix B, The sicortex-system.conf File* on [page 61](#).

1. On the workstation, save a backup copy of the `/etc/sicortex-system.conf` file and rename it `/etc/sicortex-system.conf.backup`.
2. Edit the `/etc/sicortex-system.conf` file to modify the following parameters to have the values shown below, uncomment the lines, and remove any leading spaces:

```
sca.cluster.head-node = sca-m0n8
sca.cluster.io-nodes = sca-m0n8
sca.node.sca-m0n8.eth1.address = dhcp
```

If you want to assign `sca-m0n8` a static address instead of `dhcp`, then edit these lines as shown below instead:

```
sca.cluster.head-node = sca-m0n8
sca.cluster.io-nodes = sca-m0n8
sca.node.sca-m0n8.eth1.address = x.x.x.x
sca.node.sca-m0n8.eth1.netmask = x.x.x.x
sca.node.sca-m0n8.eth1.gateway = x.x.x.x
```

Replacing the x's with the appropriate values.

3. Run `scboot` to reboot the nodes.

Giving all Nodes LAN Access Via the Head Node

You can give all nodes access to the site LAN by routing them through the head node's eth1 port. Routing the traffic for the nodes through the head node is more efficient than routing it through the SSP.

See *Appendix B, The sicortex-system.conf File* on [page 61](#) for the contents of the default configuration file.

1. On the workstation, edit the `/etc/sicortex-system.conf` file to modify these parameters, uncomment these lines, and remove leading spaces:

```
sca.cluster.external-network = default
sca.cluster.default-router = sca-m0n8
sca.node.sca-m0n8.interfaces = eth1
sca.node.sca-m0n8.eth1.address = dhcp
sca.node.sca-m0n8.router = nat
```

2. Run `schoot` to reboot the nodes.

Using Node 9, 10, or 11 as the Router to the LAN

1. Plug a PCIExpress® card into one of the three PCIExpress slots on the SC072-PDS.
2. Execute the same instructions as in *Using the Head Node as the Router to the LAN* on [page 22](#), but in all cases where the instructions specify `m0n8` which is the head node, instead specify either `m0n9`, `m0n10`, or `m0n11`, based on where you plugged in the PCIExpress® card.

Setting up Extra Routes

If you want to set up extra routes for your system to use, edit the following portion of the `sicortex-system.conf` file, using the procedure below.

```
# Specify additional routes (optional). This is a space-separated list of CIDR
# network numbers (a.b.c.d/e). Routes to each of these will be added through
# the per-interface gateway addresses (below) - not the default router address
# (above) - and traffic from internal nodes will be distributed across the
# router nodes.
# sca.cluster.external-network = 10.0.5.0/24 10.6.0.0/16
```

1. In the section of `sicortex-system.conf` shown above, un-comment the `sca.cluster.external-network` line by removing the comment character and leading spaces.

For example:

```
sca.cluster.external-network = 10.0.5.0/24 10.6.0.0/16
```

2. Although the key name is singular, you can specify a space-separated list of networks to add multiple extra routes, if you wish.

On the head node and router nodes, the system will add these routes using the interface-specific gateway addresses, so if you also specified this line in `sicortex-system.conf`:

```
sca.node.sca-m0n8.eth1.gateway = 10.0.0.1
```

Then the effect would be similar to executing the following commands on `m1n6`:

```
route add -net 10.0.5.0/24 gw 10.0.0.1
route add -net 10.6.0.0/16 gw 10.0.0.1
```

On the interior nodes, these routes are added using the router nodes you previously specified as gateways. This distributes the traffic evenly among the routers, as is done for the default route.



Note that the interface-specific gateways used for these specific routes might not be the same as the default router defined in the previous section. This provides maximum flexibility, but using multiple external routers like this increases routing complexity and should be done with caution.

Specifying the DNS Servers to Use

1. Edit the file `/etc/sicortex-system.conf` to set the `sca-cluster.dns-servers` parameter to the addresses of the DNS servers to use on your site, as shown in the example below:

```
# If not running as a self-contained system, the address(es) of
# the dns server(s) to use. NB that there is no specific
# checking to make sure that network routing is correct, etc.
#
# [We should be getting these from dhcp if possible...]
sca.cluster.dns-servers = 10.0.0.23 10.0.0.11 10.0.0.36
```

Setting up as a Standalone System

You can configure the SC072-PDS as a standalone system. *Standalone* means that the system does not depend on the network or the site for any services such as user authentication. The system may or may not be connected to the network.

User accounts On a standalone system, user accounts are administered on the workstation.

- The SC072-PDS may or may not be connected to the LAN for network access.
- The system provides all its own services, such as user authentication.
- The workstation acts as the NIS server for the nodes.
- Users are authenticated through `passwd` files on the workstation which act as the NIS database.
- The nodes act as NIS clients.

File system You can handle file systems in an analogous way, by letting the workstation export part of its file system for the nodes.

For details on configuring the workstation to export part of its file system for the nodes, see *Configuring Network File Systems* on [page 30](#).

Managing User Access to the System

You can set up the SC072-PDS in one of two ways, to authenticate users logging into the system:

- **Via a site server**—If your site already manages user accounts with an LDAP or NIS user authentication server, you can configure the SC072-PDS to validate user logins against that server on your LAN:
 - Using NIS—Network Information Service
 - Or using LDAP—Lightweight Directory Access Protocol
- **Standalone**—If no LDAP or NIS server is available on your LAN, you can configure the SC072-PDS to handle its own user authentication locally.

- Using `passwd` files on the SSP—the SSP acts as an NIS server to the nodes

Configuring User Authentication for a Site Server

This section explains how to configure your SC072-PDS to use a site NIS or LDAP server for user authentication.

Configuring the workstation

1. Use the Red Hat Administration tools (**System** → **Administration** → **Authentication**) to configure the workstation to be either a NIS client or an LDAP client to the site NIS or LDAP server.
2. Set up the nodes accordingly, using one of the procedures below.

Configuring the nodes

Configure the nodes as NIS clients or LDAP clients, depending on the user authentication method you are using.

