Cray SMW 5.1.UP02 Software Release Announcement

The Cray SMW 5.1.UP02 update release is available. This software announcement applies only to Cray XT6, Cray XT6m, Cray XE6, Cray XE6m, and Cray XE5 System Management Workstations (SMWs).

The SMW 5.1.UP02 release supports

- New or initial software installations on the following platforms:
  
  Cray XE6, Cray XE6m, Cray XT6, Cray XT6m, and Cray XE5 systems

- Upgrade installations from SMW 5.0 on the following platforms:
  
  Cray XT6 and Cray XT6m systems

- Update installations from a previous version of SMW 5.1:
  
  Cray XT6, Cray XT6m, Cray XE6, Cray XE6m, Cray XE5 and Cray XE5m systems

- DEFERRED: Migration support from SMW 4.0 to SMW 5.1.UP02 for Cray XT5 and Cray XT4 systems is deferred until a successful customer test has been completed. Look for a future announcement on support of CLE 2.2 migrations for Cray XT5 and Cray XT4 systems in the February 2011 timeframe.

The emphasis for the SMW 5.1 release is documented in the Cray System Management Workstation (SMW) 5.1 Software Release Overview, S-2482-51. Information about additional features provided in SMW 5.1 updates is included in this README file (see below).

The SMW 5.1.UP02 release was tested with the Cray CLE 3.1.UP02 release on Cray XE (Gemini) systems and Cray XT (SeaStar) systems, limited to supported platforms mentioned above. Sites can upgrade to SMW 5.1.UP02 prior to upgrading to Cray CLE 3.1.UP02. However, the recommended software combination is SMW 5.1.UP02 release and Cray CLE 3.1.UP02.

NOTE: The Cray Gemini congestion management changes are a part of both the SMW 5.1.UP02 and CLE 3.1.UP02 releases. It is recommended that sites upgrade to both SMW 5.1.UP02 (or later) and CLE 3.1.UP02 (or later).

NOTE: The SMW 5.1.UP02 Release REQUIRES an upgrade to SLES 11 SP1 base OS.

Read the SMW 5.1.UP02 Errata before installing this software package. In
particular, review the Errata item "High Speed Network Parameter Changes Needed for Cray XE5 Systems".

RELEASE OVERVIEW
----------------

- 5.1.UP02 - The following new features have been added for this update. (Previous features and fixes included in prior updates are included in this release.)

  o SLES 11 SP1 base OS upgrade REQUIRED.

    SLES 11 support ends with SMW 5.1.UP02. Novell provides security fixes for the SLES 11 release, but only until the end of 2010. Starting in January of 2011, security fixes will be supplied only for SLES 11 SP1 or later. You must upgrade to SLES 11 SP1 before upgrading to SMW 5.1.UP02. FN 5743 announced this requirement and plan for an extra 60 minutes to upgrade to SLES 11 SP1. For more information about upgrading from SLES 11 to SLES 11 SP1, refer to FN 5743.

  o Workaround for mzlogmanagerd consuming SMW CPU resources. The SMW for Cray XE systems now plays a key role in congestion management and mzlogmanagerd consumes a large portion of SMW CPU resources. In the event of a message flood, this can cause L0's to auto-throttle. mzlogmanagerd is no longer automatically started at boot time for both Cray XE and XT systems.

  o Improvements have also been made in the following areas

    - Reduced L0 timeouts at boot
    - Network Resiliency

  o Fix for HT lockups on SeaStar systems. Starting with SMW 4.0.UP03, sync flood is turned on. This resulted in HT lockup issues for large (~20 cabinet) systems. Fixes are in SMW 5.1.UP02 and CLE 3.1.UP02 releases to resolve

    Bug 763222 - System failures due to HT lockup
    Bug 764733 - XT6 system failure due to HT lockup
    Bug 765295 - System crashed due to Node HT lockup followed by HSN collapse
    Bug 767114 - HT lockup, unable to halt PPC, unable to clear or scrub SS memory
    Bug 765217 - modprobe of RCA hung during node boot
    Bug 767216 - Update xtfwstat and xtfwlog for the latest SS firmware
Cray recommends running SMW 4.0.UP03A in combination with the CLE 2.2.UP03A release.

- An updated "Installing Cray System Management Workstation (SMW) Software", S-2480-5102, is provided; it also provides procedures to upgrade the SMW base operating system from SUSE Linux Enterprise Server 11 (SLES 11) to SLES 11 Service Pack 1 (SLES 11 SP1).

