Cray XT™ Series 2.0 Software
Release Overview
S–2425–20
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSD V3 Accounting</td>
<td>19</td>
</tr>
<tr>
<td>VisIt Application for Catamount/CVN Nodes</td>
<td>19</td>
</tr>
<tr>
<td>Optional PBS Pro 8.1 Product</td>
<td>20</td>
</tr>
<tr>
<td>Programming Environment Products</td>
<td>20</td>
</tr>
<tr>
<td>Additional Enhancements to System Administration and System Maintenance</td>
<td>22</td>
</tr>
<tr>
<td>General System-Administration Enhancements</td>
<td>22</td>
</tr>
<tr>
<td>Release Upgrade, Configuration, and Installation Enhancements</td>
<td>22</td>
</tr>
<tr>
<td>Release Switching</td>
<td>24</td>
</tr>
<tr>
<td>Linux 2.6.16 Kernel on Service Nodes</td>
<td>24</td>
</tr>
<tr>
<td>Enhanced Boot Procedure</td>
<td>24</td>
</tr>
<tr>
<td>Native IP Default</td>
<td>25</td>
</tr>
<tr>
<td>ARP_OVERHEAD Parameter Added in xt.conf File</td>
<td>25</td>
</tr>
<tr>
<td>Realm-Specific IP Addressing (RSIP) for Cray XT CNL Compute Nodes</td>
<td>25</td>
</tr>
<tr>
<td>Enhancements to the System-Management Graphical User Interface</td>
<td>26</td>
</tr>
<tr>
<td>xtdumpsys Data Collection May Be Customized</td>
<td>27</td>
</tr>
<tr>
<td>Additional New and Enhanced Commands</td>
<td>27</td>
</tr>
<tr>
<td>Improvements to System Resiliency</td>
<td>29</td>
</tr>
<tr>
<td>Login or Network Node Warm Reboot</td>
<td>29</td>
</tr>
<tr>
<td>Enhancements to System Monitoring</td>
<td>30</td>
</tr>
<tr>
<td>Enhancements to Input/Output and Devices</td>
<td>31</td>
</tr>
<tr>
<td>Lustre File System 1.4.9</td>
<td>32</td>
</tr>
<tr>
<td>(Deferred implementation) Lustre File System Automatic Failover</td>
<td>32</td>
</tr>
<tr>
<td>Logical Volume Manager (LVM) Module Provided</td>
<td>34</td>
</tr>
<tr>
<td>Support for Engenio Disk Devices</td>
<td>35</td>
</tr>
<tr>
<td>Resolved Critical and Urgent Software Problem Reports (SPRs)</td>
<td>35</td>
</tr>
</tbody>
</table>

**Compatibilities and Differences [3]**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users Must Recompile Applications</td>
<td>37</td>
</tr>
<tr>
<td>The -lsma64 Compiler Option No Longer Required</td>
<td>37</td>
</tr>
<tr>
<td>MPICH2 Edge Color All-to-all Algorithm for Non-power-of-two Ranks</td>
<td>37</td>
</tr>
<tr>
<td>Change to Upper Limit for Quantum for Catamount/CVN Nodes</td>
<td>38</td>
</tr>
</tbody>
</table>

S–2425–20
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man Pages in <code>roff</code> Format</td>
<td>38</td>
</tr>
<tr>
<td>Additional System Management Compatibility Issues and Differences</td>
<td>38</td>
</tr>
<tr>
<td>Release 2.0 Upgrade-related and Configuration-related Changes</td>
<td>38</td>
</tr>
<tr>
<td>Supported Upgrade Path</td>
<td>38</td>
</tr>
<tr>
<td>Upgrade Packages Provided as Downloadable ISO Images</td>
<td>39</td>
</tr>
<tr>
<td>New LUN Size Upper Limit</td>
<td>39</td>
</tr>
<tr>
<td>The <code>/etc/lustre/llrd.conf</code> File Requires Listing All Lustre MDS and OSS Nodes</td>
<td>39</td>
</tr>
<tr>
<td>Upgrading System Software Requires SDB Update</td>
<td>39</td>
</tr>
<tr>
<td>New Boot Images Directory for CNL Images</td>
<td>39</td>
</tr>
<tr>
<td>Native IP Is Default as of Cray XT Series 2.0 Release</td>
<td>39</td>
</tr>
<tr>
<td>Last Release Supporting IPPO</td>
<td>39</td>
</tr>
<tr>
<td>Operational-related Changes</td>
<td>40</td>
</tr>
<tr>
<td>Location of <code>xtbootimg</code>, <code>xtptltrace</code>, <code>smwhotbackup</code>, and <code>smwsnapshot</code> Commands Changed</td>
<td>40</td>
</tr>
<tr>
<td><code>xtnidname</code> Command Obsolete</td>
<td>40</td>
</tr>
<tr>
<td>Boot Manager Parameter String Maximum Increased</td>
<td>40</td>
</tr>
<tr>
<td>CNL and Catamount/CVN Operational Differences</td>
<td>40</td>
</tr>
<tr>
<td>Cray RAS and Management System (CRMS) Evolving to Hardware Supervisory System (HSS)</td>
<td>40</td>
</tr>
</tbody>
</table>

## Documentation [4]

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrayDoc Documentation Delivery System</td>
<td>41</td>
</tr>
<tr>
<td>Accessing Product Documentation</td>
<td>41</td>
</tr>
<tr>
<td>Books Provided with This Release</td>
<td>42</td>
</tr>
<tr>
<td>Man Pages Provided with This Release</td>
<td>44</td>
</tr>
<tr>
<td>Additional Documentation Resources</td>
<td>46</td>
</tr>
<tr>
<td>TotalView Documentation From TotalView Technologies, LLC</td>
<td>47</td>
</tr>
<tr>
<td>PathScale Compilers Documentation From PathScale LLC</td>
<td>47</td>
</tr>
<tr>
<td>Cray Glossary</td>
<td>47</td>
</tr>
</tbody>
</table>

## Release Package [5]

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supported Upgrade Path</td>
<td>49</td>
</tr>
</tbody>
</table>

*S–2425–20*
This document gives customers a overview of the Cray XT series 2.0 general availability (GA) software release.

1.1 Emphasis for the Cray XT Series 2.0 Release

The Cray XT series 2.0 release provides the following key enhancements.

- **CNL compute-node kernel.** Customers may now choose to run a new kernel on Cray XT series compute nodes. This new kernel is called CNL. CNL provides support for application execution without the overhead of a full operating system image. The initial release of CNL delivers the same functionality as Catamount plus Linux symmetric multiprocessing (SMP), a larger set of supported system calls than the Catamount compute-node kernel, and standard networking. The Cray XT series 2.0 GA release supports 124 cabinets of Cray XT dual-core compute nodes running CNL.

- **ALPS.** The Application Level Placement Scheduler (ALPS) is provided for systems running CNL. ALPS provides application placement, launch, and management functionality and cooperates with third-party batch systems for application scheduling.

- **Compute-node selection.** Users now have more control over the selection of the compute nodes on which to run their applications. Users select nodes on the basis of desired characteristics. These characteristics are called node attributes, which enable a placement scheduler to schedule jobs based on the node attributes.

- **PBS Pro 8.1 release.** The optional batch system product PBS Pro 8.1 release is supported. The PBS Pro product is available directly from Cray Inc. and is based on the Altair PBS Professional 9.0 release.

- **Lustre file system upgrade.** The Lustre file system is upgraded to the 1.4.9 version from Cluster File Systems.

- **(Deferred implementation) Lustre file system automatic failover.** The Lustre file system automatic failover capability is supported for compute nodes running CNL.

- The Cray XT programming environment is now released independently; a new installation guide, *Cray XT Series Programming Environments Installation Guide*, is provided.
- **Separate Cray System Management Workstation (SMW) release media.** The Cray System Management Workstation (SMW) software is provided on separate media. A new CRAYSMWinstall.sh script, a new SMWinstall installation program, and a new installation guide, *Cray System Management Workstation (SMW) Software Installation Guide*, are provided.

- **Alternative boot root.** After the Cray System Management Workstation (SMW) release software is installed, customers may now install the Cray XT series 2.0 release software to an alternative boot root and shared root while the Cray XT system remains operational. Because the configuration steps may be performed on the alternative location, the required amount of dedicated system time is reduced.

- **System sets.** System administrators can create system sets. A system set is a complete set of all file systems and boot images, making it a complete, bootable system. A system set provides a way to group several disk partitions on the boot RAID into a set, which is given a label. By configuring system sets, a system administrator can easily switch between different Cray XT software releases.

- **Cray X2 compute nodes.** An update of this release will support Cray X2 compute nodes. Cray X2 compute nodes are Cray proprietary vector processors delivered as 4-way SMP compute nodes. For more information, see the *Cray Programming Environment Releases Overview and Installation Guide*; also, the *Cray X2 Compute Node Release Overview* will be provided with the update for Cray X2 compute nodes.

- **Selectable compute node OS components of Cray XT release package to be available.** Based on their software licensing agreement with Cray, customers will soon be able to specify the compute node OS component (Cray XT CNL software, Cray XT Catamount/ CVN software, or Cray X2 software) of their Cray XT release package. When this offering becomes available, it will be announced in the associated Cray XT 2.0.xx update’s README document.

For additional information about all software enhancements for this release, see Chapter 2, page 5.
1.2 Reader Comments

Contact us with any comments that will help us to improve the accuracy and usability of this document. Be sure to include the title and number of the document with your comments. We value your comments and respond to them promptly. Contact us in any of the following ways:

**E-mail:**
docs@cray.com

**Telephone (inside U.S., Canada):**
1–800–950–2729 (Cray Customer Support Center)

**Telephone (outside U.S., Canada):**
+1–715–726–4993 (Cray Customer Support Center)

**Mail:**
Customer Documentation
Cray Inc.
1340 Mendota Heights Road
Mendota Heights, MN 55120–1128
USA
This chapter describes the enhancements provided with this release. It assumes the reader currently uses or manages a Cray XT series system.

For compatibility issues and differences that you should be aware of when installing or using this release, see Chapter 3, page 37.

**Note:** The *Limitations for Cray XT Series 2.0 GA* includes a description of temporary limitations of this release. The *Cray XT Series 2.0 Release Errata* includes installation and configuration changes identified after the installation documentation for this release was packaged and lists customer-filed critical and urgent SPRs closed with this release. A printed copy of these documents is included with the release package; they are also available from your Cray representative.

In addition to the documentation noted in each feature description, see Section 4.3, page 42 for a list of revised manuals provided with this release.

**Note:** Throughout this document, the use of the term *Catamount* refers to either the Catamount kernel running on compute nodes or the Catamount Virtual Node (CVN) capability that supports dual-core processing on compute nodes running the Catamount kernel.

### 2.1 New Compute Node Kernel, CNL

This release adds a new compute node kernel, which is called CNL. CNL is based on the SUSE LINUX Enterprise Server 9 release software and the Linux 2.6.16 kernel. The initial release of CNL delivers the same functionality as Catamount plus Linux symmetric multiprocessing (SMP), a larger set of supported system calls than the Catamount compute node kernel, and standard networking.

The Cray XT series 2.0 general availability (GA) release supports 124 cabinets of dual-core compute nodes running CNL.
2.1.1 User-visible Changed Functionality

Users will notice the following differences between using Catamount/CVN and CNL on compute nodes:

- New ALPS commands are provided to launch and manage CNL applications. The aprun command launches CNL applications. The apstat command provides status information about CNL applications. The apkill command sends a signal to a specified CNL application. For additional information about using ALPS commands, see Section 2.2, page 12.