Setting up NIS Clients on the Nodes

1. Right-click on the workstation and choose **Open Terminal**.

2. Log in to the head node:

```
ssh head
```

3. Change directory:

```
cd /etc
```

4. Edit the file `/etc/nsswitch.conf` to set these keywords to NIS:

```
passwd: db files nis
shadow: db files nis
group: db files nis
```

5. Edit the `/etc/conf.d/net` file, which looks like this:

```
# This blank configuration will automatically use DHCP for
# any net.*scripts in /etc/init.d. To create a more complete
# configuration, please review /etc/conf.d/net.example and save
# your configuration in /etc/conf.d/net (this file :|!).

# Pre-configured by preinit
config_eth0=( "noop" "null" )
config_eth1=( "noop" "null" )

# NIS
nis_domain="example.com"
nis_servers="ssp"
```

6. Set the value of `nis_domain` to the domain name of the site NIS server.

7. Set the value of `nis_servers` to list one or more NIS servers on the site LAN.

8. Run `scboot` to reboot the nodes.

`scboot` propagates the NIS changes you just made on the workstation to the nodes, so that the nodes are pointed at the site NIS server.

- ☀ For users familiar with NIS, using the `-broadcast` switch for `ypbind` won't work because not all nodes, acting as NIS clients, can broadcast to find a server.

Setting up LDAP Clients on the Nodes

1. Right-click on the workstation and choose **Open Terminal**.
2. Log in to the `head` node:

```
ssh head
```

3. Change directory:

```
cd /etc
```

4. Edit the file `/etc/nsswitch.conf` to set these keywords to `ldap`:

```
passwd:  files ldap
shadow:  files ldap
group:   files ldap
```

5. Run `scboot` to reboot the nodes.

`scboot` propagates the LDAP changes you just made on the workstation to the nodes, to make the nodes LDAP clients of the site LDAP server.

Configuring Standalone User Authentication

You can configure your system *standalone*, so that even if it's connected to the network, it does not depend on the site for any services. Rather, all services such as user authentication are handled locally on the SC072-PDS.

Creating User Accounts

For standalone authentication, set up user accounts using `passwd` files. Configure the SSP on the workstation to act as an NIS server and the nodes as NIS clients.

- To create or change user accounts, you use the Red Hat **System** → **Administration** → **Users and Groups** tool. Red Hat propagates the changes to the local NIS server database (the `passwd` file) on the system. "Local" here means the NIS server and database reside on the workstation and serve the nodes.
- The nodes are NIS clients. This means you can use the **System** → **Administration** tools to manage user accounts for the entire system. We recommend that you use the default `/home/<user_name>` for the user's home directory, and make sure the user's Primary Group is `users`.
- Changes you make to Users and Groups are visible on the nodes within five minutes of applying the change (by clicking **OK**). If you want to force the nodes to update faster, run `make -C /var/yp`.
- To access the nodes, first log onto the workstation, then `ssh` to the `head` node.
- The workstation's SSP Ethernet port is preconfigured to use `dhcp`. This means that if your site LAN has a DHCP server and the workstation is connected to it, you can use `ifconfig eth1` to discover the workstation's IP address. Then you can access the workstation and `head` node remotely.
- To set up the system so users can log directly onto the head node over your site's LAN, see:

Enabling the head node's network interface on [page 23](#)

Changing the Default NIS Domain for the Workstation

The default NIS domain for the workstation is `example.com`. To change it to the appropriate NIS domain for your site:

1. On the workstation, edit the `/etc/sysconfig/network` file, and change the value of the `NISDOMAIN` parameter.
2. Also on the workstation, edit the `/opt/sicortex/rootfs/default/etc/conf.d/net` file and change the value of the `nis_domain` parameter.

Alternately, you can edit the `/etc/conf.d/net` file on the head node to change the value of `nis_domain`.

3. To activate the changes, run `scboot` to reboot the nodes.

Configuring Network File Systems

You can configure the SC072-PDS to use the file system on the workstation, or to access one or more external file systems.

- Mounting home directories so users logged into the nodes can see their home directories on a networked file server
- Setting up a Lustre file system
- Setting up and mounting all other external file systems for the SC072-PDS works the same way as on larger SiCortex system models. See the *System Administration Guide*.

Mounting Home Directories

These steps are necessary if you want to enable users to see their home directories located on a networked file server after they log onto the nodes.

The default configuration as shipped includes `/home` as a symlink to `/local/home`.

1. Edit `/etc/fstab` to create an NFS mount on the workstation to mount home directories located on the site file system. For example, at the end of the file, add this line, followed by a new line:

```
<site.file.server>:/vol/home /home nfs \
rw,tcp,noLOCK,rsize=32768,wsizE=32768, \
noatime,hard,intr 0 0
```

2. Change to the directory where the node root file system resides:

```
cd /opt/sicortex/rootfs/default
```

3. Rename the symlink:

```
mv home localhome
```

4. Create a directory called `home`, which is where the node root file system is stored on the SSP:

```
mkdir home
```

Setting up a Lustre File System on Direct-Attached Storage

To set up a Lustre file system on a direct-attached storage device, use the `lustre_catapult.sh` script in the `/opt/sicortex/script_examples/` directory.

For information on setting up Lustre file systems, see the chapter on Lustre file systems in the *System Administration Guide*,

SiCortex System Services

SiCortex runs the following services (Table 1) in the background on the SSP.

 You should never edit, stop or disable these services. These services are enabled/disabled by the SiCortex software.

Table 1. SiCortex-supplied services

Name	Function
envmond	Monitors system environmental sensors.
ev1d	Coordinates the set up of the NFS mounts.
fabricd	Monitors system logs and reports fabric problems
kernmond	Monitors system logs and reports kernel panics and memory ECC errors.
policyd	Monitors system hardware and software status, and sends alerts to Nagios.
scconserver	Serves as an intermediary to conserver. Provides the connection to the MSP.
slurmctld	SLURM control daemon running on the workstation.
syslog-ng	The system logging daemon (syslogs).
watchdogd	Checks that envmond and policyd are running.

Table 2. Required native services

Name	Function
autofs	Supports LDAP autofs maps. LDAP use flag is enabled. Also helpful for mounting site file systems.
conserver	Implements the console command, and enables communication with the console on every node.
dnsmasq	DNS forwarder and DHCP server.
local	
iptables	
xinetd	

Services Started on the Nodes

A standard set of Linux services runs on the SC072-PDS nodes. These services include `autofs`.

Chapter 4 Using the System

This chapter explains how users can log onto the SC072-PDS, compile applications, and run jobs.

In this section:

- [Logging on to the System](#)
- [Managing Jobs and Resources on the System](#)
- [Compiling and Running Applications on the Nodes](#)

Logging on to the System

Depending on how the system administrator has set up the network access to the head node, you can either:

- Log on locally through the workstation console
- Log on remotely via the workstation and then log onto the `head` node
- Log on remotely to the `head` node directly

Logging in Locally Through the Workstation

You can log into the system locally through the workstation console. No extra cable connections or configuration steps are required.

