

---

# Software Release Notes for SiCortex Systems

---

**March 25, 2009**

Version 4.0 FT

---

This document describes, in this order:

- [New Features and Other Changes](#)
- [Installing 4.0 FT Software](#)
- [Documentation Issues](#)
- [Active Issues](#) and their temporary workarounds
- [Fixed Issues](#)
- [Documentation Addenda](#)

☼ Be sure to check the SiCortex Knowledge Base and Forums at <http://www.sicortex.com/support> for late breaking information.

## 1.0 New Features and Other Changes

---

---

### Software version 4.0 FT includes these new features

---

- The SiCortex Release 4.0 Transition Guide to aid System Administrators in moving from a previous release to V4.0.
- Red Hat Linux on the SSP.
- Online delivery system allows customers to download incremental updates and patches from SiCortex.
- Updates for the PathScale compilers and the supplied math and science libraries are independent of the system software.

Updates include environment module support to enable users to dynamically load and use different versions of the PathScale compilers and the math and science libraries on the nodes. For the native environment, SiCortex provides a modulefile for each PathScale compiler and each math and science library.

Supplied `eselect` files enable System Administrators to set a system-wide default (or no default) for each version of each math and science library.

- Improved system performance
    - System OS and utilities compiled with gcc 4.2 to improve OS performance, reliability and maintainability.
    - OS memory consumption reduced by 250 MB, enabling larger on-node data sets.
    - Per processor copy of shared-library read-only data to eliminate cache contention when running multiple copies of the same process.
    - Optimized `memcpy` used in kernel to reduce execution time of kernel calls.
  - Improved performance tools
    - Stack Unwinding tool that implements `backtrace` functionality for MIPS.
    - Parallel scripting for `hpcex` and `hpcprof` provides automated processing of multiple output files to produce a unified view of multithreaded results.
    - Improved formatting of Papiex output.
-

---

**Software version 4.0 FT includes these new features (Cont'd)**

---

- New PathScale compiler features
    - Support for software pipelining
    - Support for C pragma prefetch operations
    - Support for Fortran 2003 argument association rules for CHARACTER types
    - Improved stack trace generation
    - Improved general performance, including prefetch generation
  - Updated MPI implementation based on MPICH2 version 1.1a1.
  - Added SiCortex SHMEM API.
  - Support for PVFS2. PVFS2 for SiCortex systems will be available from Argonne National Labs in Q2.
  - Standard software installation includes mounting the n32 versions of the math and science libraries on the nodes when `sbboot` runs.
  - Added GNU math and science libraries: GMP (GNU Multiple Precision Arithmetic Library) and GSL (GNU Scientific Library).
  - New tuned library functions
    - `libscm`: `atan`, `atan2`, `atan2f`, `exp`, `exp2`, `exp10`, `tan`, `tanf`
    - `libscstr/libscfstr`: `sc_bcopy`, `sc_bzero`, `sc_strcat`, `sc_strchr`, and `sc_strlen`
  - A new command, `scshutdown`, on the SSP provides the means for System Administrators, before rebooting, to shut down the nodes in an order they determine. For details, see the *SiCortex<sup>®</sup> System Administration Guide*.
  - Raid 1 supported on the SSP of SC648 and SC1458 systems.
  - The `__SICORTEX__` symbol allows programmers to segregate, in application sources, code specialized to run on SiCortex systems.
-

---

## Software version 4.0 FT includes these changes

---

- The V4.0 `sicortex.conf` configuration file replaces both V3.1 configuration files, `sicortex-system.conf` and `sicortex-install.conf`. The new configuration file, `sicortex.conf`, is located in `/opt/sicortex/etc`. This change heralds new configuration parameters and settings.
- A conversion utility, `cvt-sc-config-files`, provides the means to reproduce any customizations you made to your V3.1 `sicortex-system.conf` and `sicortex-install.conf` files in the V4.0 `sicortex.conf` configuration file.

For more details on configuration changes, see the *SiCortex<sup>®</sup> Release 4.0 Transition Guide*.

- MPI thread safety now supports `MPI_THREAD_SERIALIZED` semantics.
- `mpif90` compiler script automatically links in `-lscmpif90`.
- The `--mpidebuglib` linker command-line option causes all `mpi*` scripts (`mpicc`, `mpicxx`, `mpif77`, and `mpif90`) to use the appropriate debug library.
- The internal MTU decreased from 9,000 to 1,500 to accommodate driver changes in the SSP.
- Added support for the new infiniband option, Sun Microsystems model X4216A-Z DDR IB Host Channel Adapter PCI-e ExpressModule.

Using the X4216A-Z requires the System Administrator to edit the `openib.conf` file to load its MTLX4 driver (or the MTHCA driver if you are using the Mellanox MSEA28-2 Infiniband HCA). For details, see the *SiCortex<sup>®</sup> System Administration Guide*.