- Customers running Catamount/CVN may use the xtgenacct utility to produce accounting reports. Equivalent accounting capability is deferred for customers using CNL on their compute nodes.

- Before beginning to compile programs, the user needs to verify the target architecture is set correctly. The target architecture is used by the compilers and linker in creating executables to run on either Catamount or CNL compute nodes. (The target architecture is not necessarily the operating system currently running on compute nodes; that is determined by the system administrator at boot time.) A CNL application cannot run on compute nodes running Catamount or CVN, nor can a Catamount/CVN application run on compute nodes running CNL.

The CNL target is defined automatically at login. If any compute node is running CNL, then the target module xtpe-target-cnl is loaded, and the XTPE_COMPILE_TARGET=linux compile target environment is set. If no compute nodes are running CNL, then the target module xtpe-target-catamount is loaded, and the XTPE_COMPILE_TARGET=catamount compile target environment is set. To override a setting, use the -target= command-line option.

After you have logged in, you can change your own default compile target by using the module swap command. For example, if you have logged on to a system running CNL on the compute nodes, but you want to set your compile target for Catamount/CVN, use the following command:

```
module swap xtpe-target-cnl xtpe-target-catamount
```

- Users need to keep their Catamount/CVN executables separate from their CNL executables. CNL executables can be executed only when the compute nodes are running CNL, and Catamount/CVN executables can be executed only when the compute nodes are running Catamount or CVN.
To determine which kernel is running on the compute nodes, invoke the `xtshowcabs` command. The `xtshowcabs` and `xtshowmesh` displays now represent idle compute nodes, as described in the output display legend.

New ALPS `aprun(1)`, `apstat(1)`, and `apkll(1)` man pages and revised `yod(1)`, `xtshowcabs(1)`, and `xtshowmesh(1)` man pages are provided.

Programming considerations and CNL and Catamount/CVN differences are highlighted in the following sections of this overview. However, the *Cray XT Series Programming Environment User’s Guide* includes the complete list of programming considerations and CNL and Catamount/CVN differences.

### 2.1.2 CNL Programming Considerations

- To launch applications on CNL nodes, use the `aprun` command instead of the `yod` command. (The ALPS feature must be running to use the `aprun` command.)
- The `yod` and `aprun` utilities differ in the run-time services they provide. Applications running under Catamount or CVN rely on `yod` to provide stdin, stdout, and stderr processing; signal management; and system call-forwarding. Applications running under CNL rely on `aprun` to manage the stdin, stdout, and stderr streams and handle signal management; CNL handles system calls.
- You cannot run more than one application on a dual-core CNL compute node.
- CNL supports a single page size of 4 KB.
- CNL does not support the `dclock()` function.
- When an application fails on CNL, one core file is generated for the first failing process.
- The `aprun` utility catches and forwards the `SIGHUP`, `SIGINT`, `SIGQUIT`, `SIGTERM`, `SIGABRT`, `SIGUSR1`, and `SIGUSR2` signals to an application.
- Memory limits are defined by the node default or the `aprun -m` option. Time limits are inherited from the `aprun` process limits or specified with the `aprun -t` option. Other limits are inherited from the limits of `aprun`. All limits apply to individual processing elements. There are no aggregate application limits specified with `aprun` options.
- CNL compute nodes do not support `gdb`; see Section 5.3, page 50, for debugger information.
• The underlying support for TCP/IP sockets is available on CNL. However, a user cannot rsh or ssh into any compute nodes; rsh and ssh are not supported on CNL.

• CNL uses approximately 250 MB of memory per node. If you are running in single-core mode on a dual-core system, the remaining memory is available for the user program executable, user data arrays, the stack, libraries and buffers, and SHMEM symmetric stack heap.

  If you are running in dual-core mode, the two cores together can use the remaining memory; note that each core can use differing amounts of memory. The memory allocated to each core is available for the user executable, user data arrays, stack, libraries and buffers, and SHMEM symmetric stack heap.

2.1.3 glibc Functions CNL and Catamount/CVN Differences

Unlike Catamount and CVN, CNL provides limited support of the process control functions such as popen(), fork(), exec(), system(), and ttyname(). The exec() function can execute the scp and ksh commands and the supported BusyBox commands (see Section 2.1.4).

2.1.4 BusyBox Commands Supported on Cray XT Compute Nodes Running CNL

The following BusyBox user-level commands are supported on Cray XT compute nodes running CNL: ash, BusyBox, cat, chmod, chown, cp, cpio, free, grep, gunzip, kill, killall, ln, ls, mkdir, mktemp, more, ps, rm, sh, tail, test, vi, and zcat.

For information about supported options, see the busybox(1) man page.

For BusyBox administrator commands supported, see Cray XT Series System Management.

2.1.5 MPI CNL and Catamount/CVN Differences

The same versions of MPI run on Catamount, CVN, and CNL nodes; there are no CNL-specific changes to the MPI interface.

Note: Hybrid applications that use MPI should make MPI calls from master or sequential regions, and not from parallel regions.
2.1.6 SHMEM CNL and Catamount/CVN Differences

When running on CNL, the environment variables XT_LINUX_SHMEM_STACK_SIZE and XT_LINUX_SHMEM_HEAP_SIZE can be used to control the size (in bytes) of the stack or private heap, respectively. A process cannot request memory for the stack or private heap in excess of the total memory on a node divided by the number of CPUs on that node.

The environment variable XT_SYMMETRIC_HEAP_SIZE can be used on both Catamount/CVN and CNL nodes to control the size, in bytes, of the symmetric heap.

The revised intro_shmem(1) man page provides information about the similarities and differences between Cray SHMEM on Catamount/CVN nodes and on CNL nodes, including use of the XT_LINUX_SHMEM_STACK_SIZE, XT_LINUX_SHMEM_HEAP_SIZE, and XT_SYMMETRIC_HEAP_SIZE environment variables.

2.1.7 PAPI CNL and Catamount/CVN Differences

To run applications that use PAPI functions on compute nodes running CNL, the papi-cnl module must be loaded:

% module load papi-cnl

To run applications that use PAPI functions on compute nodes running Catamount or CVN, the papi module must be loaded:

% module load papi

2.1.8 I/O Buffering CNL and Catamount/CVN Differences

The IOBUF library is provided for compute nodes running Catamount or CVN. Compute nodes running CNL have their file I/O buffered automatically by the Linux page cache.

2.1.9 OpenMP CNL Considerations

Cray XT series systems support OpenMP shared memory model library routines, Fortran OpenMP directives, and C and C++ OpenMP pragmas for applications running under CNL.

OpenMP is not supported for applications running under Catamount or CVN.
Note: PGI compiler users and PathScale compiler users must specify the -mp option on the cc, CC, ftn, or f77 compiler driver command when compiling programs that include OpenMP directives or pragmas.

Note: Hybrid MPI/OpenMP applications that use MPI should make MPI calls from master or sequential regions, and not from parallel regions.

2.1.10 Cray XT CNL Administration Considerations

System administrators should note the following:

- To run CNL on compute nodes, the following image management functionality has been added:
  - For systems running CNL on compute nodes, a new /opt/xt-images/ directory is provided. The directory contains a master/default/ directory and a templates/default directory.
  - The new xtclone and xtpackage utilities run on the SMW. They are used to set up CNL compute node images.

The xtclone utility creates an image directory within the image repository. By default, the xtclone utility copies the contents of the master image (located in the master/default directory of the image repository) to the image directory. You can then create a template and customize it to your site's configuration before creating a boot image. The xtclone utility creates a manifest file in the root of the image directory named image.manifest that contains the absolute paths for the master source, template source, and destination directories at the time the xtclone utility was invoked. This file is a good reference for reviewing contents of an image.

The xtpackage utility facilitates the process of creating boot images from image directories. The xtpackage utility creates the directory $IMAGE_REPOSITORY/$SRC/$LOADNAME/ and creates intermediate files in it. It also creates the $(IMAGE_REPOSITORY)/$SRC/$LOADNAME.load file, which is a description of the files in the $LOADNAME directory. The xtbootimg utility uses the .load file to collect the necessary files to add this image to a unified boot image. You can use the xtpackage -o or -O option to create a "compute node only" boot image.

The xtclone(8) and xtpackage(8) man pages are included in the release package.
For compute nodes running CNL, use the `xtbootimg -L` option to specify the CNL load file path.

For compute nodes running CNL, the `/etc/hosts` file on the boot root is generated at boot time to include CNL compute nodes. Also, the installation and upgrade process modifies the `/etc/hosts` file on the boot root to include CNL compute nodes if they are not included.

- You can boot either CNL or Catamount/CVN on compute nodes. If you divide your system into more than one partition (logical machine), you can boot CNL in one partition and Catamount/CVN in another.

- Compute-node failure affects only the job running on that node. The rest of the system continues running.

- To determine which kernel is running on the compute nodes, use the `xtshowcabs` command. The `xtshowcabs` and `xtshowmesh` commands have enhanced displays that designate the kernel with which the node attributes were configured. For more information about nodes attributes, see Section 2.3, page 15.

- For systems running CNL on compute nodes, the `xtshutdown` command does not shut down compute nodes. Use the `xtgenevent` command on the SMW to shut down compute nodes, before running the `xtshutdown` command to shut down service nodes.

  **Note:** CNL compute nodes should be shut down before shutting down Lustre on the service nodes.

- To make CNL the default `osclass` node attribute, in the `/etc/attr.defaults` file on the boot node root (not the shared root), uncomment the 'DEFAULT: osclass=2' line (for CNL). This setting does not impact the booting process. The compute node operating system to boot must still be specified at boot time. For more information about the `/etc/attr.defaults` file and using the `osclass` node attribute, see Section 2.3.3.1, page 16.

- Although administrators need make no changes for the Lustre file system when running compute nodes with the CNL kernel, the following Lustre file system related differences should be noted:
  - For boot automation scripts, when running CNL on compute nodes, Lustre should start up before the compute nodes.
  - The Lustre lock recovery daemon, `llrd`, applies only when running Catamount/CVN on compute nodes.
- The Lustre table in the SDB is not used when running CNL on compute nodes.

- `libsysio` functions only when running Catamount/CVN on compute nodes. Applications running on CNL compute nodes do not use the `/etc/sysio_init` configuration file or the `sysio` library. Compute nodes running CNL mount the Lustre file system on boot and remain mounted between applications.

- For systems running CNL, when an administrator changes the mode of a compute node from batch to interactive, and vice versa, with the `xtprocadmin` or `mysql` commands, the ALPS `apbridge` daemon must be restarted on the boot node.

- The `/proc` file system does not exist on compute nodes running Catamount/CVN.

- CPA and the `yod` command function only with Catamount/CVN running on compute nodes.

- ALPS, the `aprun` command, and other ALPS related commands function only with CNL running on compute nodes.

The following administrator commands and their associated man pages were modified to support the CNL kernel: `xtbootsys(8)`, `xtbootimg(8)`, `xtcli_boot(8)`, and `xtprocadmin(8)` (extended to display node attributes).

### 2.2 Application Level Placement Scheduler (ALPS) for CNL Applications

**Note:** ALPS is not supported for Cray XT compute nodes running the Catamount kernel or the CVN capability that supports dual-core processing on compute nodes running the Catamount kernel.

The new Application Level Placement Scheduler (ALPS) is provided for compute nodes running CNL. ALPS provides application placement, launch, and management functionality and cooperates with third-party batch systems for application scheduling.

The third-party batch system makes the policy and scheduling decisions, and ALPS provides a mechanism to place and launch the applications contained within batch jobs. ALPS also supports placement and launch functionality for interactive applications.
Note: The ALPS application placement and launch functionality is provided for applications executing on compute nodes only; ALPS does not provide placement and launch functionality for service nodes.