You can log in as `root`, or as any username and password that have been created on the workstation. The example below shows the default `root` password. It is recommended that the default `root` password be changed for security purposes:

```
Username: root
Password: sicortex
```

Logging in Remotely via the Workstation to the Head Node

Configuration

Before remote users can log onto the workstation and from the workstation, log into the `head` node, the SC072-PDS system administrator must do the following:

- Connect the SSP Ethernet port on the back of the workstation to the site LAN with an Ethernet cable.
- Execute the configuration commands to put the SC072-PDS workstation on the site network.

Logging In

Use the same user name and password you normally use to log onto your site network.

1. Log on to the SC072-PDS workstation. If the system administrator has assigned a hostname, use it. Otherwise use the default workstation hostname, `mss-ssp`, or the IP address assigned to the workstation. For example:

```
user@ws101$ ssh <ws_hostname>
```

2. If you are using `ssh`, the first time you log onto the workstation, the system issues a warning about the authenticity of the workstation IP address, and asks if you want to continue. Type `yes`.
3. Enter your user password:

```
Password: <your_password>
```

```
user@ws_hostname$
```

4. Because you are already logged into the workstation, you can log on to the head node using the internal alias or node name. For example:

```
<user>@<ws_hostname>$ ssh head
```

or

```
<user>@<ws_hostname>$ ssh sca-m0n8
```

5. The first time you log onto the `head` node, the system issues another warning about the authenticity of head, and asks if you want to continue. Type `Yes`, then **Enter**.

6. Enter your user password:

```
Password: <your_password>
```

```
sca-m0n8 <user_name> $
```

7. cd to your /home/<user_name> directory

```
sca-m0n8 ~ $ cd /home/<user_name>
sca-m0n8 <user_name> $
```

Now you're ready to compile programs or run jobs.

Logging in Remotely to the Head Node

Configuration

Before remote users can log directly into the head node from elsewhere on the LAN, the SC072-PDS system administrator must do the following:

- Connect the **head** node port on the back of the workstation to the site LAN with an Ethernet cable (beside the port marked "SSP").
- Execute the configuration commands to put the head node Ethernet interface on the site network.
- Configure the site DNS server with a logical name for the head node `sca-m0n8`.

Logging In

You should be able to use the same user name and password you normally use to log onto your site network. This assumes the system administrator has set up and enabled NIS or LDAP on the SC072-PDS.

1. From outside the SC072-PDS, when you log onto the head node, you must use the logical name configured on your site LAN. For example:

```
<user@ws101>$ ssh <name-of-SC072-head-node>
```

2. If you use `ssh`, the first time you log onto the **head** node, the system issues a warning about the authenticity of host **head**, and asks if you want to continue. Type **Yes**, then **Enter**.
3. If there is no SSH key in your home directory, you'll need to enter your user password:

```
Password: <your_password>
```

```
sca-m0n8 ~ $
```

4. If the system has been configured according to the previous chapters in this book, you will already be in your home directory.

If not, `cd` to your home directory:

```
sca-m0n8 ~ $ cd /home/<user_name>
```

```
sca-m0n8 <user_name> $
```

Managing Jobs and Resources on the System

About SLURM

The SC072-PDS uses the Simplified Linux Utility for Resource Management (SLURM) for managing jobs and resources.

SLURM commands require that you specify a SLURM partition when you run a job. The SLURM partition specifies which nodes will be used to run the job. The system provides two default SLURM partitions:

- Base partition—`sca`



`sca` is the base SLURM partition. It contains all of the nodes in the system. Do not delete the `sca` SLURM partition.

- Compute partition—`sca-comp`

`sca-comp` is the predefined SLURM compute partition. It includes all nodes but avoids using `head` (`sca-m0n8`), unless a job requires it. Use `-p sca-comp` when you run user applications.

Running Jobs

Use the SLURM command `srun` to run jobs on the system.

For examples, see *Chapter 2, Running Applications* in the *SiCortex® System Programming Guide*.

Managing Jobs

Use the SLURM commands `sinfo`, `squeue` and `scancel` to monitor node resources and manage jobs launched by `srun`.

For examples, see *Chapter 2, Running Applications* in the *SiCortex® System Programming Guide*.

Compiling and Running Applications on the Nodes

See the *SiCortex® System Programming Guide* for full details on compiling, running, and managing launched jobs.

Compiling Applications

The PathScale compilers are preferable. Unless you need gcc-specific features, the Pathscale compilers generate more efficient code than the GNU compilers.

PathScale Compilers

pathf95	Fortran 77 90 95)
pathcc	C compiler
pathCC	C++ compiler

GNU compilers:

gcc	C compiler
g++	C++ compiler

1. If you haven't already done so, log on to the `head` node, then `cd` to your home directory (see [page 34](#)).
2. Compile, then run your application.

For details, see *Chapter 2, Running Applications* and *Chapter 3, Compiling and Linking Applications* in the *SiCortex® System Programming Guide*.

Running Applications

You launch jobs using SLURM's `srun` command. For examples, see *Chapter 2, Running Applications* in the *SiCortex® System Programming Guide*.

For complete details on how to use SLURM commands, see their man pages: `slurm(1)`, `srun(1)`, `sinfo(1)`, `scancel(1)`, `squeue(1)`, and `salloc(1)`.

By default SLURM distributes one job process (*task* in SLURM terminology) per node, up to the number of nodes specified on the `srun` command line, before it doubles up processes on any node.

SLURM's `srun` command has two major parameters that determine how it distributes jobs across the nodes:

- `-N <number of nodes>`
- `-n <number of processes>`

The `-m cyclic` option assigns processes across nodes in round robin fashion. Users can also choose an arbitrary assignment of nodes by specifying the `-m arbitrary` option.

For example, running `hostname` on all twelve nodes:

```
srun -p sca -N 12 hostname
```

returns all twelve node names. But, to stripe a job across all twelve nodes, use the `-m cyclic` option:

```
srun -p sca -m cyclic -N 12 <executable>
```

Chapter 5 Monitoring the System

In this section:

- [Monitoring the System](#)
- [Viewing System Logs](#)

Monitoring the System

The policy daemon, `policyd`, implements system monitoring, notifying users through entries in log file or Nagios alerts when alert conditions occur.

Conditions that cause alerts include:

- Out of range temperature readings
- Power supply voltages out of tolerance
- Node kernel panics
- Node memory ECC errors
- Node communications fabric problems
- Problems with the monitoring system (for example, `policyd` not running)

☀ Log file entries do not correspond one-to-one with Nagios alerts, and in general, the log file provides more detail than Nagios alerts.