- On the SC5832 and SC1458, NBD is the only mode for serving the root file system to the nodes. For details, see the *SiCortex<sup>®</sup> System Administration Guide*.
  - TotalView, TAU, and Vampir are no longer bundled with SiCortex software.
-

## 2.0 Installing 4.0 FT Software

---

### Pre-Installation Tasks

1. Save your existing R3.x customized system configuration files, `/etc/sicortex-system.conf` and `/etc/sicortex-install.conf`, to a location on your network.
2. Make a tarfile of the entire contents of the following directories on the SSP:

```
/opt/sicortex
/etc
/var/state
```

It's important that this tarfile uses relative paths so that users can restore these files without destroying the new system. For example:

```
cd /
tar czf $HOME/sicortex-3.1.tar.gz \
    ./etc ./opt/sicortex ./var/state
```

3. Move the tarfile off the SiCortex system.

If you leave the tarfile on your SiCortex system, it will be inaccessible when you boot R4.0 from the USB drive.

### Connect the R4.0 USB Drive and Boot the System

For Release 4.0 Field Trial, SiCortex provides a USB drive on which Release 4.0 has been installed. This setup allows you to retain your Release 3.x configuration on the SSP.

Follow these steps to connect the USB drive to your SSP and boot Release 4.0 from it:

1. Connect the USB drive to the USB port on the front or back of the SSP.
2. Rest the USB drive on the nearest available flat surface. On the larger systems, place it in the ventilation space above the SSP.
3. Ensure you are logged into the SSP as `root`.
4. Type `init 6` to reboot the SSP.
5. Edit the BIOS.

Watch the messages as the SSP boots. Select the setup option on your model that allows you to edit the order the SSP uses to select its boot device.

6. Move the USB drive to *position 1* in the boot order.
7. Close any Setup windows.
8. Reboot the SSP.

At this point, the SSP reconfigures itself based on the options you selected. Then it boots from the USB drive (where R4.0 is installed).

After a few minutes and some initial messages, you see two or three messages like this:

```
Bringing up interface eth1... [FAILED]
Bringing up interface eth2... [FAILED]
Bringing up interface eth3... [FAILED]
```

These problems will be corrected by a later step in this procedure.

9. When the SSP finishes rebooting, the console displays the Red Hat login screen. Log in as root:

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

10. Verify that the system is booted from the USB drive. Right-click on the Red Hat desktop, open a terminal window, and enter this command:

```
cat /etc/sicortex-release
```

You should see SiCortex 4.0 followed by a build number.

If you do not see this result, reboot the SSP and check your BIOS settings and your connection between the USB drive and the SSP.

11. Execute this command:

```
python /opt/sicortex/bin/sc-fix-nic-names
```

12. Execute this command and enter *y* for all questions.

```
cp /tmp/ifcfg-eth? /etc/sysconfig/network-scripts
```

13. Type this command to shut down the SSP:

```
shutdown -h now
```

14. Wait for the shutdown command to time out, or press the SSP's power button to finish shutting it down.

15. Power-cycle the processor modules to remove old information stored in the MSPs.

SC072-PDS—Press the power button to turn the system off, wait 15 seconds, then press it again to power up the system.

SC648—Lift both latches on all processor modules, then unseat all modules. Wait 3 minutes for the age-out timer to expire on the modules' internal addresses. Reseat and relatch the modules.

SC5832—Press both power breakers on the back of the system to power it down. After 3 minutes, press the power breakers again to power up the system.

16. Press the Power button on the SSP to reboot it.

17. After the SSP has finished rebooting and the Red Hat login window appears, log in as root:

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

The next step converts your R3.x system configuration files to R4.0.

## Convert Your R3.x Configuration Files to R4.0

The `cvt-sc-config-files` program on the R4.0 USB drive converts your R3.0 or R3.1 configuration files to an R4.0 configuration file.

- Input

Your existing, customized R3.X system configuration files:

```
sicortex-system.conf
sicortex-install.conf
```

- Output

A valid R4.0 `sicortex.conf` file that recreates your configuration.

When you booted your system from the R4.0 USB drive, the boot code created a default `sicortex.conf` file. This file contains important settings that you may need.

1. Before you continue, rename and save this file:

```
cp /etc/sicortex.conf /etc/sicortex.conf.default40
```

2. Run the following command:

```
/opt/sicortex/bin/cvt-sc-config-files directory-where-you-saved-your-3.1-config-
files > /opt/sicortex/etc/sicortex.conf
```

to convert your R3.x configuration files and place the output file in the directory where the R4.0 boot code expects to find it.

A few lines will print on `stderr`, either:

```
Cluster type = SC072 (SC648, ...)
```

or

```
Unable to determine cluster type; manual edits required; proceeding.
```

3. If the output reports "Unable to determine cluster type...", edit `/opt/sicortex/etc/sicortex.conf`. Uncomment the line that corresponds to your system model.

```
# machine.model = SC072
# machine.model = SC648
# machine.model = SC1458
# machine.model = SC5832
```

The output may also report either or both of the following:

```
File not found or is empty: sicortex-install.conf; proceeding.
File not found or is empty: sicortex-system.conf; proceeding.
```

If you see these errors, you either didn't enter the correct source directory or you didn't copy the files to that directory. Fix the problem and start again.

### Boot the Nodes

1. Run `scboot` to boot the nodes.

If you converted your R3.x configuration files, the system boots using your converted `sicortex.conf` file, which recreates the same system configuration you established in R3.x.

2. If you had no R3.x configuration files to convert, edit the default `sicortex.conf` file to add system configuration settings.

The available settings are very similar to the R3.x configuration settings. See `man sicortex.conf` for more information.