A markup language (XML) interface is provided by ALPS for communication with third-party batch systems. This interface is available through use of the apbasil client. ALPS uses application resource reservations to guarantee resource availability to batch system schedulers.

2.2.1 Using ALPS

The user may use commands for actual job submission, such as qsub for PBS Pro, and these commands interact with ALPS for the purposes of placing that job across the required compute nodes.

Note: Only one application can be placed per node; you cannot run two different executables on the same node at the same time.

The following commands are provided to use ALPS:

aprun Initiates the main interface by which interactive users can submit applications
apstat Provides status information about reservations, compute resources, and pending and placed applications
apkill Sends a signal to a specified application to terminate a job

See the following man pages for detailed information: intro_alps(1) (an introduction to ALPS), aprun(1), apstat(1), and apkill(1).

The xtshowcabs and xtshowmesh commands were modified to show node usage and job information reported by ALPS.

2.2.2 ALPS Administration

The apmgr command initiates an interface for ALPS by which system administrators can cancel pending reservations. For additional information, see the apmgr(8) man page.
A series of daemons make up part of ALPS. An apsys daemon is run on each
login service node. These daemons act as intermediaries for the many aprun
command invocations on a service node to talk with the placement scheduler,
apsched.

**Note:** Only one instance of the placement scheduler may run across the entire
system at a time. The ALPS daemons ensure this through the use of the shared
file system.

The apbridge and apwatch daemons provide information to apsched about
the nodes available for placement. The apinit daemon launches and manages
new applications. A master apinit daemon resides on every compute node and
initiates all new activity on that node.

The aprun command connects to the apinit daemon on the first node of an
application's allocated node set and sends a launch message containing all of
the information that the compute nodes need to launch and manage the new
application. The apinit daemon forks additional child processes as needed.
This first apinit daemon also contacts the apinit daemon on some number
of other compute nodes in order to fan out the application to other nodes. This
fan-out process propagates in a tree-like fashion until the specified number of
nodes has been contacted.

Some login node application tools, such as debuggers and performance analysis
programs, require a helper agent to run on the same set of compute nodes as an
application. ALPS provides compute-node launch assistance to these application
tools.

See the following man pages for detailed information about the ALPS daemons:
apsys(8), apsched(8), apinit(8), apbridge(8), and apwatch(8).

ALPS is automatically loaded when CNL is booted and running on Cray XT
compute nodes. To configure ALPS, use the new /etc/sysconfig/alps
system information file (in both the boot root and in the shared root) and the new
/etc/alps.conf configuration file (in the shared root).

Administrators set a node as batch or interactive in the alloc_mode column
of the SDB processor table. The apstat command displays each node
as B (for batch) or I (for interactive). When an administrator changes
the batch/interactive mode designation for a compute node (using the
xtprocadmin command or the mysql command), it is necessary to restart the
apbridge daemon in order for the change to take effect. This must be done on
the boot node. Restarting the other ALPS components (for example, on the SDB
node or on the login node if they are separate nodes) is not necessary.
2.3 Node Attributes

Users now have more control over the selection of the compute nodes on which to run their applications. Users select nodes on the basis of desired characteristics (node attributes), allowing a placement scheduler to schedule jobs based on the node attributes.

For an application to be run on CNL compute nodes, the nodes satisfying the requested node attributes are passed by the aprun utility to the ALPS placement scheduler as the set of nodes from which to make an allocation.

For an application to be run on Catamount or CVN compute nodes, the nodes satisfying the requested node attributes are passed by the yod utility to the CPA placement scheduler as the set of nodes from which to make an allocation.

The new aprun(1) and cnselect(1) man pages and the revised yod(1), xtproadmin(8), xthwinv(8), and xtbounce(8) man pages document this feature.

2.3.1 Viewing Node Attributes

The xtproadmin command is used to view node attributes. The new xtproadmin -A option lists all attributes of selected nodes. The new xtproadmin -a attr1,attr2 option lists selected attributes of selected nodes.

2.3.2 Specifying Node Attributes at Job Launch

Cray XT series systems provide two application launch utilities: yod for Catamount and CVN applications and aprun for CNL applications. Both the yod and aprun utilities provide options for automatic and manual application placement. With automatic job placement, yod or aprun distributes the application instances on the number of processors requested, selecting from the available nodes. With manual job placement, yod or aprun distributes the application instances using the user-supplied node candidate list.
To provide the application launcher with a list of nodes that have a particular set of characteristics (attributes), the user invokes the new cnselect command, which generates a candidate list of compute nodes based on user-specified selection criteria. The cnselect script uses these selection criteria to query the table of node attributes in the service database (SDB) and returns a node list to the user based on the results of the query. Available attributes are: clockmhz, pageszl2, availmem, coremask, archtype, osclass, nodeid, label0, label1, label2, and label3. For additional information about these attributes, see Section 2.3.3.1, page 16, and the cnselect(1) man page.

When launching the application, the user includes the node list using the aprun -L node_list option or using yod options (-f list nodelist_file or -list processor-list). The ALPS or CPA placement scheduler allocates nodes based on this list.

Note: The cnselect script may be used as provided, or the system administrator can modify the script to meet site-specific needs.

2.3.3 System Administration of Node Attributes

The system administrator can set the characteristics of individual compute nodes by using the /etc/attr.xthwinv and /etc/attr.defaults files or by using the xtprocadmin command.

2.3.3.1 Setting Node Attributes Using the /etc/attr.xthwinv and /etc/attr.defaults Files

Node attribute information is written to the /etc/attributes data file and loaded into the new attributes table in the SDB when the SDB is booted.

The system administrator enables node attributes by instructing the boot process to create the /etc/attributes data file when the boot node is booted and to load it into the SDB when the SDB node is booted. To instruct the boot process to generate the /etc/attributes file, the system administrator modifies the /etc/xt.conf file to include the line: SDBATTR=/etc/attributes. To instruct the boot process to load the /etc/attributes file into the SDB, the system administrator invokes the xtopview command and adds the following line to the /etc/xt.conf shared root file: SDBATTR=/etc/attributes.

Data for the /etc/attributes file comes from two other files: /etc/attr.xthwinv, which contains information to generate the hardware attributes for each node, and /etc/attr.defaults, which contains default values for additional attributes not generated from the attr.xthwinv file.
The hardware attributes listed in the `attr.xthwinv` file apply to all nodes and include:

- **clockmhz**: The processor clock speed in megahertz.
- **availmem**: The amount of physical memory on the node.
- **coremask**: A bit mask that shows which cores are available on a node. For a single-core processor, the value is 1. For a dual-core processor where both cores are available, the value is 3.

To generate the `/etc/attr.xthwinv` file, the system administrator invokes the `xthwinv` command on the System Management Workstation (SMW), redirecting the output to the `/etc/attr.xthwinv` file on the boot node, for example, `ssh smw 'xthwinv s0' > /etc/attr.xthwinv`, which is run from the boot node.

In addition to hardware characteristics, administrators can specify additional attributes in the `/etc/attr.defaults` file. This file can contain attribute settings for attributes in the following list. The attributes can be applied to all nodes or to a given set of nodes.

- **archtype**: The architecture type: reserved for future architecture=1; Cray XT series=2; default=2.
- **osclass**: The compute node's operating system: Catamount=1; CNL=2; default=1. For a machine that sometimes boots CNL and sometimes boots Catamount, you can modify the automation boot scripts to set `osclass` appropriately.
- **pageszl2**: The log base 2 of the page size. For example, if `pageszl2` is 12, then the page size is 4K ($2^{12} = 4096$, or 4K).

- **label0**, **label1**, **label2**, **label3**: Each label is a string of up to 32 characters; the string cannot contain any spaces or shell-sensitive characters.

To create the `attr.defaults` file, the system administrator copies the example file provided in `/opt/xt-boot/default/etc/attr.defaults.example` and then edits the file to modify the existing attribute settings and to create site-specific attributes as needed.
In addition to the attributes in the `/etc/attr.defaults` file, there are two keywords that allow a system administrator to describe the node or set of nodes to which attributes are assigned. For global default-attribute values that apply to the entire system, the line that specifies an attribute must begin with the DEFAULT: keyword. For example:

```
DEFAULT: osclass=1
```

The `nodeid` keyword assigns attributes to a specific node or set of nodes and overrides a default setting. For values that apply only to certain nodes, the line that specifies the attributes must begin with `nodeid=[RANGE]`, where `RANGE` is a comma-separated list of nodes and ranges that have the form `m-n`. No spaces are allowed between comma-separated list of nodes and ranges. When listing multiple attributes for a set of nodes, separate the attributes by a single space, for example, `nodeid=234,245-248 archtype=2 osclass=1`.

The following example uses labels to assign groups of compute nodes to specific user groups without the need to partition the system:

```
nodeid=101-500 label0=physicsdept
nodeid=501-1000 label1=csdept
nodeid=50-100,1001 label2=biologydept
```

**Note:** System administrators should not set hardware attributes (memory size, clock speed, and cores) in the `attr.defaults` file because the values will be overwritten by those already specified in the `/etc/attr.xthwinv` file.

### 2.3.3.2 Setting Attributes Using the `xtprocadmin` Command

System administrators can also use the `xtprocadmin -a attr=value` command to temporarily set certain site-specific attributes. Using the `xtprocadmin -a attr=value` command to set certain site-specific attributes is not persistent across reboots. Attribute settings that are intended to be persistent across reboots must be specified in the `attr.defaults` file.

The `xtprocadmin` command can be used on Catamount/CVN nodes with no further actions because CPA and yod always check the database before doing a launch. However, for CNL nodes, when an administrator changes the batch/interactive mode designation for a compute node in the `alloc_mode` column of the SDB `processor` table (using the `xtprocadmin` or `mysql` commands), it is necessary to restart the ALPS `apbridge` daemon in order for ALPS to be aware of changes made to the SDB. This must be done on the boot node. Restarting the other ALPS components (for example, on the SDB node or on the login node if they are separate nodes) is not necessary.
2.4 New `yod` Command Option

The new `yod -tracemalloc` option reports activities related to memory allocation and deallocation, allowing users to pinpoint memory leaks.

2.5 BSD V3 Accounting

The GNU process accounting has been upgraded to 6.4. It may be enabled for Cray XT series service nodes. The package name is `acct`, which is activated using the `acct` boot script.

The GNU 6.4 process accounting utilities process V2 and V3 format records seamlessly, even when the data is written to the same file. Output goes to an accounting file, which by default is `/var/account/pacct`. The accounting utilities provided for administration use are: `ac`, `lastcomm`, `accton`, and `sa`. The related man pages are accessible by using the `man` command.

2.6 VisIt Application for Catamount/CVN Nodes

The VisIt application can be run on Cray XT Catamount/CVN nodes. In order to run the VisIt application, there must be a socket off-load (SOLD) facility available to the compute node clients. This socket off-load facility is provided by a separate RPM that is installed by default when installing the Cray XT 2.0 release package. The RPM creates a user library and provides the `sold` executable for user access.

Caution: The socket off-load (SOLD) facility is supported for use ONLY with the VisIt application.

For more information about configuring and using the VisIt application with the SOLD facility, see the VisIt and SOLD on Catamount Cray white paper available through CRInform.
2.7 Optional PBS Pro 8.1 Product

The optional PBS Pro 8.1 batch system product for Cray XT series systems is available directly from Cray; it is based on the Altair PBS Professional 9.0 release. This PBS Pro 8.1 release contains several new features, including:

- Job arrays, which are mechanisms for grouping related work. Job arrays allow a user to submit, query, modify, and display a set of jobs as a single unit. This new feature is particularly useful for users who submit and manage very large numbers of related jobs.