Depending on the conditions it encounters, `policyd` may send notification to Nagios, reboot an individual node, or shut down the Processor module. `policyd` always logs events in `policyd.log`.

Depending on the nature and severity of the problem, the workstation administrator can decide to reboot one or more nodes or declare one or more nodes (or links) disabled to exclude them from the boot process until they can be fixed.

The only Field Replaceable Units (FRUs) on the SC072-PDS are the DIMMs. If any other component, such as a node, has persistent failures, contact SiCortex Customer Support to arrange for repairs.

Nagios Alerts

The SC072-PDS is designed to send alerts to a Nagios server running on the site LAN.

You must set `--nagios-server` to the Nagios server running on your site LAN in both `/etc/conf.d/policyd` and `/etc/conf.d/watchdogd` for Nagios operation:

```
POLICYD_OPTS="--nagios-server <nagios-server.company-com>  
WATCHDOG_OPTS="--nagios-server <nagios-server.company-com>
```

These alerts are delivered to Nagios as passive service checks. For example:

```
***** Nagios *****  
Notification Type: PROBLEM  
Service: SiCortex Power  
Host: msp-ssp  
Address: 10.4.0.24  
State: WARNING  
Date/Time: Tues Aug 03 04:33:42 EST 2008  
Additional info:  
1196270098 env sca-ssp0 MspEnv_Power_Po1_02 948 mV
```

Policyd Log File Entries

Log file entries from `policyd` vary depending on the source and on the event or condition. Here are some examples that could appear in `/var/log/policyd.log`:

```
msp-ssp log # tail policyd.log  
[2008-07-23 13:16:36] WARNING: sicortex.policyd.Temperature: TIMEOUT on sca-ssp0  
MspEnv_Temp_AD_04: no reading for 180 seconds  
[2008-07-23 13:16:36] WARNING: sicortex.policyd.Temperature: TIMEOUT on sca-ssp0  
MspEnv_Temp_AD_05: no reading for 180 seconds  
[2008-07-23 13:30:14] INFO: sicortex.policyd: 1216834213 env sca-ssp0 MspEnv_Temp_Node_05 45250  
mC  
[2008-07-23 13:30:14] INFO: sicortex.policyd.Temperature: RECOVERY on sca-ssp0  
MspEnv_Temp_Node_05  
Why? 1216834213 env sca-ssp0 MspEnv_Temp_Node_05 45250 mC
```

 mC designates thousandths of a degree Celsius.

Viewing System Logs

Use the **System** → **Administration** → **System Log** tool on the workstation to view the system logs listed here:

Log	Location and Format	Description
Workstation syslogs	/var/log/messages-<yyyymmdd>	Syslog messages from the workstation accumulate in this log. A new log file is created daily. The version of software that booted the System is recorded in this log.
MSP syslogs	/var/log/msp-messages-<yyyymmdd>	This log contains the combined console and syslog messages for the MSP. During the preinitialization phase of the boot process, node messages are recorded here. Some temperature, voltage, and current information collected by the MSP may also appear in this log. However, the main log to check for environmental events is policyd.log.
Node syslogs	/var/log/nodes-<yyyymmdd>/sca-m0n[0-11].log	Each node has its own syslog, and messages apply to the entire node. Indicates whether a node is up or down.
MGTnet state	/var/log/mgtstate/sca-mgt0.log	Logs state of mgtnet port on msp0.
scboot	/var/log/scboot.log	Logs information about the last scboot.
Diagnostics	/var/log/diagcomm_client.log	Logs MSP RPC protocol messages generated by envmond.
MSP environmental stats	/var/log/msp-env/<yyyymmdd>	If present, logs MSP environmental syslog data. Includes temperatures, voltages, and current readings from temperature sensors and PoL regulators on the Processor module collected by the MSP.
MSP console	/var/log/msp.console	Logs the output of the MSP console.
msp0 state	/var/log/mspstate/sca-msp0	Logs the runtime state of the SSP; reported by dnsmasq, the MSP's uBoot loader, and the MSP's uClinux dhcp client.
Raw power data	/var/log/RRD/Power/sca-msp0	Logs the raw power data.
Raw temp data	/var/log/RRD/Temperature/sca-msp0	Logs the raw temperature data.

Policyd Log File Entries

Log	Location and Format	Description
sca.vars	/var/log/scboot/sca.vars	Dump of the shell variables and functions that were defined when scboot ran.
scconserver	/var/log/scconserver.log	Generated by the SiCortex console server multiplexer, scconserver.

Use the Linux `tail` command to view the system logs listed below:

Log	Location and Format	Description
Diagnostics	/var/log/sca/diagcomm.log	Logs MSP RPC protocol messages generated by scboot.
MFD	/var/log/sca/mfd.log	The Master Fabric Daemon logs all messages related to the hardware components implementing the interconnect fabric in this log.
Node console	/var/log/sca/sca-m0n[0-11].console	Conserver creates one console log for each node that contains the log of the node's virtual console. It logs ECC errors and kernel panics from the node and output from the startup scripts.
Master clock agent	/var/log/sca/master_clock_agent.log	The master clock agent supervises the synchronization of kernel timer interrupts and <code>MPI_WTime()</code> as part of the process of booting the nodes. In the stdout of <code>scboot</code> , the message "global clock sync complete" displays in the normal case. If problems occur, this message will instead say "global clock sync failed: <message>", and the details of the problem can be found in <code>master_clock_agent.log</code> .
SLURM jobs	/var/log/slurm-comp.log	All messages regarding SLURM jobs are logged here.
envmond	/var/log/envmond.log	Logs messages concerning the operation of the envmond daemon, such as startup, shutdown, internal errors, and so forth.
ev1d	/var/log/ev1d.log	Cluster event monitoring: barriers, sequencing, locking
kernmond	/var/log/kernmond.log	Logs messages concerning the operation of the kernmond daemon, such as startup, shutdown, internal errors, and so forth.

Log	Location and Format	Description
mspenv	/var/log/mspenv.log	Generated by mspenv (a command line utility used by envmond to communicate with the MSP). Logs the process of making RPC calls to the MSP, as opposed to the results of the calls.
policyd	/var/log/policyd.log	Logs messages from policyd about system monitoring conditions and events. policyd is main environmental monitoring daemon on the system. Major temperature, voltage, and current events are logged to this file.
powerutil	/var/log/powerutil.log	Logs status of the (Valere) power supply generated by the command line power utility, powerutil.
watchdogd	/var/log/watchdogd.log	Logs messages concerning the operation of the watchdogd daemon, such as startup, shutdown, internal errors, and so forth.

You can use the `tail` command on any log file to check the status of an operation.