---

## 3.0 Documentation Issues

---

Issue 9561	V4.0 FT user documentation not included on installation media
Description	<p>The 4.0 FT user documentation is posted on the SiCortex Support web site at <a href="http://www.sicortex.com/support/user_documentation">http://www.sicortex.com/support/user_documentation</a>. The user documents installed with the 4.0 FT software describe V3.x software.</p> <p>Version 4.0 FT of the SiCortex<sup>®</sup> System Administration Guide is not yet fully up to date and ready for release. When ready, it will be posted on the SiCortex Support web site.</p>
Workaround	<p>For differences in administering the system in V4.0 FT, see the <i>SiCortex<sup>®</sup> Release 4.0 Transition Guide</i> on the SiCortex Support web site.</p>

## 4.0 Active Issues

### 4.1 SOFTWARE INSTALLATION Issues

<b>Issue 7884</b>	<b>emerge outputs confusing messages when it runs ldconfig</b>
Description	<p>When you run emerge to install software, you may see messages like this at the end of the install process:</p> <pre>&gt;&gt;&gt; Regenerating /etc/ld.so.cache... /sbin/ldconfig: /usr/lib/libperl.so.1.5.8 is not an ELF file - it has the wrong magic bytes at the start.  /sbin/ldconfig: /usr/lib/libperl.so.1 is not an ELF file - it has the wrong magic bytes at the start.  /sbin/ldconfig: /usr/lib/libperl.so is not an ELF file - it has the wrong magicbytes at the start.  &gt;&gt;&gt; dev-util/boost-build-1.35.0-r1 merged.</pre>
Workaround	Ignore the messages containing ...- it has the wrong magic bytes at the start. These messages are harmless and do not indicate an error occurred.
<b>Issue 9391</b>	<b>Cross-Development Toolkit tarball missing from installation media</b>
Description	The installation media does not include the <code>si-cortex-toolchain-6.0.2.tgz</code> tarball, which contains the Cross-Development Toolkit and installs it on a compliant x86 workstation.
Workaround	If you want to install the Cross-Development Toolkit on a compliant x86 workstation, contact your SiCortex Customer Support representative.

## 4.2 SC072 Issues

<b>Issue 7196</b>	<b>Default hostname incorrect for standalone operation</b>
Description	The default hostname is localhost.localdomain. During bootup, sbboot issues the warning message—Hostname: Host name lookup failure—but completes the boot process. This results in booting an improper configuration.
Workaround	Either connect eth1 to the network before you boot the nodes, or set up the System for standalone operation as described in the <i>SC072-PDS User's Guide</i> .
<b>Issue 7322</b>	<b>sopathcc incorrectly directs user to load a module when it can't find pathcc</b>
Description	When it cannot find pathcc, sopathcc displays this message:  pathcc not found You may need to load a toolchain module. For example, 'module load pathscale-mips'
Workaround	The culprit is a problematic PATH variable that you need to fix.
<b>Issue 9220</b>	<b>SLURM fails to prioritize nodes to avoid using <i>head</i> unless a job requires it</b>
Description	The partition sca-comp includes the head node sca-m0n8. Though node prioritization on sca-comp is configured so that jobs run on sca-m0n8 only when necessary, SLURM ignores this prioritization and treats sca-m0n8 as any other compute node.
Workaround	When appropriate, explicitly exclude sca-m0n8 from the job. For example:  srun -p sca-comp -x sca-m0n8 -n 7 myapp

### 4.3 SYSTEM Issues

<b>Issue 3873</b>	<b><i>syslog chip # attn input buffer full messages can be confusing</i></b>
Description	<p>The MSP continues to buffer console output from the nodes when the SSP services, conserver or scconserver, stop working. When the buffers fill up, the MSP creates a syslog entry in /var/log/msp-messages&lt;year&gt;&lt;month&gt;, like this:</p> <pre>007-09-26T02:51:45-04:00 &lt;kern:warning&gt; scx-msp32 kernel: chip 4 attn input buffer full</pre>
Workaround	<p>If this happens during normal operation, the MSP stops accepting console output from the nodes. If it happens during booting, the boot process may hang.</p> <p>Do not kill or restart scconserver without a specific reason, such as a hanging boot process.</p>

Issue 4475	Default size of core dumps may be insufficient
Description	The System ships with a default size for core dumps that may not be adequate for the needs of some applications.
Workaround	<p>Users can increase the soft limit of core dump files for their application.</p> <p>A safe strategy for users running large rank parallel jobs is to create a wrapper shell script that sets the <code>corelimit</code> to <code>unlimited</code> on only one or a few ranks in such a job:</p> <ul style="list-style-type: none"> <li>To create a wrapper shell script that sets the core dump limit to <code>unlimited</code> on the first ten ranks only and runs the program <code>user_app</code> (and any arguments), put the following code in a file, for example, <code>core.pl</code>: <pre data-bbox="467 814 1372 1081">#!/usr/bin/perl  my \$procID = int \$ENV{'SLURM_PROCID'}; my \$corelimit = 0; if (\$procID &lt; 10){     \$corelimit = unlimited; } exec "bin/bash -c 'ulimit -c \$corelimit; user_app [app_args]'";</pre> </li> </ul> <p>Then invoke the <code>core.pl</code> script to run <code>user_app</code> like this:</p> <pre data-bbox="467 1155 841 1186">srun -p sc1 -n20 ./core.pl</pre> <p>Other alternatives are:</p> <ul style="list-style-type: none"> <li>Overriding the default core file size of processes launched by <code>srun</code> <pre data-bbox="435 1323 1388 1386">ulimit -c unlimited # set the core dump size of the current shell srun -p sc1-comp1 n8 myprogram --propagate=CORE</pre> <p>The <code>--propagate=CORE</code> option causes <code>srun</code> to propagate to the nodes the core dump size (soft limit) of the current shell, which was previously set to <code>unlimited</code>.</p> </li> <li>Setting a C program's core dump size <p>A C program can use the <code>setrlimit()</code> function to set the soft limit of its own core dump size.</p> </li> </ul>