- Enhancements to job resource requests and placement.

- Scheduler enhancements, including backfill.

- Improved stability.

Note: Under the Cray XT series 1.5 release, PBS Pro 5.3.5xt supported the `qsub size` parameter to specify the number of compute nodes (each with one socket) that a user wanted to reserve for a job. Under the Cray XT series 2.0 release, the `size` parameter is no longer valid for PBS Pro 8.1 and later versions. The `size` parameter is replaced with a set of resources a user can reserve. Information about the options to use in place of the `size` parameter is included in the Resource Types section of the `pbs_resources(7B)` man page.

Also, with previous releases, customers running PBS Pro had the `xt-pbs` module automatically loaded at login. Because PBS Pro is an optional product, beginning with the PBS Pro 8.1 release for Cray XT series systems, the module name changed to `pbs`, and system administrators or users must load the `pbs` module:

```
module load pbs
```

For information about loading the `pbs` module, configuring your environment to have it automatically loaded at login, and for detailed information about differences between using the PBS Pro 5.3.5xt and PBS Pro 8.1 releases, see the `PBS Pro Release Overview, Installation Guide, and Administration Addendum`.

2.8 Programming Environment Products

The following Cray XT series Programming Environment products are available. For a list of products included in the Cray XT series Programming Environment release package, see Section 5.5, page 50.
• Products covered by the Cray XT PGI Programming Environment license:
  – PGI Compilers
  – GNU Compiler Collection (GCC)
  – Modules
  – xt gdb module for using the GNU debugger
  – toolsup (dwarf and elf libraries)
  – Scientific libraries (Libsci), which includes the Iterative Refinement Toolkit (IRT)
  – Libscifft library
  – IOBUF library for systems running Catamount or CVN on compute nodes
  – MPT library, which includes MPI and SHMEM
  – AMD Core Math Library (ACML)
  – Java Developer Toolkit
  – Performance Application Programming Interface (PAPI)
• Cray products that require a separate Cray license; each requires its own license:
  – Cross Compiler for Cray XT series systems
  – CrayPat performance analysis tool
  – Cray Apprentice2 performance data visualization tool
• Optional, licensed third-party products (Cray provides a Cray XT support package):
  – PathScale EKOPath Compilers (licensed through Cray and distributed by PathScale LLC)
  – TotalView Debugger (licensed and distributed by TotalView Technologies, LLC)
2.9 Additional Enhancements to System Administration and System Maintenance

This section describes additional software enhancements for Cray XT series system administration and system maintenance. The following topic areas are discussed.

- General administration enhancements
- System resiliency improvements
- I/O and devices enhancements

For compatibility issues and differences that you should be aware of when upgrading your Cray XT series system to this release, see Section 3.6, page 38.

2.9.1 General System-Administration Enhancements

The following general system-administration enhancements are provided in the Cray XT series 2.0 release.

2.9.1.1 Release Upgrade, Configuration, and Installation Enhancements

In addition to specific enhancements made to support the new CNL kernel (see Section 2.1.10, page 10), the following improvements are included as part of the Cray XT series 2.0 release.

- The `fdisk` utility is now used instead of the `sfdisk` utility to back up current SMW software.

- The Cray System Management Workstation (SMW) software is now provided on separate media. A new `CRAYSWMinstall.sh` script and a new `SMWinstall` installation program are provided. Also, a new installation guide, *Cray System Management Workstation (SMW) Software Installation Guide*, is provided. Customers may now upgrade their SMW software independently.

- The Cray XT programming environment is now an independent release. A new installation guide, *Cray XT Series Programming Environments Installation Guide*, is provided.
After you have installed the Cray System Management Workstation (SMW) release software, you now have the ability to install the remainder of the Cray XT series 2.0 release software to an alternative boot root and shared root while your Cray XT system remains operational. Because the configuration steps may be performed on the alternative location, the required amount of dedicated Cray XT system time is reduced.

The new XTinstall.conf configuration file is provided with this release. Based on the settings you choose in the XTinstall.conf file, the XTinstall.conf file then updates other configuration files, eliminating some manual configuration steps.

The new system set file, /etc/sysset.conf, is provided with this release. A system set is a complete set of all file systems and boot images; it provides a way to group several disk partitions on the boot RAID into a set, which is given a label. For detailed information see the sysset.conf(5) man page and the Cray XT Series Software Installation and Configuration Guide.

A new installation program, XTinstall, is also provided with this release. You can do a full installation or an upgrade from a 1.5 update to the base 2.0 release or from the base 2.0 release to a 2.0 update to a system set by using the label of that system set as a parameter to the XTinstall program. (The sysset.conf file must be created prior to running the XTinstall program.) For example, it is possible to have two or more UNICOS/lc operating system versions on the boot RAID from which you choose the one to boot, or you can use system sets to have a PRODUCTION system set and a TEST system set.

The xtcli bootCfg or xtcli partCfg commands can specify which boot image to use to boot a system. This boot image is related to the system set in that it has information (in the parameters file) about the root disk device for the boot node and the mount point for the shared root file system. A boot image can be a disk partition in a system set, or it can be a file in a directory (which must be in the same relative location on both the SMW and the boot root).

Based on your software licensing agreement with Cray, you will soon be able to specify the compute node OS component (Cray XT CNL software, Cray XT Catamount/CVN software, or Cray X2 software) of your Cray XT release package. When this offering becomes available, it will be announced in the associated Cray XT 2.0.xx update’s README document.
2.9.1.2 Release Switching

You can change which version of SMW software, including the firmware flash level, and which version of UNICOS/lc is used when your Cray XT system is booted. The new `xtsmwrelswitch` command and `xtrelswitch` script change settings on the SMW in preparation to boot the Cray XT system. They can work in conjunction with multiple boot images and system sets.

The `xtsmwrelswitch` command manages the active CRMS release or the UNICOS/lc release. To switch to a different UNICOS/lc release, the `xtsmwrelswitch -a OS=` command is used. The `xtsmwrelswitch` command then runs the `xtrelswitch` script in the background. The `xtrelswitch` script performs the UNICOS/lc release switching by manipulating symbolic links in the file system and by setting the default version of module files that are loaded at logon.

The new `xtsmwrelswitch(8)` and `xtrelswitch(8)` man pages are provided.

2.9.1.3 Linux 2.6.16 Kernel on Service Nodes

Although the operating system on the service nodes is based on the SUSE LINUX Enterprise Server 9 release, the kernel running on service nodes has been upgraded to the Linux 2.6.16 kernel from SUSE LINUX Enterprise Server 10 release. Linux 2.6.16 is the same kernel version that is provided for CNL compute nodes.

2.9.1.4 Enhanced Boot Procedure

The boot image now contains boot images for service nodes and for Catamount and CNL compute nodes. The boot system is told which image/load file to use when booting a particular node or set of nodes.

The `xtcli boot_type` option allows you to specify the operating system to boot on a given node.

The `xtbootsys crms_boot_loadfile` and `xtbootsys crms_boot_sdb_loadfile` automation file procedures were added.

The SMW boot manager and the `xtbootimg` command were modified to support this feature.

Revised `xtcli_boot(8)` and `xtbootsys(8)` man pages are provided.
2.9.1.5 Native IP Default

Effective with the Cray XT series 2.0 release, native IP (SSIP) is the default IP used by Cray XT nodes to communicate with each other. Effective with the Cray XT series 2.1 release, IP over Portals (IPPO) will no longer be supported.

2.9.1.6 ARP_OVERHEAD Parameter Added in xt.conf File

The ARP_OVERHEAD parameter was added in the xt.conf file, which resides in /opt/xt-boot/default/etc/. This parameter can be used to further increase the size of ARP tables.

2.9.1.7 Realm-Specific IP Addressing (RSIP) for Cray XT CNL Compute Nodes

Note: RSIP is not supported for Cray XT compute nodes running the Catamount kernel or the CVN capability that supports dual-core processing on compute nodes running the Catamount kernel.

Realm-Specific IP Realm-Specific internet Protocol (RSIP) is an IP address translation technique that can be used to safeguard many private Internet addresses behind a single public Internet address. RSIP allows CNL compute nodes and the service nodes to share the IP addresses configured on the external Gigabit and 10 Gigabit Ethernet interfaces of network nodes. By sharing the external addresses, customers may rely on their system’s use of private address space and do not need to configure compute nodes with addresses within the site’s IP address space.

Note: This release supports RSIP for IPv4 TCP and UDP transport protocols only; IP Security and IPv6 protocols are not supported.

Also, because multicast addresses are not supported on the HSN, this implementation does not support RFC 3105, which describes the use of Service Location to locate RSIP servers. (The Service Location Protocol (SLP) requires multicast addressing to function.)

The Cray XT implementation of RSIP consists of two main pieces: the RSIP client and RSIP server or gateway. The RSIP client is run on internal nodes such as compute nodes to allow those nodes to reach external IP networking resources. The RSIP server is run on service nodes and provides RSIP support for the RSIP clients as well as the IP gateway function for IP packets to and from those RSIP clients. The xtrsipcf command generates and optionally installs the necessary RSIP client and server configuration files. The procedure to manually install server configuration files is described in Cray XT Series System Management. (Support for automatic installation of server configuration files is deferred.)
Using the information stored in the Central Data Repository (CDR), all CNL compute nodes are configured to be RSIP clients by default. Service nodes without their own IP networking hardware can be configured as RSIP clients using the `xtrsipcfg` command with the `-a` option. All Cray XT CNL compute node clients reference a common configuration file, `/etc/krslip.conf`. A CNL compute node is configured as a client node when the node's identifier (NID) is included in `/etc/krslip.conf`. This configuration file is installed on the boot image for the compute node.

RSIP servers must be selected based on machine-specific networking hardware configuration. RSIP servers can only be run on service nodes that have a local external IP interface, such as a 10 GbE network interface card (NIC). Only one external interface will be used per service node. RSIP servers should not be run on service nodes providing Lustre services. Sites are highly encouraged to set up dedicated network nodes for the purpose of running RSIP servers.

Each client reserves a number of ports from the server. Currently, this value is set to 64. Each server will be configured to allow a total of 51,808 ports for RSIP clients, which allows for 809 clients per server. Future releases will allow the RSIP client port consumption to scale more closely with actual use.

For each server, the networking hardware configuration must be updated and the RSIP service must be enabled. The RSIP server configuration file, `/etc/rsipd.conf`, contains the configuration information used by the RSIP server, `rsipd`. To enable the RSIP server service, use the `chkconfig` command on each RSIP server node.

For more information about RSIP, see Cray XT Series System Management and the `rsipd(8)`, `xtrsipcfqg(8)`, and `rsipd.conf(5)` man pages.

**2.9.1.8 Enhancements to the System-Management Graphical User Interface**

The Java SE Runtime Environment (JRE) 6 (also called version 1.6.0 or 1.6) is installed on the SMW for using the XTGUI.
The following graphical user interface enhancements were made for this release:

- Console log output can now be displayed from selected nodes. A new filtered console information tab displays console log information for the nodes you select. Choosing nodes reduces the volume of log output to examine. You can create one or more log filters.

To view filtered console log output, select **Components** from the component detail tab, then the **View Console Output** option, and then the filtered console tab that appears.

The XTGUI help system includes directions for using this feature.

- The **Show Active Cmds** menu option was replaced with the **Show Sessions** menu option. Choosing this menu causes a window to pop up that contains a table of active sessions. If you right-click on a row in the table, a pop-up menu is shown, which has two options: **Show Session Members** and **Abort Session**. Choosing **Show Session Members** causes another window to pop up that lists which components are affected by the selected active session. Choosing **Abort Session** causes the session to be aborted.