<code>tail <file.log></code>	Shows the last few lines in the log file
<code>tail -n 100 <file.log></code>	Shows the last 100 lines in the log file
<code>tail -f <file.log></code>	Shows the end of the log file in real time and displays new lines as they are written.

Policyd Log File Entries

Chapter 6 Updating System Software

In this chapter:

- [Checking Your Current Software Version](#)
- [Installing a New Release](#)
- [Installing from DVD](#)

Checking Your Current Software Version

To check the current software version installed on the system, look at the `/etc/redhat-release` and `/etc/sicortex-release` files. For example:

```
[root@PDS ~]# cat /etc/*-release
Red Hat Enterprise Linux Client release 5.2 (Tikanga)
SiCortex Release 5.1.0_rc23-*r98* (V3.0 (RedHat, Andrea) RC23)
```

Installing a New Release

New systems ship with all software pre-installed, so there is no need to load the DVDs.

Follow the procedure below if you need to:

- Update to a new release on an SC072-PDS you already own.
- Reload the system software if your system has become damaged.

Install DVDs

The DVD distribution consists of two DVDs:

- DVD-1 SiCortex Red Hat® Enterprise Linux® 5 Software
- DVD-2 SiCortex Workstation and Node Software

Installing from DVD

☀ **This procedure replaces all information on the disk, so back up all key data and configuration settings before starting this installation.**

To install the software from DVD:

1. Ensure the system is powered up.
2. Place DVD-1 into the SC072-PDS drive and press the reset button on the front panel.
3. If there is a "kickstart" error, click "reboot."
4. Wait for the procedure to finish. After about 25 to 30 minutes, the Congratulations screen will request a reboot.
5. Click the reboot screen button, listen for a beep, and then remove DVD-1 immediately. (If you delay, the **Red Hat®** Enterprise Linux 5 software install process repeats.)
6. Allow the reboot to continue. (Ignore crash kernel, eth and NFS failure messages; these are benign.)
7. Log in as username: **root** password: **si cortex**
8. Place DVD-2 in the DVD drive.
9. Wait 15 seconds for a pop-up window to appear, but do not interact with it.
10. Right-click the desktop, and open a terminal from the pop-up menu.
11. Type **bash /me**(press Tab) **S**(press Tab) **I**(press Tab), then press ENTER. The system completes the correct command for you. For example:

```
bash /media/SiCortex\ Linux\ 3.1\ \ (r105)\ \ 2_2/INSTALL.sh
```
12. When it is done (30 minutes) the machine reboots automatically.
13. When it reboots, remove DVD-2.

14. Log in:

```
username: root
password: sicortex
```

15. Right-click the desktop to open another terminal.

16. Boot the nodes:

```
scboot
```

17. After a few minutes of setup plus about 200 seconds on the timer, the nodes should be done booting.

18. Type `sinfo`. Twelve nodes should be `idle`.

19. Test the installation. Log into the `head` node (`sca-m0n8`) and `ping` a node:

```
ssh head
```

```
password: sicortex
```

When the authenticity of the host is challenged and you are asked if you want to continue, reply Yes.

Then `ping` any node:

```
ping -c 1 sca-m0n5
```

You should see that `ping` statistics for that node.

20. If the installation proves successful, restore any data you backed up to its proper location.

21. Run a SLURM job. For example, if you run `hostname`:

```
srun -p sca -N12 hostname
```

You should see all 12 hostnames. They will not appear in numerical order.

Chapter 7 Troubleshooting

In this section:

- [Diagnosing Failed Nodes and Links](#)
- [Declaring Nodes or Links Disabled](#)
- [Restoring Disabled Components](#)
- [Cancelling Jobs](#)

Diagnosing Failed Nodes and Links

You can tell when nodes are down by querying SLURM using the `sinfo` or the `squeue` command.

Both commands report unresponsive nodes by appending an asterisk to the last reported state of the node. For example, `idle` means that the node is idle and responsive. `idle*` means that the node last reported its state as idle, but is currently unresponsive.

Figure 2, “Steps for diagnosing failed nodes and links,” on page 50 shows the steps for diagnosing failed nodes and links in a flow chart format.

For a more extended discussion of dealing with disabled components, see [Concept—Mapping the Interconnect Fabric](#) and [Concept—Declaring Disabled Nodes, Modules, Links](#) in the *System Administration Guide*.