Issue 4563	System may fail to recover if NBD offline too long
Description	<p>An unusable node may indicate problems occurring elsewhere in the System.</p> <p>Transient fabric errors can cause nodes to become unusable. A long lag time (minutes) in fabric recovery can prevent the NBD device driver from recovering properly. Because the root file system on the SC5832 is served by NBD, this situation causes affected nodes to become unresponsive, thus unusable.</p> <p>Other possible causes of unusable nodes are:</p> <ul style="list-style-type: none"> <li>• Kernel software bugs</li> <li>• Transient hardware problems (power glitches, double bit RAM errors, etc.)</li> <li>• Hardware faults (solder joint failures, chip failures, etc.)</li> <li>• Hardware/software problems elsewhere in the System that affect the node's ability to communicate</li> </ul>
Workaround	<p>Depending on the cause, disable or reboot the afflicted node.</p> <p>To diagnose the cause, check the node's console log file, <code>ssp:/var/log/scx/scx-m#n#.console</code>, for</p> <ul style="list-style-type: none"> <li>• Kernel panic messages <p>These usually indicate a software bug, but could indicate transient hardware trouble. For example, spin lock lockups usually indicate a software problem, but are sometimes attributable to a transient hardware fault that hit a processor while it was holding a lock.</p> </li> <li>• ECC-related Kernel panic message <p>DRAM ECC errors can be transient, but if they occur too frequently, disable the node until you replace the faulty DIMM.</p> <p>Non-DRAMM ECC errors tend to be transient or caused by power glitches.</p> </li> <li>• Fabric error messages <p>Also check the <code>ssp:/var/log/scx/mfd.log</code> file for fabric link errors.</p> </li> </ul> <p>(Cont'd next page)</p>

Issue 4563	System may fail to recover if NBD offline too long
Workaround (Cont'd)	<ul style="list-style-type: none"> <li>• NBD messages</li> </ul> <p>NBD messages appearing without fabric messages on the same node typically means there is fabric trouble elsewhere on the path NBD is using. Check by scanning the console logs of other nodes and the <code>ssp:/var/log/scx/mfd.log</code> file.</p> <p>When node hardware, typically memory, is faulty, it's generally best to disable the node, rather than to reboot it. Disabling a particular link has little affect on typical application communications. Disabling links (or nodes) does reduce the redundancy of communication paths available to other nodes, potentially preventing communication-critical applications and benchmarks from reaching their full potential.</p> <p>If you cannot resolve the problem, contact SiCortex Technical Support at <a href="mailto:support@sicortex.com">support@sicortex.com</a>, or call 978.897.0214.</p>
Issue 5985	Policyd power shut-down messages can be confusing
Description	<p>Each time a sensor goes out-of-range, <code>policyd</code> issues a command to shut down the CPU module. If multiple sensors go out-of-range, <code>policyd</code> issues multiple shutdown commands. This behavior doesn't hurt anything—the CPU module remains powered down.</p> <p><code>Policyd</code> logs each shutdown event to <code>/var/log/policy</code>. The first event logged is the one that shut down the CPU module; subsequent shutdown events have no effect.</p>
Workaround	None.

<b>Issue 6635</b>	<b>chrootfs does not clean up manual mounts</b>
Description	<p>Any mount created within <code>chrootfs</code> manually by the user will not be cleaned up when the <code>chrootfs</code> is terminated, and will result in some alarming looking complaints about not being able to clean up the <code>chrootfs</code> mounts.</p> <p>In the future, a solution will be implemented that provides full cleanup logic.</p> <p>Don't forget that in the <code>chrootfs /etc/mtab</code> is a symlink to <code>/proc/mounts</code>, so the mounts don't show up relative to the <code>chroot</code> in there.</p>
Workaround	After you use <code>chrootfs</code> , you must clean up any mounts you created.

<b>Issue 6978</b>	<b>Default NFS I/O sizes set in <code>fstab</code> may be insufficient</b>
Description	The System ships with <code>rsize</code> and <code>wsizes</code> in <code>/opt/sicortex/rootfs/default/etc/fstab</code> defaulted to 1024, which may not be sufficient for some sites.
Workaround	You may want to increase the size of both to 32768. To do so, edit <code>fstab</code> to add the appropriate entry for your site.

<b>Issue 8540</b>	<b><code>powf</code> missing from <code>libstdc++</code> in <code>gcc</code></b>
Description	In previous releases, some functions, such as <code>POWF</code> , were incorrectly included in <code>libstdc++</code> . These functions belong in <code>libm</code> , and in some cases the <code>libm</code> functions are more accurate than their <code>libstdc++</code> counterparts. This may prevent C++ programs compiled or linked under previous releases from running under 4.0FT.
Workaround	If you encounter this problem with an existing C++ program, recompile it to correctly resolve these references.