The XTGUI also adds new displays to support the new System Environment Data Collections (SEDC) feature; see Section 2.9.2.2, page 30, for descriptions of these displays.

### 2.9.1.9 xtdumps sys Data Collection May Be Customized

An new example file, `example.xtdumps-plugin`, is included in the `/opt/cray/etc` directory and provides techniques to help you customize your own xtdumps plugin so it can collect additional data during an xtdumps session. The `example.xtdumps-plugin` file uses explicit (absolute) paths when collecting data from the SDB syslog file.

The following plugin file data array members were added: `config,use_tftp,partition,l1list,partition,status,partitionl0list,partition,nodelist, and partition,sslist.`

For additional enhancements to the xtdumps, see Section 2.9.1.10.

### 2.9.1.10 Additional New and Enhanced Commands

In addition to the commands noted in this document that changed to support specific new features, the following new and enhanced commands are provided for this release.
• **xtbootsys:**
  - The new --partition partition-name command-line option specifies which partition to initialize and boot.
  - The new CRMS_PARTITION environment variable indicates the default partition name if the --partition partition-name command-line option is not used. Valid values of the CRMS_PARTITION environment variable are of the form p#, where # is in the range 0 through 31.

• **xtbounce:**
  - The new --partition partition-name command-line option specifies which partition to restart.
  - The new CRMS_PARTITION environment variable indicates the default partition name if the --partition partition-name command-line option is not used. Valid values of the CRMS_PARTITION environment variable are of the form p#, where # is in the range 0 through 31.
  - The new -o stop_after key word is used to indicate that xtbounce should terminate after certain stages, such as after halting nodes.

• The new xtclass -r retries | --retries retries option specifies the number of times to retry the request. The default is 6.

• The new xtdaemonconfig command dynamically configures CRMS daemons. Each configurable daemon has a unique list of configurable items. For detailed information, see the xtdaemonconfig(8) man page.

• The new xtdiagstat -D option increases the utility’s debug level.

• The new xtdiagtar -D option increases the utility’s debug level.

• **xtdumpsys:**
  - The new --interactive option causes xtdumpsys to prompt the user before performing certain operations.
  - The new --partition partition-name command-line option specifies which partition to dump.
  - The new CRMS_PARTITION environment variable indicates the default partition name if the --partition partition-name command-line option is not used. Valid values of the CRMS_PARTITION environment variable are of the form p#, where # is in the range 0 through 31.
The following additional environment variables were added: 

- XTDUMPSYS_PARALLEL, XTDUMPSYS_X2_LDUMP, XTDUMPSYS_XTHB_CYCLES, 
- XTDUMPSYS_XTHB_TIMEOUT, XTDUMPSYS_XTMEM2FILE_MAXIDS, 
- XTDUMPSYS_XTPTLTRACE_MAXIDS, and XTDUMPSYS_XTRSH_MAXIDS.

- Several new options were added to the xtptltrace command; see the xtptltrace(8) man page.

- The new --restart-daemons, --start-daemons, and --stop-daemons options restart, start, and stop the background xtconsole, xtconsumer, and xtnetwatch daemons, respectively.

- The rca-helper tool, which is only called from scripts, is now documented in the rca-helper(8) man page.

  **Note:** The xtnidname command is obsolete and can return incorrect data on certain system configurations. See the rca-helper(8) man page for information about replacing the functionality that was provided with this command.

- The new xtsortdiaglog utility sorts temporary log files that are generated by the diagnostics manager. See the xtsortdiaglog(8) man page for additional information.

### 2.9.2 Improvements to System Resiliency

The following enhancements have been made to improve system resiliency.

**Note:** The documentation provided with this release documents the Cray RAS and Management System (CRMS). However, the hardware monitoring and management aspect of the CRMS is being renamed to the Hardware Supervisory System (HSS). The HSS provides all of the CRMS hardware monitoring and management functions. The HSS is also being implemented on future Cray platforms.

#### 2.9.2.1 Login or Network Node Warm Reboot

Login or network nodes cannot be rebooted via a direct node shutdown/reboot. They must be restarted via the CRMS system using the xtbootsys --reboot listid SMW command. CRMS must be used so that the proper kernel is pushed to the node.
After a warm reboot of a login or network node, the administrator may need to restart ALPS (/etc/init.d/alps start), if running CNL on compute nodes, and remount any Lustre file systems (as necessary).

For additional information, see the xtbootsys(8) man page.

2.9.2.2 Enhancements to System Monitoring

The new System Environment Data Collections (SEDC) CRMS manager monitors the system’s health and records the environmental data and status of hardware components such as power supplies, processors, temperature, and fans. SEDC can be set to run at all times or only when a client is listening. By default, this manager is configured to scan the system hardware components only when a client is listening.

There are three major components to the SEDC manager: L0/L1 SEDC daemons, the SMW SEDC server, and the SEDC UI client. L0/L1 SEDC daemons scan the hardware to provide the detailed system environment data, such as fan speed, temperatures, and voltages. The SMW SEDC server manages the system scanning configurations and provides the scanning configuration to L0/L1 SEDC daemons. All L0/L1 SEDC daemons are controlled by the SMW SEDC server. SEDC UI clients subscribe to the scanning result events from L0 and L1 SEDC daemons and present data in a readable format.

The SEDC configuration file (/opt/cray/etc/sedc_srv.ini by default) configures the SEDC server. The /opt/cray/etc/sedc_srv.ini file has entries that control the location of the files to be saved, the size of each file, and the number of files to save. In this file, several other options are configurable; for example the system administrator can create sets of different configurations as groups so that the L0/L1 daemons can scan components at different frequencies. The default /opt/cray/etc/sedc_srv.ini file provided lists all of the items that may be scanned by SEDC. You can change the SEDC configuration file path by setting the CRMS_SEDC_CONF environment variable.

The SEDC manager, sedc_manager, is started by the /etc/init.d/rsms script. After the sedc_manager is started, it reads the sedc_srv.ini configuration file to configure the SEDC monitoring. The sedc_manager saves separate files for each group defined within the sedc_srv.ini file. The file is named to match the name of each group.
If you modify the configuration file while the sedc_manager is running, you must send a -SIGHUP signal to the sedc_manager process, which causes the sedc_manager to reread the sedc_srv.ini file, update the L1/L0 SEDC scanning processes, and close all log files and then reopen them, using the latest configuration information.

For more information about the SEDC manager, see Cray XT Series System Management and the sedc_manager(8) man page.

To view SEDC data, either run the XTGUI or the xtsedcvviewer command-line interface.

Two types of charts may be displayed through the XTGUI:

- A chart that shows data values for a single SEDC scan item, for example, cabinet temperature; the chart shows the current temperature for all cabinets in the system, arranged along a single horizontal line.

- A chart that shows multiple scan items within the same graph, for example, cabinet exhaust temperature and cabinet inlet temperature. On a chart that displays multiple scan items, each scan item is displayed with a separate horizontal axis, and a different color; and a legend indicates which axis corresponds to which scan item.

Positioning the mouse pointer above a data point displays the component name and scan value for the data point.

For more information about these charts, see Cray XT Series System Management or the XTGUI online help system.

### 2.9.3 Enhancements to Input/Output and Devices

The following I/O and devices enhancements are provided in the Cray XT series 2.0 release.
2.9.3.1 Lustre File System 1.4.9

The Lustre file system version 1.4.9 is provided with the Cray XT series 2.0 release. Lustre enhancements include these features:

- Runs on the Linux 2.6.16 kernel. This is transparent to users. Administrators need make no changes.
- Provides improved NFS export of Lustre file systems.
- Accommodates a LUN partition size of up to 8 TB.
- Supports MPI-IO for compute nodes running CNL. To enable file-locking for Linux clients when mounting the Lustre file system on compute nodes running CNL, add the `-o flock` option to the `mount` command.
- Supports Lustre file system quotas.

2.9.3.2 (Deferred implementation) Lustre File System Automatic Failover

Note: The automatic failover feature is supported for compute nodes running CNL. Support for Catamount/CVN compute nodes will be provided in a future Cray XT series release.

Note: Implementing the automatic failover feature requires specific cabling and configuration of the storage devices and file system.

Lustre OSS/MDS failover is a service that switches to a standby server when the primary server fails or the service is temporarily shut down for maintenance. Lustre failover is now automatic.

The automatic Lustre failover framework includes the `xt-lustre-proxy` process, the SDB, and a set of database utilities. The Lustre configuration and failover states are kept in the SDB. Lustre database utilities and `xt-lustre-proxy` are used for Lustre startup/shutdown and failover management.
The `xt-lustre-proxy` is a Linux process that is responsible for Lustre automatic failover in the event of a Lustre service failure. During the boot process, an `xt-lustre-proxy` process starts on each MDS or OSS. It starts up primary Lustre services on the MDS or OSS and then monitors the health of the services through the CRMS system. If there is a node-failure or service-failure event, CRMS notifies the `xt-lustre-proxy` on the backup node to start up the backup services. The primary and backup configuration is specified in the SDB. To avoid both primary and backup services running at the same time, the `xt-lustre-proxy` on the backup node issues a node reset command to shut down the primary node before starting the backup services. The proxy also marks the primary node as dead in the database so that if the primary node is rebooted while the backup system is still running, the `xt-lustre-proxy` on the primary node does not start the services.

**Database tables and utilities:** The system administrator sets up the following SDB tables: `lustre_failover`, `lustre_service`, and `filesystem`. Each SDB table can be populated by executing the related Lustre database table utility. Each utility uses a formatted data file to generate database entries. Each line in a formatted data file should contain a record of comma-delineated pairs of the form `field1=val1, field2=val2`, and so on.

To populate the `lustre_failover` table, the system administrator creates the `lustre_failover` formatted data file to generate database entries and then executes the `xtlustrefailover2db` utility, which uses the `lustre_failover` formatted data file to generate the database entries.

To populate the `lustre_service` table, the system administrator creates the `lustre_serv` formatted data file to generate database entries and then executes the `xtlustreserv2db` utility, which uses the `lustre_serv` formatted data file to generate the database table entries.

To populate the `filesystem` table, the system administrator creates the `filesys` formatted data file to generate database entries and then executes the `xtfilesys2db` utility, which uses the `filesys` formatted data file to generate the database entries.

The `xtlusfoadmin` command displays Lustre automatic failover database tables. It can be used to modify or display fields of a given Lustre automatic failover database table. When it is used to make changes to database fields, failover operation is impacted accordingly. For example, the `xtlusfoadmin` command is used to set file systems active or inactive or to enable or disable the Lustre failover process for Lustre services.
**Backing up SDB table content:** The following set of utilities can be used to dump the database entries to a datafile:

- The `xtdb2lustrefailover` utility, which dumps the `lustre_failover` table in the SDB to the `lustre_failover` data file.
- The `xtdb2lustreserv` utility, which dumps the `lustre_service` table in the SDB to the `lustre_serv` data file.
- The `xtdb2filesys` utility, which dumps the `filesystem` table in the SDB to the `filesys` data file.

**System startup:** The `xt-lustre-proxy` process starts Lustre. It uses `lconf`, the Lustre configuration .xml file, and the SDB to determine the correct OSS or MDS to start. The mechanism of starting up Lustre using `lconf` command is not needed. However, `lconf` can still be used directly to bypass the automatic failover framework.

**System shutdown:** The `xt-lustre-proxy` process replaces `lconf` for a system shutdown. The `xt-lustre-proxy` process accepts only a `SIGTERM` to do a graceful shutdown. Any other way of termination, such as `SIGKILL`, triggers the failover process and results in a failure event delivered to the backup node. The backup node then issues a `node reset` command to shut down the primary node. Use the `pkill xt-lustre-proxy` command to prevent the failover.