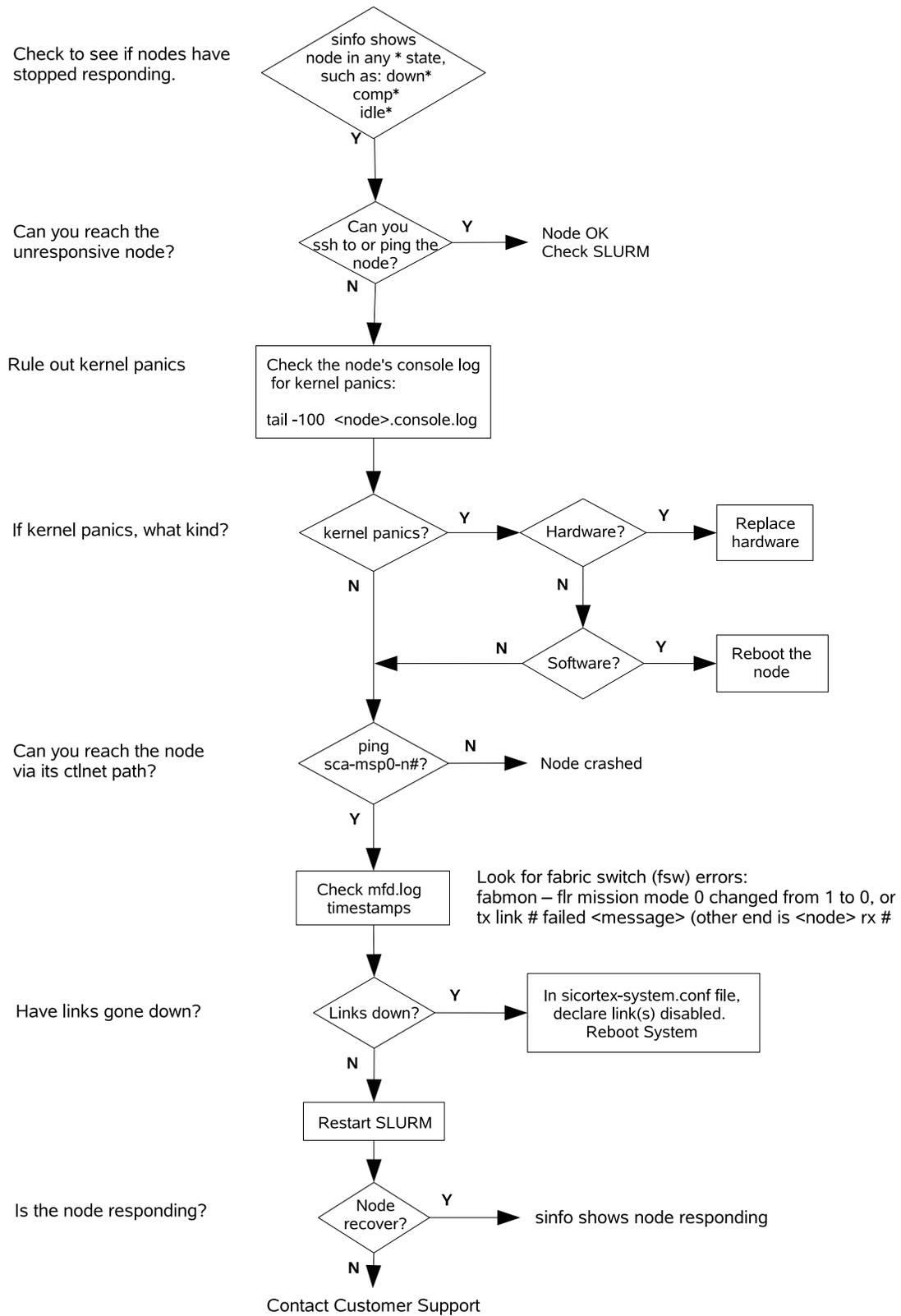


Figure 2. Steps for diagnosing failed nodes and links

Declaring Nodes or Links Disabled

Occasionally nodes or links may malfunction, such that affected nodes fail to boot or they become unreachable through `ssh`.

`scboot` uses the `scbootmon` utility and `mfd_watcher` to monitor and report on the progress of the boot process. The following example shows a boot process that fails:

```
Loading and booting linux
bamf: Loading vmlinux
bamf: Loading bootk
Finished loading linux (kernel boot initiated)
-----scboot-monitor-----
secs  kernel fabric initfs slurm
  30     3     0     0     0
err: all 3 nodes checked in, but no router available (none are gateways) - BOOT FAILED
  87     3     3     0     0
```

1. After `scboot` finishes, you can type the SLURM `sinfo` command to check which nodes have booted successfully:

While `scboot` is executing, nodes will still be going through the boot process. During this time the nodes are shown as having the states `down`, `down*` and `idle*` until they become available.

When `scboot` finishes, you can use the SLURM `sinfo` command to check if all of the nodes are finished booting. The fully booted state, where all nodes are ready to run jobs via SLURM, looks like this:

```
sysadmin@ssp ~ $ sinfo -p sc1
PARTITION AVAIL  TIMELIMIT NODES  STATE NODELIST
sca        up    infinite   12   idle  sca-m0n[0-11]
```

If you see a state like the one below, showing `down`, `down*` or `idle*` instead of `idle`, then the nodes are not currently in contact with the main SLURM controller, `slurmctld`, most likely because they are still booting:

```
sysadmin@ssp ~ $ sinfo -p sc1
PARTITION AVAIL  TIMELIMIT NODES  STATE NODELIST
sca        up    infinite   12   idle* sca-m0n[0-11]
```

```
sysadmin@ssp ~ $ sinfo -p sc1
PARTITION AVAIL  TIMELIMIT NODES  STATE NODELIST
sca        up    infinite   12   idle* sca-m0n[0-11]
```

If you run this command repeatedly (or use the `sinfo -i` flag, causing it to run continuously), you will see nodes assume the `idle` state over time. If a small set of nodes fails to reach the `idle` state

significantly after the rest, these nodes may be experiencing problems.

```
sysadmin@ssp ~ $ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
sca        up    infinite    1   down  sca-m0n1
sca        up    infinite    2   down* sca-m0n3,sca-m0n5
sca        up    infinite    9   idle  sca-m0n0,sca-m0n2,sca-m0n4,sca-m0[n6-11]
```

Wait for sbboot to finish

sbboot provides feedback on how many nodes have reached each stage of the boot process. It exits when the nodes are ready to be used for running jobs.

When nodes persistently are unable to boot and reach the `idle` state, you declare them disabled in the `sicortex-system.conf` file, so `sbboot` excludes them the next time you boot the system.

To declare nodes or links disabled:

1. Edit the `Other cluster-wide configuration` section of the `sicortex-system.conf` file. Uncomment, remove leading spaces from, and set the following parameters:

```
sca.cluster.disabled-nodes = m0nN
```

and

```
sca.cluster.disabled-links = m0nN -rxN |txN
```

2. You can specify either the receiving link (`rxN`) or the transmitting link (`txN`).

For example:

```
# -----
# Other cluster-wide configuration

# If some nodes are broken or misbehaving, they can be
# masked out of
# the system configuration by adding them to this list.
#
# Disabled nodes and links
sca.cluster.disabled-nodes = m0n4, m0n2
sca.cluster.disabled-links = m0n3-rx1
# -----
```

When you reboot the nodes, `sbboot` informs SLURM (`slurmctld`) which nodes and links are disabled (if any), and SLURM puts those nodes in the `Drain` state.

If a job specifies `-N <node_number>` that exceeds the number of nodes in the `idle` state, SLURM will still queue jobs on nodes in the `drain` state and wait for those nodes to return to the `idle` state. Because this could take some time, users should be warned of nodes or links that have been disabled.

The `sinfo` command shows the state of nodes in a partition:

```
$ sinfo -p sca-comp
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
sca       up    infinite   6   idle sca-m0n[0,7-11]
sca       up    infinite   2   drain sca-m0n[2,4]
sca       up    infinite   4   alloc sca-m0n[1,3,5-6]
```

Restoring Disabled Components

If a node persistently fails on your SC072-PDS, and the system is under warranty, return the system to SiCortex for repair or replacement. Nodes are not field-replaceable units (FRUs) on the SC072-PDS.

If one or more DIMMs persistently fail, DIMMs can be replaced in the field. Contact SiCortex Customer Service for assistance.

Cancelling Jobs

You may sometimes find that a job is not executing properly and you need to cancel it.