<b>Issue 8682</b>	<b><code>powerutil</code> doesn't work</b>
Description	The <code>powerutil</code> command, which provides the means to measure a system's power consumption, doesn't operate properly on SC648, SC1458, or SC5832 systems.
Workaround	None.

<b>Issue 8866</b>	<b>Stopping/starting policyd generates no reports in /var/log/watchdogd.log</b>
Description	Stopping/starting policyd generates no reports in /var/log/watchdogd.log
Workaround	None.

<b>Issue 9210</b>	<b>Looped symlinks cause programs that use libraries in /lib64 or /usr/lib64 on the nodes to fail</b>
Description	Natively compiling programs that use /lib64 or /usr/lib64 libraries or running programs that use /lib64 or /usr/lib64 dynamic libraries will fail unless the looped symlinks /lib64/lib64 and /usr/lib64/lib64 are removed.
Workaround	<p>On the SSP, remove the looped symlinks:</p> <ol style="list-style-type: none"> <li>1. <code>cd /opt/sicortex/rootfs/default/lib64</code></li> <li>2. Verify the error condition: <code>cd lib64</code> Should return "Too many levels of symbolic links"</li> <li>3. <code>rm -f lib64</code></li> <li>4. <code>cd /opt/sicortex/rootfs/default/usr/lib64</code></li> <li>5. Repeat steps 2 and 3 for /usr/lib64.</li> <li>6. <code>scboot</code> the nodes.</li> </ol>

<b>Issue 9411</b>	<b>Incorrect Ethernet cabling between nodes and SSP can cause scboot to fail on SC1458/648 systems with 2-port Enet NIC installed in the SSP</b>
Description	<p>If you have an SSP with a two-port add-in Ethernet card and <code>scboot</code> fails with a message like this</p> <pre>all 240 nodes checked in, but no router available</pre> <p>most likely <code>m0n6</code> (port 2 on Processor Module 0) is not connected to the port that the SSP identifies as <code>eth2</code>.</p>
Workaround	<p>To check, run the command <code>ethtool eth2</code> on the SSP, and ensure that <code>Link detected: yes</code> appears on the last line of output.</p> <p>If a link is detected on <code>eth3</code> instead of <code>eth2</code>, reconnect the Ethernet cable to the other port on the SSP's add-in card and rerun <code>scboot</code>.</p>

<b>Issue 9448</b>	<b>Nodes missing correct MODULEPATH for Math &amp; Science libraries</b>
Description	The modulefiles for the math and science libraries are installed in <code>/usr/share/Modules/modulefiles</code> , which is not in the default MODULEPATH, so the module commands cannot find these modulefiles.
Workaround	In your <code>.rc</code> file or in a shell, set the MODULEPATH accordingly:  <code>export MODULEPATH=\$MODULEPATH:/usr/share/Modules/modulefiles</code>

## 4.4 BOOTING Issues

<b>Issue 3763</b>	<b>Recovering from a failed attempt to reboot the MSPs</b>
Description	An attempt to reboot the MSPs—either an automatic reboot by <code>scboot</code> or a manual attempt using <code>scboot</code> 's <code>--start_msp=force</code> flag—may fail because <code>scboot</code> may timeout before the MSPs reboot. This happens because the MSP code in flash has a backoff algorithm for retrying DHCP requests. If the MSP misses two DHCP OFFER packets, it won't try again until after the <code>scboot</code> timeout.  ☀ <code>scboot</code> prints the names of the MSPs that timed out. Output such as:  <pre> ... scx-msp24: rebooted and ready ... scx-msp8: missing (powered off or disconnected) Caught signal, cleaning up </pre> indicates that a number of MSPs rebooted and are ready, but at least one timed out and is considered missing or disconnected.
Workaround	To recover, ping <code>sc*-msp##</code> (where <code>##</code> is the id of the recalcitrant MSP) to make sure the MSP has come back up, then invoke <code>scboot</code> without the <code>--start_msp=force</code> flag.

<b>Issue 4297</b>	<b>Error messages from boot software load failures are confusing</b>
Description	<p>When the boot software fails a parallel load of <code>vmlinux</code>, it returns an error message that looks like this:</p> <pre>sc1-msp2: Failed to load tftp://msp-ssp//scboot_tmp/sc1/vmlinux sc1-msp2: MSP: RPC error: LoadElf error: nodemask = 0x4000</pre> <p>This message means that the boot software failed to complete setup and initialization of the DIMMs on a node, or the setup and initialization completed unsuccessfully.</p> <p>☼ The <code>nodemask</code> value indicates which node or nodes were involved. In the example, <code>nodemask = 0x4000</code> identifies node 14 on module 2 (<code>m2n14</code>).</p>
Workaround	<p>Try these tactics to boot the System:</p> <ul style="list-style-type: none"> <li>• Reboot the MSP, using <code>--start_msp=force</code>.</li> <li>• Reseat the processor module in its slot, then reboot the MSP.</li> <li>• Mark the node (in this example (<code>m2n14</code>) as unusable in the <code>etc/sicortex.conf</code> file, then reboot the MSP.</li> </ul> <p>See the <i>SiCortex<sup>®</sup> System Administration Guide</i> for details on how to reboot an MSP and on how to create and edit the <code>sicortex.conf</code> file.</p> <p>Email <a href="mailto:support@sicortex.com">support@sicortex.com</a> if nodes continuously fail to boot.</p>
<b>Issue 5721</b>	<b>During a warm boot, scboot may exit before the node is booted</b>
Description	<p>When rebooting a single node using</p> <pre>scboot -p &lt;partition&gt; --nodes m#n#</pre> <p>the <code>scboot</code> command exits before the node has finished booting.</p>
Workaround	<p>Monitor the state of the node, and wait for it to return to the <code>Idle</code> state before submitting jobs to it.</p> <p>You can use <code>sinfo</code> (<code>watch sinfo -p &lt;partition&gt;</code>) to see if the node is in the <code>Idle</code> state, or look at the node's console log on the SSP to see if <code>scboot</code> has finished booting the node.</p>