Status of failover is recorded in console and log messages.

The Cray XT series 2.0 release includes the following man pages: `xt-lustre-proxy(8)`, `xtlusfoadmin(8)`, `xtlustrefailover2db(8)`, `xtdb2lustrefailover(8)`, `xtlustreserv2db(8)`, `xtdb2lustreserv(8)`, `xtfilesys2db(8)`, and `xtdb2filesys(8)`.

2.9.3.3 Logical Volume Manager (LVM) Module Provided

The Logical Volume Manager (LVM) kernel module for service nodes is included with the Cray XT series 2.0 release package, but the module is not loaded by default. For additional information, see your Cray service representative.
2.9.3.4 Support for Engenio Disk Devices

A Cray XT series system uses one or more RAID (redundant array of independent disks) subsystems for disk storage. With the Cray XT series 2.0 release, support is added for the following Engenio devices from LSI Logic Corporation.

- Engenio 6998 devices for parallel RAID
- Engenio 3992 devices for system boot RAID and for parallel RAID

2.10 Resolved Critical and Urgent Software Problem Reports (SPRs)

The list of customer-filed critical and urgent SPRs closed with the Cray XT series 2.0 release is included in the Cray XT Series 2.0 Release Errata provided with the Cray XT series 2.0 release package.
This chapter describes compatibility issues and functionality changes to be aware of after upgrading from the Cray XT series 1.5 release.

For temporary limitations of this release and changes identified after the documentation for this release was packaged, see the *Cray XT Series 2.0 Release Limitations*.

### 3.1 Users Must Recompile Applications

Because of changes made with the Cray XT series 2.0 release, users must recompile applications when moving from the base Cray XT series 1.5 release or from a 1.5 update to the Cray XT series 2.0 release. Customers running the 2.0 limited availability (LA) release running CNL on compute nodes do not need to relink when moving to the 2.0 general availability (GA) release.

**Caution:** Not recompiling an application can result in undefined behavior and may cause the application to fail or hang or may cause a Cray XT series node to fail.

### 3.2 The `lsma64` Compiler Option No Longer Required

The `lsma64` compiler option was used during a transition period. It should not be used any longer for either Catamount or CNL applications because the `-default64` compiler option is now supported for SHMEM. The `-default64` option directs the linker to use the default64 version of the SHMEM library.

### 3.3 MPICH2 Edge Color All-to-all Algorithm for Non-power-of-two Ranks

The send-to-one-rank-receive-from-another-rank for all-to-all when the message size is greater than 32 KB was replaced with an edge color algorithm. If the `MPICH_ALLTOALLV_SENDRECV` environment variable is set, the edge color algorithm is used for transfers greater than 512 bytes and less than or equal to 32 KB as well. This results in modest improvements in all-to-all transfer rates at high-process counts.
3.4 Change to Upper Limit for Quantum for Catamount/CVN Nodes

The behavior has changed for the use of the SET_QUANTUM mbox option to the process control thread (PCT). Previously, a user program could set the quantum (SET_QUANTUM) to be up to 60 seconds, but this would cause a PCT heartbeat failure. As a result, the user quantum that can be set by the user has decreased from 60 seconds to 5 seconds. In CVN mode, both processes are always set to the same value.

3.5 Man Pages in roff Format

Cray developed man pages included with the Cray XT series 2.0 release package are now provided in roff format. This change was made to improve usability, both printed form and online display.

3.6 Additional System Management Compatibility Issues and Differences

In addition to the feature information described in Chapter 2, system administrators should also note the following compatibility issues and differences when upgrading to the Cray XT series 2.0 release.

For temporary limitations of this release and changes identified after the documentation for this release was packaged, see the Cray XT Series 2.0 Release Limitations.

3.6.1 Release 2.0 Upgrade-related and Configuration-related Changes

The following information is provided to help you prepare for installing the Cray XT series 2.0 release. Installation procedures are included in the Cray System Management Workstation (SMW) Software Installation Guide, the Cray XT Series Software Installation and Configuration Guide, and the Cray XT Series Programming Environments Installation Guide, which are provided with the Cray XT series 2.0 release package.

3.6.1.1 Supported Upgrade Path

The System Management Workstation (SMW) 3.0.1 release is provided. The SMW 3.0.1 release includes the Cray RAS and Management System (CRMS). You must be running at least version 1.5.27 to upgrade to the SMW 3.0.1 release.

The UNICOS/lc 2.0 release is provided. You must be running at least version 1.5.27 to upgrade to the UNICOS/lc 2.0 release.
3.6.1.2 Upgrade Packages Provided as Downloadable ISO Images

Upgrade release packages for UNICOS/lc and the SMW software are now provided as ISO images and may be downloaded from a Cray website.

3.6.1.3 New LUN Size Upper Limit

The LUN size upper limit has increased to 8 TB.

3.6.1.4 The /etc/lustre/llrd.conf File Requires Listing All Lustre MDS and OSS Nodes

For systems running Catamount/CVN on compute nodes, to reduce instances of Lustre MDS clean-up problems during eviction, the /etc/lustre/llrd.conf file should be updated to include a list of all the Lustre MDS and OSS nodes.

3.6.1.5 Upgrading System Software Requires SDB Update

Customers upgrading to this release must update the SDB database schema from version 4.6 to 4.8. This procedure is documented in the Cray XT Series Software Installation and Configuration Guide.

3.6.1.6 New Boot Images Directory for CNL Images

For systems running CNL on compute nodes, a new /opt/xt-images/ directory is provided. The directory contains a master/default/ directory and a templates/default directory. By default, the xtclone utility copies the contents of the master image (located in the master/default directory of the image repository) to the image directory.

3.6.1.7 Native IP Is Default as of Cray XT Series 2.0 Release

Effective with the Cray XT series 2.0 release, native IP (SSIP) is the default IP.

3.6.1.8 Last Release Supporting IPPO

Effective with the Cray XT series 2.1 release, IP over Portals (IPPO) will no longer be supported.
3.6.2 Operational-related Changes

3.6.2.1 Location of xtbootimg, xtptltrace, smwhotbackup, and smwsnapshot Commands Changed

The xtbootimg, xtptltrace, smwhotbackup, and smwsnapshot commands on the SMW now reside in /opt/cray/bin. The xtptltrace command previously resided in /opt/cray/sys/diag/bin. The xtbootimg, smwhotbackup, and smwsnapshot commands previously resided in /opt/xt-boot/default/bin/snos64.

3.6.2.2 xtnidname Command Obsolete

The xtnidname command is obsolete and can return incorrect data on certain system configurations. See the rca-helper(8) man page for information about replacing the functionality that was provided with this command.

3.6.2.3 Boot Manager Parameter String Maximum Increased

The boot manager on the SMW now accepts a parameter string to a maximum of 512 characters. The previous maximum was 256 characters.

3.6.2.4 CNL and Catamount/CVN Operational Differences

For CNL and Catamount/CVN operational differences, see Section 2.1.10, page 10.

3.6.2.5 Cray RAS and Management System (CRMS) Evolving to Hardware Supervisory System (HSS)

The Cray RAS and Management System (CRMS) is evolving to the next generation Cray hardware and software management system, which is called the Hardware Supervisory System (HSS). HSS is also an integrated system of hardware and software that monitors Cray system components, manages hardware and software failures, controls startup and shutdown processes, manages the system interconnection network, and displays the system state to the administrator. The HSS interfaces with all major hardware and software components of a Cray XT series system and will be built upon to support future Cray platforms. This evolution is transparent and requires no changes from administrators.
This chapter describes the documentation that supports the Cray XT series 2.0 software release.

4.1 CrayDoc Documentation Delivery System

The CrayDoc documentation delivery system, along with product documentation, is provided with each Cray software release. The CrayDoc software runs on any operating system based on UNIX systems or systems like UNIX including Mac OS X, Linux, BSD, and anywhere else that Perl and Apache can be compiled from source code with freely available (GNU) tools. The installation and administration of the CrayDoc server software and Cray documentation are described in the CrayDoc Installation and Administration Guide.

4.2 Accessing Product Documentation

With this release, Cray provides books, man pages, and third-party documentation. These documents are provided in the following ways:

- CrayDoc, the Cray documentation delivery system that allows you to quickly access and search Cray books, man pages, and in some cases, third-party documentation. Access this HTML and PDF documentation via CrayDoc at the following URLs:
  - The local network location defined by your system administrator
  - The CrayDoc public website: docs.cray.com

- Man pages. Access man pages by entering the man command followed by the name of the man page. For more information about man pages, refer to man(1) man page by entering:

  `% man man`

- Third-party documentation not provided through CrayDoc. Access this documentation, if any, according to the information provided with that product.
4.3 Books Provided with This Release

The books provided with this release are listed in Table 1 and in Table 2, which also note whether each book was updated. Many books are provided in HTML and all are provided in PDF.

Note: Two additional documents are provided with this release. The Limitations for Cray XT Series 2.0 GA includes a description of temporary limitations of this release. The Cray XT Series 2.0 Release Errata includes installation and configuration changes identified after the installation documentation for this release was packaged and lists customer-filed critical and urgent SPRs closed with this release. A printed copy of these documents is included with the release package; they are also available from your Cray representative. You should also contact your Cray representative about Cray XT system-related information addressed in Field Notices (FNs).

Table 1. Cray-developed Books Provided with This Release

<table>
<thead>
<tr>
<th>Book Title</th>
<th>Number</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cray XT Series Software Release Overview (this document)</td>
<td>S–2425–20</td>
<td>Yes</td>
</tr>
<tr>
<td>Cray XT Series Software Installation and Configuration Guide</td>
<td>S–2444–20</td>
<td>Yes</td>
</tr>
<tr>
<td>Cray XT Series System Overview</td>
<td>S–2423–20</td>
<td>Yes</td>
</tr>
<tr>
<td>Cray XT Series System Management</td>
<td>S–2393–20</td>
<td>Yes</td>
</tr>
<tr>
<td>Cray XT Series Programming Environments Installation Guide</td>
<td>S–2465–20</td>
<td>New</td>
</tr>
<tr>
<td>CrayDoc Installation and Administration Guide</td>
<td>S–2340–41</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 2. Third-party Books Provided with This Release

<table>
<thead>
<tr>
<th>Book Title</th>
<th>Number</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lustre Operations Manual</td>
<td>S–6540–149</td>
<td>Yes</td>
</tr>
<tr>
<td>PGI User’s Guide</td>
<td>S–6516–70</td>
<td>Yes</td>
</tr>
<tr>
<td>PGI Fortran Reference</td>
<td>S–6518–70</td>
<td>Yes</td>
</tr>
<tr>
<td>PGI Tools Guide</td>
<td>S–6517–70</td>
<td>Yes</td>
</tr>
<tr>
<td>PGI Server 7.0 and Workstation 7.0 Installation and Release Notes</td>
<td>S–6539–70</td>
<td>Yes</td>
</tr>
<tr>
<td>AMD Core Math Library (ACML)</td>
<td>S–6511–31</td>
<td>Yes</td>
</tr>
<tr>
<td>PAPI User’s Guide</td>
<td>S–6515–35</td>
<td>Yes</td>
</tr>
<tr>
<td>PAPI Programmer’s Reference</td>
<td>S–6514–35</td>
<td>Yes</td>
</tr>
<tr>
<td>SuperLU Users’ Guide</td>
<td>S–6532–10</td>
<td>No</td>
</tr>
<tr>
<td>FLEXlm End Users Guide</td>
<td>S–6508–95</td>
<td>No</td>
</tr>
</tbody>
</table>

If your site has ordered the PBS Pro product for your Cray XT series system, the following books are also provided. All PBS Pro books are provided in PDF. The PBS Pro Release Overview, Installation Guide, and Administration Addendum is provided in PDF and HTML format.