When you cancel a job using the `scancel` command, you may notice that the nodes do not return to `idle` immediately. For example:

```
sysadmin@ssp015:~>squeue
JOBID PARTITION  NAME      USER  ST      TIME  NODES NODELIST(REASON)
   80      sca     sleep  sysadmin  R       0:06     4  sca-m0n[0-11]

sysadmin@ssp015:~>scancel 80

sysadmin@ssp015:~>squeue
JOBID PARTITION  NAME      USER  ST      TIME  NODES NODELIST(REASON)
   80      sca     sleep  sysadmin  CG       0:06     4  sca-m0n[0-11]

sysadmin@ssp015:~>sinfo -p sca
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
sca       up    infinite   4   CG  sca-m0n[0-11]

later...
```

Cancelling Jobs When the System Is Working

```
sysadmin@ssp015:~>slinfo -p sca
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
sca         up    infinite    4    idle sca-m0n[0-11]
```

SLURM jobs pass through a series of states, typically:

PENDING (PD) → RUNNING (R) → COMPLETING (CG) → COMPLETED (CD)

See the `squeue(1)` man page for more information on job states.

Jobs that are PENDING or RUNNING may be canceled with the `scancel` command.

However, when you cancel a job, it does not immediately vanish from the system. Instead, it advances to the COMPLETING state for any post-processing or cleanup steps.

`squeue` continues to show the job until it leaves the COMPLETING state.

If you try to `scancel` a job that is in the CG state, SLURM gives you the message:

```
scancel: error: Kill job error on job id 61: Job/step
already completing or completed
```

This can be confusing to users, who intuitively think that if they cancel a job it should disappear immediately.

The rationale for handling cancelled jobs this way is that SLURM needs to track the state of every node in every job to make sure that resources allocated to the job are properly released.

There are two common scenarios for using `scancel`, described below.

Cancelling Jobs When the System Is Working

You may have the need to cancel jobs when the system is working fine. The cluster is working, the fabric is working, the nodes are working, SLURM is working, and the user's job is working. The user decides—for whatever reason—to cancel the job. The user does an `scancel`, the nodes go to CG, and a few minutes later the job is all gone.

Cancelling Jobs When the System Is Having Problems

The more urgent use for `scancel` is when some part of the system is having problems—cluster, fabric, nodes, SLURM, network, or a job. The system administrator starts cancelling jobs, trying to get the cluster back to a stable, idle state. The jobs go to CG, but something is broken, so the nodes can't report job completion back to `slurmctld` on the SSP, so SLURM keeps the jobs in CG, waiting for completion messages that aren't arriving. There are timeouts that will eventually remove jobs from CG, but depending on how you configured SLURM, they may be very long (hours) or infinite.

It is unfortunate that when things go wrong and you most need `scancel` to just delete all the jobs off the system and make your nodes available again, `scancel` is least likely to do what you want.

If you are sure that things really are broken and you just want to return the nodes to `idle`, the recommended best practice is to do the following.

You must be logged in as `root` to execute these commands:

1. Type:

```
squeue
```

to get a list of nodes that are stuck in the CG state.

2. Type the following commands:

```
scontrol update nodeName=sca-m0n1 state=down
scontrol update nodeName=sca-m0n1 state=resume
```

for each stuck node. When the node state goes to `down`, SLURM immediately terminates all jobs on the node.

You can use SLURM wildcards in these commands, for example:

```
sca-m0n[1-6]
```

When the node state goes to `resume`, SLURM sets its state to `idle`.

Cancelling Jobs When the System Is Having Problems

Appendix A System Administration Defaults

The SC072-PDS workstation ships with the following **System** → **Administration** configuration parameters.

Only the parameters that are set by SiCortex software or are directly related to setting up and configuring the system in the supported modes are included in these tables. All other values are the standard Red Hat defaults.

Table 3. Authentication

Tab	Parameter	Value
User Information	NIS	disabled
	LDAP	disabled
Authentication	LDAP	disabled
Options	Use shadow passwords	enabled
	Password hashing algorithm	enabled

Table 4. Date & Time

Tab	Parameter	Value
Date & Time	Date	enabled
	Time	enabled
Network Time Protocol		
	Enable NTP	disabled
	Advanced Options	
	Synchronize system clock	disabled
	before starting services	
	Use Local Time Source	enabled
Time Zone	Time Zone	New York
	System clock uses UTC	enabled

Table 5. Network

Tab	Device	Parameters	Value	
Devices	eth0	Activate device when computer starts	disabled	
		Statically set IP address	address: 172.31.0.1	
			subnet mask: 255.255.255.0	
	eth1	Activate device when computer starts	enabled	
		Automatically obtain IP address setup with:	dhcp	
		Automatically obtain DNS information from provider:	enabled	
		eth2	Activate device when computer starts	disabled
			Statically set IP address	address: 172.31.150.1
				subnet mask: 255.255.255.0
Hardware	eth0-eth2	Realtech Semiconductor Ethernet		
DNS		Hostname	localhost.localdomain	
		Primary DNS	10.0.0.23	
		Secondary DNS	10.0.0.11	
		Tertiary DNS	10.0.0.36	
		DNS Search Path	sicortex.com	

Tab	Name	Alias	IP Address
Hosts	misp-ssp	misp-ssp.scsystem	172.31.0.1
	sca-misp0	sca-misp0.scsystem	172.31.0.100
	sca-misp0-n0		172.31.100.100
	sca-misp0-n1		172.31.100.101
	sca-misp0-n2		172.31.100.102
	sca-misp0-n3		172.31.100.103
	sca-misp0-n4		172.31.100.104
	sca-misp0-n5		172.31.100.105
	sca-misp0-n6		172.31.100.106
	sca-misp0-n7		172.31.100.107
	sca-misp0-n8		172.31.100.108
sca-misp0-n9		172.31.100.109	
sca-misp0-n10		172.31.100.110	

Tab	Name	Alias	IP Address
	sca-m0n11		172.31.100.111
	sca-m0n0	sca-m0n0.scsystem	172.31.200.200
	sca-m0n1	sca-m0n1.scsystem	172.31.200.200
	sca-m0n2	sca-m0n2.scsystem	172.31.200.200
	sca-m0n3	sca-m0n3.scsystem	172.31.200.200
	sca-m0n4	sca-m0n4.scsystem	172.31.200.200
	sca-m0n5	sca-m0n5.scsystem	172.31.200.200
	sca-m0n6	sca-m0n6.scsystem	172.31.200.200
	sca-m0n7	sca-m0n7.scsystem	172.31.200.200
	sca-m0n8	sca-m0n8.scsystem head	172.31.200.200
	sca-m0n9	sca-m0n9.scsystem	172.31.200.200
	sca-m0n10	sca-m0n10.scsystem	172.31.200.200
	sca-m0n11	sca-m0n11.scsystem	172.31.200.200
	mgt0-ssp0	ssp0 ssp ssp.scsystem	172.31.150.1
	sca-mgt0	sca-mgt0.scsystem	172.31.150.100

Table 6. Security Level and Firewall

Tab	Parameter	Value
Firewall	Firewall	enabled
	Trusted services: sshx	enabled
SELinux	SELinux setting	enforcing

Table 7. Symlinks

Symlink	Points to	Explanation
home	/local/home	By default, the installed system software creates /local on the workstation, and also exports it to the nodes. The home symlink to /local/home allows the user to log on to the workstation or the nodes as himself or herself, and see his or her home directory in both cases.