<b>Issue 8330</b>	<b>scboot's --show_settings option can stop a running system</b>
Description	The --show_settings option does not work correctly, overriding existing configuration settings.
Workaround	Do not use the --show_settings option.

<b>Issue 9358</b>	<b>scboot's --start_msp=skip option doesn't work properly</b>
Description	The --start_msp=skip option fails, generating an error message like this:  <pre>Checking Module Service Processors unrecognized num Diagcomm failure: ['MSP', 0, -1, 'diagcomm_connect() poll failed (msp0 (sc1-msp0:1235))'] Diagcomm failure: ['DiagComm', -1, -1, 'Error: MSP0 selected but not connected: nodeMask = 1'], rev 08</pre>
Workaround	Do not use the --start_msp=skip option.

<b>Issue 9382</b>	<b>scboot outputs network configuration information</b>
Description	While running, scboot prints some network configuration information, for example:  <pre>... Starting syslog-ng: [ OK ] {'external-network': 'default', 'internal-domain': 'scsystem', 'dns-search-list': 'sicortex.com', 'default-router': 'gw.sicortex.com', 'netblock': '172.31', 'io-nodes': 'sci-m0n6', 'dns-servers': '10.0.0.23 10.0.0.11 10.0.0.36', 'head-node': 'sci-m0n6'} ...</pre>
Workaround	You can ignore these messages.

## 4.5 COMPILING/LINKING Issues

<b>Issue 5817</b>	<b>gcc sometimes cannot link objects compiled by pathcc/pathCC</b>
Description	When linking objects built with the PathScale compiler, gcc may complain of missing references.
Workaround	When this happens, add <code>-lscstr</code> to the end of the gcc link line. For details, see the <i>SiCortex<sup>®</sup> System Programming Guide</i> .
<b>Issue 5988</b>	<b>OpenMP executables produced with PathScale compilers fail the DUMA memory checker</b>
Description	Currently, DUMA does not work with OpenMP executables produced using the PathScale compilers
Workaround	Don't use DUMA on OpenMP applications.
<b>Issue 9410</b>	<b>n32 on the nodes is broken</b>
Description	After running <code>mount-n32</code> on the nodes, you can compile neither n32 nor n64 applications.
Workaround	Don't run <code>mount-n32</code> on the nodes.

## 4.6 LUSTRE™ Issues

Issue 4597	Lustre mount fails when too many connections go through an I/O node
Description	<p>Under some circumstances, an I/O node that is configured for NAT routing can become overloaded trying to handle too many sockets. Because of the large number of sockets opened by a Lustre client, mounting an external Lustre file system can trigger socket overload.</p> <p>When socket overload occurs, Lustre on some client nodes will complain of communication errors while trying to talk to the server. Approximately 1600 sockets per I/O node triggers communication errors.</p>
Workaround	<p>If you encounter socket overload, increase the number of I/O nodes.</p> <p>☼ On an SC5832, mounting a 4-server Lustre file system from every client node requires eight I/O nodes.</p>

## 4.7 FabriCache Issues

<b>Issue 7053</b>	<b>The SLURM partition created by <code>fc_create</code> must be manually deleted after a reboot</b>
Description	<p>Because the FabriCache file system does not survive a reboot, you must recreate it by running, as root, the <code>fc_create</code> script on the SSP after each <code>scboot</code>.</p> <p>The <code>fc_create</code> script also creates a SLURM partition named <code>&lt;fcname&gt;_clients</code>, which contains those nodes that have access to the created FabriCache file system. This partition enables users to launch jobs on the pool of FabriCached nodes.</p>
Workaround	<p>Ensure that the <code>&lt;fcname&gt;_clients</code> partition has been cleaned up. Before you reboot, run on the SSP as root:</p> <pre>fc_destroy &lt;slurm_partition&gt; &lt;fcname&gt;</pre> <p>(where both arguments match those used with <code>fc_create</code> to create the partition).</p> <p>If this command fails, delete the partition manually after rebooting, but before running <code>fc_create</code> again:</p> <pre>scontrol delete PartitionName=&lt;fcname&gt;_clients</pre> <p>Then run <code>fc_create</code> on the SSP.</p>
<b>Issue 7278</b>	<b><code>fc_create</code> fails to run on a Default SLURM partition</b>
Description	<p>With <code>Default=YES</code> set for the SLURM partition on which you want to create a FabriCache file system, the <code>fc_create</code> command fails with an error.</p>
Workaround	<p>Set <code>Default=NO</code> for the partition, then run <code>fc_create</code>.</p>