Table 3. PBS Pro Books Provided with This Release

<table>
<thead>
<tr>
<th>Book Title</th>
<th>Number</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS Pro Release Overview, Installation Guide, and Administration Addendum</td>
<td>S–2438–81</td>
<td>Yes</td>
</tr>
<tr>
<td>PBS Professional 9.0 User Guide</td>
<td>S–6500–90</td>
<td>Yes</td>
</tr>
<tr>
<td>PBS Professional 9.0 External Reference Specification</td>
<td>S–6501–90</td>
<td>Yes</td>
</tr>
<tr>
<td>PBS Professional 9.0 Administrator Guide</td>
<td>S–6502–90</td>
<td>Yes</td>
</tr>
<tr>
<td>PBS Professional 9.0 Quick Start Guide</td>
<td>S–6510–90</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: If you are installing another batch system, please contact the appropriate vendor for the necessary installation, configuration, and administration information.
4.4 Man Pages Provided with This Release

- Application launch commands: `aprun(1)` for applications running on CNL compute nodes; `yod(1)` for applications running on Catamount/CVN compute nodes
- System view commands: `xtshowcabs(1), xtshowmesh(1)`
- Cray-specific MPI man page: `intro_mpi(1)`
- Cray SHMEM man pages: start with `intro_shmem(1)`
- Single-system view (SSV) man pages: `xhostname(1), xtkill(1), xtpsi(1), xtwho(1)`
- Cray System Management Workstation (SMW) man pages
- UNICOS/lc man pages
- Lustre file system automatic failover man pages: `xt-lustre-proxy(8), xtlusfoadmin(8), xtlustrefailover2db(8), xtdb2lustrefailover(8), xtlustresv2db(8), xtdb2lustresv(8), xtfilsys2db(8), xtdb2filsys(8)`
- Modules software package man pages: `module(1), modulefile(4)`
- Man pages are provided for the following third-party products:
  - PGI compiler commands: `cc(1), CC(1), ft(1), and f77(1)`
  - GNU compiler commands: `gcc(1) and g77(1)`
There are two sets of man pages for each compiler suite:

- PGI compilers:
  - cc(1), CC(1), ftn, and f77(1) tell users how to compile and link
  - pgcc(1), pgCC(1), pgf95(1), and gf77(1) document PGI-specific options

- PathScale compilers:
  - cc(1), CC(1), and ftn tell users how to compile and link
  - pathcc(1), pathCC(1), pathf95(1), and eko(1) document PathScale-specific options

- GCC compilers:
  - cc(1), CC(1), ftn, and f77(1) tell users how to compile and link
  - gcc(1), g++(1), gfortran(1), and g77(1) document GCC-specific options

- MPICH2
- LAPACK
- ScaLAPACK
- BLACS
- PAPI
- Lustre
- SUSE LINUX

**Note:** Third-party and open source man pages are not packaged in CrayDoc; use the `man` command on your Cray XT series system to access these man pages.

If your site ordered CrayPat, man pages are provided: start with `craypat(1)`.

If your site ordered Cray Apprentice2, the `app2(1)` man page is provided. In addition, the Cray Apprentice2 GUI provides online help.

If your site ordered PBS Pro, man pages are provided on the software DVD only: start with `pbs(1B)`. 
4.5 Additional Documentation Resources

Table 4 lists additional resources for obtaining documentation not included with this release package.

Table 4. Additional Documentation Resources

<table>
<thead>
<tr>
<th>Product</th>
<th>Documentation Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNU compilers</td>
<td>Documentation for the GNU C and Fortran compilers is available at <a href="http://gcc.gnu.org/onlinedocs/">http://gcc.gnu.org/onlinedocs/</a></td>
</tr>
<tr>
<td>MPICH2</td>
<td>Additional documentation is available in HTML and PDF formats from the Argonne National Laboratory website at <a href="http://www-unix.mcs.anl.gov/mpi/mpich2">http://www-unix.mcs.anl.gov/mpi/mpich2</a>. Additional information about the MPI-2 standard is available at <a href="http://www.mpi-forum.org/docs/docs.html">http://www.mpi-forum.org/docs/docs.html</a></td>
</tr>
<tr>
<td>ScaLAPACK</td>
<td>The ScaLAPACK Users’ Guide and ScaLAPACK tutorial are available in HTML format at <a href="http://www.netlib.org/scalapack/">http://www.netlib.org/scalapack/</a></td>
</tr>
<tr>
<td>FFTW 2.1.5 and FFTW 3.1.1</td>
<td>Additional FFTW 2.1.5 and FFTW 3.1.1 documentation is available at <a href="http://www.fftw.org">http://www.fftw.org</a></td>
</tr>
<tr>
<td>SuperLU</td>
<td>Additional SuperLU documentation is available at <a href="http://crd.lbl.gov/~xiaoye/SuperLU/">http://crd.lbl.gov/~xiaoye/SuperLU/</a></td>
</tr>
<tr>
<td>PAPI</td>
<td>Additional PAPI documentation is available at <a href="http://icl.cs.utk.edu/papi/">http://icl.cs.utk.edu/papi/</a></td>
</tr>
<tr>
<td>MySQL</td>
<td>MySQL documentation is available at <a href="http://www.mysql.com/documentation">http://www.mysql.com/documentation</a></td>
</tr>
<tr>
<td>FLEXlm</td>
<td>Additional FLEXlm documentation is available at <a href="http://www.macrovision.com/products/legacy_products/flexport/index.shtml">http://www.macrovision.com/products/legacy_products/flexport/index.shtml</a></td>
</tr>
<tr>
<td>glibc</td>
<td>glibc documentation is available at <a href="http://gcc.gnu.org/onlinedocs">http://gcc.gnu.org/onlinedocs</a></td>
</tr>
<tr>
<td>GNet</td>
<td>GNet documentation is available at <a href="http://www.gnetlibrary.org">http://www.gnetlibrary.org</a></td>
</tr>
<tr>
<td>GLIB</td>
<td>GLIB documentation is available at <a href="http://developer.gnome.org/doc/API/2.0/glib/index.html">http://developer.gnome.org/doc/API/2.0/glib/index.html</a></td>
</tr>
<tr>
<td>RPM</td>
<td>RPM documentation is available at <a href="http://www.rpm.org">http://www.rpm.org</a></td>
</tr>
</tbody>
</table>
4.6 TotalView Documentation From TotalView Technologies, LLC

TotalView documentation for Cray XT series systems is available from TotalView Technologies. For information about TotalView documentation, refer to http://www.totalviewtech.com/Documentation/alltvdocs.htm.

4.7 PathScale Compilers Documentation From PathScale LLC

The PathScale compiler documentation is available from PathScale LLC. For information about PathScale compiler documentation, refer to http://www.pathscale.com/docs.html.

4.8 Cray Glossary

A Cray Glossary of terms specific to Cray XT series systems is included with CrayDoc. The entire Cray Glossary is available on the CrayDoc public website:

http://docs.cray.com
5.1 Supported Upgrade Path

The System Management Workstation (SMW) 3.0.1 release is provided. The SMW 3.0.1 release includes the Cray RAS and Management System (CRMS). You must be running at least version 1.5.27 to upgrade to the SMW 3.0.1 release.

The UNICOS/lc 2.0 release is provided. You must be running at least version 1.5.27 to upgrade to the UNICOS/lc 2.0 release.

5.2 Optional Products Offered Directly From Cray Inc.

The Cray XT series 2.0 software release supports the following optional products offered directly from Cray Inc.:

- PBS Pro 8.1
- CrayPat version 3.2.3 or later
- Cray Apprentice2 version 3.2.3 or later
- Cross Compiler for Cray XT series systems version 2.0
5.3 TotalView Offered From TotalView Technologies, LLC

You can order a special implementation of the TotalView 8.2 debugger for Cray XT series systems from TotalView Technologies, LLC. You cannot order TotalView directly from Cray Inc.

TotalView provides source-level debugging of MPI applications. For information about ordering, installing, using, and maintaining TotalView, refer to http://www.totalviewtech.com/productsTV.htm.

5.4 PathScale Compilers Offered From PathScale LLC

You can order the PathScale 3.0 compilers from Cray Inc.; your order will be fulfilled directly by PathScale LLC. Cray provides the PathScale-sup package, which contains module file support for the PathScale 3.0 compilers on Cray XT series systems. For information about the PathScale EKOPath 3.0 compilers, see http://www.pathscale.com/ekopath.html.

5.5 Contents of the Release Package

This release package includes:

- System Management Workstation 3.0.1
- UNICOS/1c 2.0, which includes:
  - Linux kernel 2.6 and SUSE LINUX 9.0 Service Pack 2 running on service nodes
  - Catamount/CVN 2.0 compute node software and CNL 2.0 compute node software

  **Note:** Based on their software licensing agreement with Cray, customers will soon be able to specify the compute node OS component (Cray XT CNL software, Cray XT Catamount/CVN software, or Cray X2 software) of their Cray XT release package. When this offering becomes available, it will be announced in the associated Cray XT 2.0.xx update's README document.

- Application Level Placement Scheduler (ALPS) 1.0
- Cray XT 2.0 file system, which is based on the Lustre 1.4.9 file system from Cluster File Systems, Inc.
Also included with the UNICOS/lc release package are these related products:

- GNU glib 2.4.2
- GNet 2.0.5 network library
- Modules 3.1.6 user environment management utility
- MySQL 4.0 database manager
- FLEXlm 9.5 license manager

- Programming Environment 2.0, which includes:
  - PGI 7.0.7 Fortran, C, and C++ compilers and tools
    
    **Note:** PGI 7.0.7 requires the FLEXlm license manager, which controls the number of simultaneous users.
  - The GNU 4.2.1 C and Fortran 90 compilers
  - Cray MPICH2 1.0.2 library of MPI-2 routines
  - ACML 3.6 library of BLAS, LAPACK, and FFT routines
  - IOBUF library (iobuf) version 1.0.5 for systems running Catamount or CVN on compute nodes
  - Cray XT series LibSci 10.0.1 library, which includes BLAS (including the LibGoto), LAPACK (including LibGoto), ScalAPACK, BLACS, SuperLU, and Iterative Refinement Toolkit (IRT) of ScalAPACK, BLACS, SuperLU routines, and the Cray LibSci FFT interface
  - FFTW 2.1.5 and FFTW 3.1.1
  - Performance API (PAPI) 3.5
  - GNU debugger remote proxy server, xt gdb, for Cray XT series systems

- CrayDoc software suite and the documentation, described in Chapter 4, page 41

- A printed copy of this release overview

- A printed copy of the *Cray System Management Workstation (SMW) Software Installation Guide*

- A printed copy of the *Cray XT Series Software Installation and Configuration Guide*
5.6 Licensing

Cray licenses the following as separate products for Cray XT series systems under a Cray license agreement:

- Cray XT OS binary (which provides rights to UNICOS/lc operating system and its components)
  
  **Note:** Source Code Option: The Cray XT OS license is binary by default. Certain U.S. customers may be eligible to obtain a Cray XT OS buildable source license for an additional fee. For more information regarding source code for the Cray XT OS, contact your sales representative.