Services

Tab	Parameter	Value
Background Services	conserver	disabled
	dnsmasq	disabled
	envmond	enabled
	ev1d	disabled
	kernmond	disabled
	policyd	disabled
	scconserver	disabled
	slurmctld	enabled
	sshd	enabled
	syslog	enabled
	syslog-ng	enabled
	watchdog	disabled
	xinetd	disabled
	yplib	disabled
	yppasswdd	enabled
	ypserv	enabled
	ypxfrd	disabled
yum-updatesd	enabled	

System Log

Tab	Parameter	Value

These are listed in the Viewing System Logs topic

Users and Groups

Tab	Parameter	Value
Groups	slurm	-----

Appendix B The sicortex-system.conf File

The settings in the `sicortex-system.conf` file determine much of the configuration of the system. For details on how to modify these settings, see *Chapter 3, Managing the System* on [page 15](#).

The contents of the `sicortex-system.conf` file, as installed, are:

```
#
# This is the main configuration file for the SiCortex system.
#   - Contains options specific to the SiCortex software release
#   - For options that begin with a partition, you must reboot the
#     partition for changes to take affect.
#   - /etc/sicortex-system.conf should be a symlink to
#     /var/state/etc/sicortex-system.conf. This allows the system
#     configuration to be persistent across different software release
#     installs on the same SSP.

# This header is necessary for the config parser module. Do not alter.
[DEFAULT]

#
# Warning! When editing this file, use comments only at the beginning
# of a line! Trying to do something like
#
# keyword = value    # comment
#
# will not do what you want!
#
#-----
# Basic cluster and boot parameters

#
# The way we serve the node rootfs. For 648/5832 systems, this
# defaults to httpnbd. For smaller systems, it defaults to nfs.
#
sca.boot.rootfs-mode = nfs

#
# The node kernel's log level to use while booting
#
# sca.boot.log-level    = 8

#
# Additional arguments to append to the node kernel command line
#
# sca.boot.append-kargs = initcall_debug

#
# The netblock to use for internal network. Must not be null!
#
sca.cluster.netblock = 172.31

#
# If not running as a self-contained system, the address(es) of the dns
```

```

# server(s) to use. NB that there is no specific checking to make sure
# that network routing is correct etc.
#
# [We should be getting these from dhcp if possible...]
# sca.cluster.dns-servers = 10.0.0.23 10.0.0.11 10.0.0.36

# -----
# Configuration for specialized nodes, ie head nodes etc

#
# Which node is the head node, ie has external ethernet interfaces, users are
# expected to log in to it etc. If not set, there is no designated head node.
#
# IMPORTANT! When specifying a head node, you must also configure their
# network interfaces, in the network configuration section, below.

# sca.cluster.head-node = sca-m0n8

#
# Which nodes are IO gateway nodes.
#
# IMPORTANT! When specifying IO node(s), you must also configure
# their network interfaces, in the network configuration section, below.

# sca.cluster.io-nodes = sca-m0n8

#-----
# Network configuration

# Specify the default gateway for the cluster
# If "ssp" is specified, the SSP is used as the default gateway.
#

# sca.cluster.external-network = default
# sca.cluster.default-router = gw.sicortex.com

#
# Network configuration for exterior nodes
#
# sca.node.sca-m0n8.interfaces = eth1

# This node is a router; values are yes/nat/no
# sca.node.sca-m0n8.router = nat

# If you are using static addresses, you need to specify
# address/netmask/gateway for each interface on each node.
# sca.node.sca-m0n8.eth1.address   = 10.4.2.27
# sca.node.sca-m0n8.eth1.netmask  = 255.255.0.0
# sca.node.sca-m0n8.eth1.gateway  = 10.4.0.1

# If you are using dhcp, you need to specify it once per interface on each
# node
# sca.node.sca-m0n8.eth1.address = dhcp

# -----
# Other cluster-wide configuration

# If some nodes are broken or misbehaving, they can be masked out of
# the system configuration by adding them to this list.
#
# Disabled nodes, modules and links
# sca.cluster.disabled-nodes = m0n4

# -----

```

Index

B

- Base partition
 - described 36
 - sca 14
- Booting the nodes
 - scboot 14
- Booting the system
 - booting the nodes 14
 - default **root** password 13, 33
 - logging onto the head node 14
 - overview 13, 49
 - powering up 13

C

- Changing the **root** password on the SSP 14
- Compiling and running applications
 - logging onto the workstation as a user 34
 - overview 33
- Compiling and running applications on the nodes
 - compiling an application 37
 - managing jobs 36
 - running an application 37
- Compiling applications on the nodes 37
- Connecting the System to a network 16
- Creating user accounts
 - overview 17

D

- default **root** password 13, 33
- DHCP server setup requirements 16

H

- Head node
 - described 10
 - logging on as **root** 14
 - logging on as user 34

L

- Logging on
 - head node as **root** 14
 - head node as user 34
 - Red Hat workstation as user 34

M

- Managing running jobs 36
- Module Service Processor (MSP), described 10

P

- Power start button, location of 13
- Power switch, location of 13
- Powering up the system
 - power start button, location of 13
 - power switch, location of 13

R

- Red Hat workstation
 - default **root** password 13, 33
 - power start button 13
 - power switch 13
 - root** log on 13, 33
 - user log on 34
- Running applications using SLURM 37

S

- sca, base partition 14
- scboot, node boot strapping program 14
- Setting up and managing SLURM partitons and resources 27
- SLURM
 - job management 38
 - launching applications 37
 - monitoring and controlling running jobs 36
 - setting up and managing partitons and resources 27
 - srun** 37
- SSP log in
 - changing the **root** password 14
- System architecture
 - base partition, described 36
 - head node, described 10
 - internal components, diagram of 10
 - MSP, described 10
 - overview 9
- System management
 - connecting to a network 16

- overview 15
- setting up and managing SLURM partitions
and resources 27
- user accounts, creating 17

U

- User documentation

- accessing **iii**
 - location of **iii**

- User log on

- head node 34
 - Red Hat workstation 34