## 5.0 Fixed Issues

<b>Issue 2750</b>	<b>Processes fail to exit under MPI_Abort</b>
Description	<p>MPI_Abort is executed when a process encounters an error condition. A call to MPI_Abort by any process in an MPI application should guarantee quick termination of all processes initiated by the command that started the job.</p> <p>By default, SLURM didn't terminate all processes when a single process aborted. This behavior was designed to accommodate loosely-coupled applications, in which the failure of a single process need not cause the entire job to fail. For an MPI job, however, this behavior often caused the job to hang and create zombie MPI processes. The SLURM upgrade fixed this problem.</p>
<b>Issue 3474</b>	<b>Name Server lookups made from the nodes may not find site-wide systems</b>
Description	<p>For example, <code>\$ nslookup ws121</code> didn't find the host (<code>sicortex.com</code>), but <code>\$ nslookup ws121.sicortex.com</code> did.</p> <p>The fix requires the System Administrator to put appropriate entries for <code>cluster.dns-servers</code> and <code>cluster.dns-search-list</code> in the <code>sicortex.conf</code> file. For details, see the <i>SiCortex<sup>®</sup> System Administration Guide</i>.</p>
<b>Issue 4631</b>	<b>PathScale compiler provides no diagnostics for identifying mismatched object files in an archive</b>
Description	<p>You can't mix regular objects with IPA objects in the same archive file. The PathScale linker complained</p> <pre>cannot mix regular and ipa objects in same archive</pre> <p>but did not identify which objects were compiled incorrectly.</p>

---

**Fixed Issues -**

---

<b>Issue 4667</b>	<b>During reboot, scbootmon sometimes prematurely signals completion</b>
Description	<p>During reboot, SLURM didn't always receive the message that nodes were down in time to update <code>sinfo</code> with current information. When this happened, based on outdated <code>sinfo</code> information, <code>scbootmon</code> incorrectly signaled the system was booted and ready when it actually wasn't.</p> <p>When this happened, you couldn't run anything on the System because <code>scboot</code> was still booting it.</p>
<b>Issue 4791</b>	<b>Boot up hangs on SC5832 if ev1d runs out of file descriptors</b>
Description	<p>When you booted the SC5832 too many times in quick succession, <code>ev1d</code> sometimes ran out of file descriptors, causing the boot procedure to hang.</p>
<b>Issue 5089</b>	<b>uri server configuration not supported in /etc/ldap.conf on the SSP</b>
Description	<p>Specifying the site LDAP server in <code>/etc/ldap.conf</code> using the <code>uri</code> directive, <code>uri ldap://&lt;SERVER&gt;/</code>, didn't work.</p>
<b>Issue 5970</b>	<b>PathScale cross-compilers can't find gcc</b>
Description	<p>The PathScale C/C++ cross-compilers didn't have the correct pathname for <code>gcc</code>.</p>
<b>Issue 5990</b>	<b>Running executables linked with the GotoBLAS library fails bigphysarea allocation</b>
Description	<p>Running executables linked with <code>-lgotoblas</code> generated this warning:</p> <p style="padding-left: 40px;">GotoBLAS Warning...Bigphysarea alloation was failed.</p>

---

**Fixed Issues -**

---

<b>Issue 6163</b>	<b>scslocate doesn't work on the nodes</b>
Description	<p>The <code>scslocate</code> command didn't work on the nodes because the <code>locate(1)</code> database was not built on the nodes.</p> <p>Red Hat provides <code>/etc/cron.daily/mlocate.cron</code>, which serves the same purpose as <code>scslocate</code>.</p>
<b>Issue 6326</b>	<b>User accounts using LDAP unable to launch SLURM jobs</b>
Description	<p>SLURM jobs launched by <code>root</code> worked, but SLURM jobs launched by regular users failed like this:</p> <pre>\$ srun -p sca -N 1 hostname srun: error: Task launch failed on node sca-m0n0: Unspecified error srun: error: Application launch failed: Unspecified error</pre> <p>This error occurred when certain parameters in the <code>/etc/nsswitch.conf</code> file on the nodes were set incorrectly. To fix, after you configured the workstation to use LDAP for user authentication, you had to edit the <code>/etc/nsswitch.conf</code> file on the nodes to set these parameters:</p> <pre>passwd: files ldap shadow: files ldap group: files ldap</pre> <p>then reboot the nodes.</p>
<b>Issue 6450</b>	<b>scboot may appear to fail if syslog-ng on the SSP runs out of network connections</b>
Description	<p>Lacking network connections, <code>syslog-ng</code> could not receive boot confirmation from the nodes, which caused <code>scboot-monitor</code> to report failure booting the nodes, even if the nodes booted successfully.</p> <p>A <code>scboot-monitor</code> reporting a <code>slurm</code> count greater than a <code>kernel</code>, <code>fabric</code>, or <code>initfs</code> count indicated the lack of network connections.</p>