- Lustre Parallel File System (contractual rights to Lustre are included with the Cray XT OS license for some initial customers)

- Cray XT PGI Programming Environment (licensed by number of simultaneous users)

- PathScale EKOPath Compiler Suite (licensed by number of simultaneous users)

- Cray XT Linux Cross Compiler Programming Environment (optional product)

- PBS Pro Batch Subsystem (optional product)

- CrayPat Performance Collector (optional product)

- Cray Apprentice2 Performance Analyzer (optional product licensed by number of simultaneous users)

The `PAPILicnotices(7)` and `superlulicnotices(7)` man pages list the license notices for the software that Cray supplies for the Cray XT Series Programming Environment in conjunction with the software and documentation copyright distribution requirements. The `gnulicnotices(7)` man page lists the public license notice for the GNU Free Documentation used in the UNICOS/lc release.
For more information about licensing and pricing, contact your Cray sales representative or send e-mail to crayinfo@cray.com.

Customers outside the United States and Canada must sign a Letter of Assurance before software can be shipped to them. For questions about whether you have signed this agreement, or questions about which software requires this letter, send e-mail to crayinfo@cray.com.

5.7 Ordering Software

This release package is distributed by order only to customers who have signed a license agreement for the Cray software that includes this product. The most current revision of the release package is supplied. To receive any upgrades to a given Cray product, the customer must also have a signed support agreement for this Cray software.

You can order the release package from the Cray Software Distribution Center in any of the following ways:

**E-mail:**
-orderdsk@cray.com

**CRInform (for subscribers):**
-crinform.cray.com

Click on the Order Cray Software link.

**Telephone (inside U.S., Canada):**
1–800–284–2729 (BUG CRAY), then 605–9100

**Telephone (outside U.S., Canada):**
+1–651–605–9100

**Fax:**
+1–651–605–9001

**Mail:**
Software Distribution Center
Cray Inc.
1340 Mendota Heights Road
Mendota Heights, MN 55120–1128
USA

Software will be shipped by ground service or 5-day international service.
This chapter describes the customer services that support this release.

6.1 Technical Assistance with Software Problems

If you experience problems with Cray software, contact your Cray service representative. Your service representative will work with you to resolve the problem. If you choose to have full- or part-time support on site, your on-site personnel are your primary contacts for service. If you have elected not to have on-site support, call or send e-mail to the Cray Customer Support Center:

E-mail:
support@cray.com

Telephone (inside U.S., Canada):
1–800–950–2729 (CRAY)

Telephone (outside U.S., Canada):
+1–715–726–4993

Fax:
+1–651–605–9001

6.2 CRInform System

The CRInform system is the information and problem-reporting system for Cray customers who are CRInform subscribers.

You are a CRInform subscriber if your site has a software license agreement and software support agreement. Access CRInform at:

crinform.cray.com
Ask your system administrator for your password. Some of the things a subscriber to CRInform can do include:

- Report software problems (SPRs)
- Request technical assistance (RTAs)
- Communicate with other Cray system users
- Read about software problems reported at other sites
- Learn about solutions to various problems
- Order Cray software
- View Cray Service Bulletin

The CRInform program automatically logs events pertinent to your Cray system site as news items, so you do not have to search through the system for new information. The logged events include Software Problem Report (SPR) or Request for Technical Assistance (RTA) activity, new orderable software, new issues of the Cray Service Bulletin, new field notices (FNs), new software release documents, new software problem fix information, new marketing information, and new CRInform program information. You can also get automatic e-mail notification of any or all of the news items.

6.3 Training

To find out more about Cray training, contact your Cray representative or contact us in any of the following ways:

**E-mail:**
wwwtng@cray.com

**Web:**
www.cray.com/training/

**Fax:**
+1–715–726–4991

**Mail:**
Technical Training
Cray Inc.
P.O. Box 6000
Chippewa Falls, WI 54729–0080
USA
6.4 Cray Service Bulletin (CRSB)

The CRInform site provides access to the online Cray Service Bulletin, also called CRSB, which contains descriptions of software problems, information about service procedures or agreements, and announcements of product upgrades and future products for the private use of Cray customers.

6.5 Cray Public Website

The Cray public website offers information about a variety of topics and is located at:

www.cray.com
A  
Accessing Cray documentation, 41  
Accounting, GNU process, 19  
ACML, 51  
ALPS placement scheduler, apsched, 14  
apbridge daemon, 14  
apinit daemon, 14  
apkill command, 6, 13  
apmgr command, 13  
appl application  
management, 12  
placement, 12  
scheduling, 12  
Application Level Placement Scheduler (ALPS), 12  
Application Level Placement Scheduler (ALPS), 12  
aprun command, 6, 13, 15  
apstat command, 6, 13  
apsys daemon, 14  
apwatch daemon, 14  
ARP_OVERHEAD parameter, 25  
Automation file procedure, 24  

B  
Batch system, 12, 20  
BLACS, 51  
BLAS, 51  
Books, accessing, 41  
Boot image, 23  
Boot Manager  
parameter string maximum, 40  
Boot procedure, 24  
Booting release version, 24  

C  
C compiler, 51  
C++ compiler, 51  
Catamount, 50  
Catamount/CVN release package, 23, 50  

CNL  
administration, 10  
programming considerations, 7  
user-visible functionality, 6  
CNL and Catamount/CVN differences, 8  
BusBox commands, 8  
I/O buffering, 9  
MPI, 8  
OpenMP, 9  
PAPI, 9  
SHMEM, 9  
CNL compute node kernel, 5  
CNL release package, 23, 50  
cnselect command, 16  
Collecting data during xtdumpsys session, 27  
Comaptibilities, 37  
Compiler option -lsma64, 37  
Compiler target environment variable, 6  
Compilers, 51  
Compute node kernel, CNL, 5  
Configuration steps automated, 23  
Console log output, 26  
Contact information, 55  
Training, 56  
Cray Apprentice2, 49, 52  
Cray RAS and Management System (CRMS), 29  
Cray Service Bulletin, 56–57  
Cray websites, 57  
CRInform, 55  
training, 56  
Cray websites, publications, 41  
Cray websites, support, 55  
CrayPat, 49, 52  
CRInform, 55  
CRMS, 50  
upgrade requirement, 38, 49  
CRMS manager, 30  
CRSB, 57
Customer services, 55
Customer Support Center, 55
Customer-filed critical and urgent SPRs, 35
Customs, 53

D
Debugger, TotalView Technologies TotalView, 50
documentation, 47
Default IP, 39
DEFAULT: keyword, 18
Differences, 37
Disk devices, 35
Distribution Center, 53
Documentation, 41
Documentation, accessing, 41

E
Engenio 3992 devices, 35
Engenio 6998 devices, 35
Enhancements, 5
Environment variable
  MPICH_ALLTOALLVW_SENDRECV, 37
Errata, 5, 42
/etc/attr.defaults file, 11, 16
/etc/attr.xthwinv file, 16
/etc/attributes file, 16
/etc/hosts file
  CNL nodes, 11
/etc/lustre/llrd.conf file, 39
/etc/sysconfig/alps system information file, 14
/etc/sysset.conf file, 23
/etc/xt.conf file, 16
example.xtdumpsys-plugin file, 27
Export license, 53

F
Failover
  Lustre file system, 32
  fdisk utility, 22
FFT, 51
Field notices (FNs), 56
FLEXlm license manager, 50
Fortran compiler, 51

G
glibc, 51
Glossary, 47
GNet, 50
GNU compilers, 51
GNU glibc, 51
GNU process accounting, 19
GUI, 26

H
Hardware environmental information, 30
Hardware Supervisory System (HSS), 29

I
Installation program
  SMWinstall, 22
  XTinstall, 23
Installation-related enhancements, 22
Installing to alternative boot root and shared root, 23
IOBUF library (iobuf), 51
IP default, 39
IP Over Portals (IPPO) support, 39
IPPO, 25

K
Kernel, 50
Kernel version, 24
Key enhancements, 1

L
LAPACK, 51
Letter of assurance, 53
LibSci, 51
Licensing, 52–53
Limitations, 5, 42
Linux, 50
Logical Volume Manager (LVM) kernel module, 34
Login node
  warm reboot, 29
-1sma64 compiler option, 37
LUN size, 32, 39
Lustre, 50
Lustre file system, 32
  automatic failover, 32
  database tables, 33
  database utilities, 33
Lustre man pages, 45

M
Man pages, 44
Man pages format, 38
Man pages, accessing, 41
Man pages, third party, 44
Microkernel, 50
Modules utility, 50
MPICH_ALLTOALLV_SENDRECV environment variable, 37
MPICH2, 51
MySQL, 50

N
Native IP, 25, 39
Network node
  warm reboot, 29
New features, 5
Node attributes, 15–16
  hardware attributes, 17
  job launch, 15
  setting, 16, 18
  site-specific attributes, 17
  viewing, 15
nodeid keyword, 18

O
Open source man pages, 45
Operating system, Catamount, 50
Operating system, Cray Linux, 50
Operating system, SUSE LINUX, 50
Operating system, UNICOS/lc, 50

/reopt/cray/etc/sedc_srv.ini configuration file, 30
/reopt/xt-images/ directory, 10, 39
Optional products, 49
Ordering software, 53

P
PAPI, 51
PathScale ECKOPath compilers
  documentation, 47
PathScale EKOPath compilers, 50
PathScale support package, 50
PBS Pro, 49
  pbs module, 20
  size resource, 20
PBS Pro batch system, 12, 20
PGI compilers, 51
Pricing, 53
Problems, 55
Process accounting, 19
Programming Environment products, 20
Publications, 41
Publications, accessing, 41

Q
Quantum upper limit, 38

R
rca-helper command, 40
rca-helper tool, 29
Reader comments, 3
Realm Specific IP Addressing (RSIP), 25
Recompile requirement, 37
Release package, 49–50
Release switching, 24
Request for Technical Assistance (RTA), 56

S
ScaLAPACK, 51
SDB update requirement, 39
SDBATTR=/etc/attributes, 16
sedc_manager, 30
Cray XT™ Series 2.0 Software Release Overview

SET_QUANTUM mbox option, 38
sfdisk utility, 22
smwhotbackup command location, 40
SMWinstall installation program, 22
smwsnapshot command location, 40
Socket off-load (SOLD) facility, 19
Software Problem Report (SPR), 56
SQL update
  update_sdb_4_8.sql, 39
Subscriber, CRInform, 55
SuperLU, 51
Support agreement, 52–53, 55
Support Center, 55
SUSE LINUX, 50
SUSE LINUX man pages, 45
System Environment Data Collections (SEDC), 30
System Management Workstation (SMW)
  upgrade requirement, 38, 49

T
Technical support, 55
Third-party man pages, 45
TotalView, 49–50
documentation, 47
TotalView Technologies TotalView, 49
Training, 56

U
UNICOS/lc, 50
  upgrade requirement, 38, 49
UNICOS/lc operating system enhancements, 5
Upgrade path supported, 38
Upgrade release package
downloadable, 39
Upgrading requirement, 39

V
VisIT application, 19

X
xt-lustre-proxy, 33
xtbootimg command location, 40
xtbootsys
  new options, 28
xtbounce
  new options, 28
xtclass command, 28
xtcli boot_type option, 24
xtclone command, 10
xtdaemonconfig command, 28
xtdiagstat command, 28
xtdiagtar command, 28
xtdumpsys
  example plugin file, 27
xtdumpsys--interactive option, 28
XTGUI
  Java version, 26
SEDC, 31
xthwinv command, 17
XTinstall installation program, 23
XTinstall.conf file, 23
xtlusfoadmin command, 33
xtnidname command
  obsolete, 40
xtopview command, 16
xtpackage command, 10
XTPE_COMPILE_TARGET target environment
  variable, 6
xtprocadmin command, 15, 18
xtptltrace command, 29
xtptltrace command location, 40
xtrelswitch script, 24
xtsedcviewer, 31
xtshowcabs command, 7, 13
xtshowmesh command, 7, 13
xtsmwrelswitch command, 24

Y
yod command, 15
  -tracemalloc option, 19