---

**Fixed Issues -**

---

<b>Issue 6647</b>	<b>scboot complains "scand unresponsive" for a disabled module</b>
Description	On a system with one or more placeholder cards, even though you properly declared the placeholder cards, scboot displayed error messages when it tried to start scand on the missing processor modules.
<b>Issue 6836</b>	<b>When compiling Fortran programs on the nodes, include -I /usr/include on the command line</b>
Description	When compiling programs like POP, which require the NETCDF module, you needed to use the command line option -I to direct the pathf95 compilers to look in /usr/include for the NETCDF module.
<b>Issue 6993</b>	<b>scld32 incorrectly passes -mabi=n32 to the linker</b>
Description	scld32 incorrectly passed -mabi=n32 to the linker
<b>Issue 7191</b>	<b>scboot ignores the -r flag</b>
Description	If you tried to boot an alternate rootfs using scboot's -r flag, the boot failed without generating error messages. The -r flag is no longer supported. Settings in the system.conf file handle booting from an alternate rootfs. See the <i>SiCortex® System Administration Guide</i> for details.
<b>Issue 7525</b>	<b>scboot needs root's PATH</b>
Description	You needed root access to boot the nodes. In previous releases, the su or sudo command was sufficient to get root, but in V3.1, scboot needed programs in /opt/sicortex/bin.  This directory is on root's path, but wasn't typically on a user's path. If /opt/sicortex/bin wasn't on the path when scboot ran, scboot failed.

<b>Issue 8023</b>	<b>Date/Time drift between workstation and nodes can render srun unable to launch jobs</b>
Description	<p>On the SC072 PDS, the date and time on the nodes became out-of-sync with the date and time on the workstation over time. When this occurred, srun could not launch jobs, for example:</p> <pre>\$ srun -p sca -N 12 date srun: error: Task launch failed on node sca-m0n0: Job credential expired</pre>

## 6.0 Documentation Addenda

<b>Issue 3522</b>	<b>Linking MPI codes statically fails with link errors</b>
Description	<p>SiCortex does not support statically-linked MPI programs. Link problems occur in <code>glibc</code> and in the SLURM libraries.</p> <p>However, you can link with the static version of the MPI library and link with the dynamic version of other libraries.</p>

<b>Issue 5997</b>	<b>GPTL does not support applications that call fork or exec</b>
Description	GPTL does not support applications that call the <code>fork</code> or <code>exec</code> functions.

<b>Issue 6089</b>	<b>Red Hat® System Logs tool cannot open all SiCortex log files</b>
Description	<p>The <i>System Logs</i> tool fails to recognize <code>mfd.log</code>, <code>slurm-comp.log</code>, <code>policyd.log</code>, <code>ev1d.log</code>, <code>kernmond.log</code>, <code>envmond.log</code>, <code>mshenv.log</code>, and <code>watchdog.log</code> as log files.</p> <p>To read these log files, open them from the command line.</p>

<b>Issue 6557</b>	<b>Installation of some cross-development performance tools incomplete</b>
Description	<p>The hpcviewer in the HPCToolkit requires a JRE to run. However, the installation software does not provide a JRE.</p> <p>To fix, download and install a JRE before you try to run hpcviewer. For details, go to:</p> <p><a href="http://www.sicortex.com/support/knowledge_base/v3_1_release_note_updates/sc072_pds_installing_a_java_runtime_environment">http://www.sicortex.com/support/knowledge_base/v3_1_release_note_updates/sc072_pds_installing_a_java_runtime_environment</a>.</p>

<b>Issue 6772</b>	<b>Installing 32-bit Application Support Requires Two steps</b>
Description	<p>To support building n32 applications, you had to install the n32 buildroot on the SSP and mount it on the nodes.</p> <p>Now the installation software installs the n32 buildroot, but you still have to mount it manually on the nodes each time you reboot. To do so, create the following script in the /opt/sicortex/config/local.d directory:</p> <pre>#!/bin/bash # Standard boilerplate: grab boot args, preclude running on SSP if [ ! -L /var/state/boot_args ]; then     # Don't run this on the SSP.     exit 0 fi # # Wait our-module-number seconds on each node, # to avoid all nodes trying to mount n32 at once, # which would overload the SSP. # source /var/state/boot_args sleep \${SCv_module_id} /usr/sbin/mount-n32</pre> <p>Then make this script executable with <code>chmod +x</code>, so it will run as part of the boot process.</p>

<b>Issue 6894</b>	<b>Video driver sometimes blanks out the Login screen on bootup</b>
Description	<p>On bootup, the video driver sometimes blanks out the monitor when displaying the Login screen.</p> <p>When this happens, press CTRL ALT Backspace to restart the X11 server, the graphical interface to the PDS monitor.</p> <p>To permanently fix this problem, download and install the ATI Catalyst 8.10 Proprietary Linux x86 Display Driver from <a href="http://support.amd.com/us/gpudownload/linux/8-10/Pages/radeon_linux.aspx">http://support.amd.com/us/gpudownload/linux/8-10/Pages/radeon_linux.aspx</a>.</p> <p>Follow these steps to install and configure the driver on the SSP:</p> <ol style="list-style-type: none"><li>1. Run the ATI driver installer.</li><li>2. Run the aticonfig script: <code>aticonfig --initial -f</code></li><li>3. Log off then log back on to restart the X server. The screen comes up in low resolution mode.</li><li>4. Set the display resolution to whatever you require.</li><li>5. Reboot the system to ensure a clean configuration.</li></ol>
<b>Issue 8930</b>	<b>V4.0 libscmpi binary incompatible with previous versions</b>
Description	<p>The MPI header files in V4.0 have changed enough that binary compatibility with previous versions is not guaranteed.</p> <p>Recompile and relink any pre v4.0 MPI programs.</p>