- "Migrating from System Management Workstation (SMW) 4.0 to SMW 5.1 Software", S-0033-5102, supports upgrading from SMW 4.0 to SMW 5.1.UP02 with the CLE 2.2.UP03 release software or a special CLE 2.2.UP02 release package. S-0033-5102 is provided to support ONLY Cray XT systems that are upgrading to Cray XE systems; migrating SMW 4.0 software to SMW 5.1 software for Cray XT systems is deferred until testing is completed.

**NOTE:** Support for migration from SMW 4.0 is currently deferred until testing has completed.

**xtdiscover** changes:

- The default boot node name and default SDB node name provided by xtdiscover changed for Cray XE systems:

  - for Cray XT (SeaStar) systems:
    
    default boot node name is c0-0c0s0n0
    default SDB node name is c0-0c0s0n3

  - for Cray XE (Gemini) systems:
    
    default boot node name is c0-0c0s0n1
    default SDB node name is c0-0c0s2n1

- The following Caution was added to the xtdiscover(8) man page:

  Caution: Using the --testconfig option after a system is installed results in loss of node information because there is only one HSS database; --testconfig does not operate on a separate test database configuration.

- The steps to save and restore the HSS database are now provided in the NOTES section of the xtdiscover(8) man page.

**xtcli boot cfg -i cpio_path** option:
The maximum length of a cpio path is limited to 80 characters, and it must begin with the / character.

- Xtbounce changes:

  - Two new entries to xtbounce initialization file variables:

    - `timeout_retry_count`: Number of times to retry the timed out stage; default is 5.

    - `timeout_retry_delay`: Amount of time to wait between retries of stages that have timed out; default is 5 seconds.

- During the gather_cab_pwr_states phase of xtbounce, if the HSS software on an L1 and any of its L0s is out of sync, error messages such as the following will be printed during the xtbounce:

  
  
  ***** gather_cab_pwr_states *****

  18:28:42 - Beginning to wait for response(s)

  ERROR: rs_phys_node2ascii() failed for node struct 0xb7e70150080700f8
  ERROR: rs_phys_node2ascii() failed for node struct 0xb7e70150080700f8
  ERROR: rs_phys_node2ascii() failed for node struct 0xb7e70150080700f8
  ERROR: rs PHYS_node2ascii() failed for node struct 0xb7e70150080700f8

  If this occurs, it indicates that the L0 software is at a different revision than the L1 software. xtbounce will print a list of cabinets for which this error has occurred. The message will be like:

  ERROR: power state check error on 2 cabinet(s)
  WARNING: unable to find c0-0 in err_cablist
  WARNING: unable to find c0-2 in err_cablist

  This error is an indication that when the HSS software was previously updated, the L1s and the L0s were not updated to the same version.

  To correct this error, cancel out of xtbounce (with ctrl-C), wait approximately five minutes for the <xtbounce> related activities on the L0s to finish, then reboot the L1(s) and their associated L0s to get the HSS software synchronized. Following this, the xtbounce may be run once again.

- After a system is shut down and an "abnormal" condition occurs, such as when a node dies abnormally during the shutdown, the HSS system may experience time outs during a subsequent xtbounce.
These time outs will occur because some nodes in the system, particularly service nodes, appear halted from the CLE perspective but not from the HSS point of view.

When a node halts, it stops sending RCA heartbeats to the L0 controller. As a result, the HSS determines that a booted node has died, and l0sysd performs some postmortem analysis of the node. This analysis can take several minutes to complete. During this interval, l0sysd does not respond to any events.

If this situation occurs, wait at least five minutes from the time the system completes the CLE shutdown before you initiate a bounce/boot cycle. It is good practice to always wait five minutes after shutdown is completed before booting the system.

o Cray XE systems: rtr_manager is no longer used; rtr manages the process.

o Cray XE systems: xthwerrlog can now watch errors as they occur or parse the logfile. The -f option is no longer required; if not supplied, errors are watched in 'real time.' All current options except -e|--end and -s|--start are supported in the new mode.

o New commands:

hssds_init -- Creates the Cray HSS data store

nid2nic -- For Cray XE systems: Prints all nid-to-nic address mappings

xtclear_link_alerts -- For Cray XE systems: Use xtclear_link_alerts before using xtwarmswap --add to clear alert flags on any nodes, Geminis, or links that relate to the blade being added.

xtresview -- For Cray XE systems: Displays the current state of cabinets, blades, and links, and whether any have failed or been warm swapped out

o xtbootdump changes:

- Added the ability for xtbootdump to pass skip options to xtdumpsys (--skip option)

- Now recognizes timed out nodes

o xtdimminfo change:
The output changed: socket # and core # were added to the comment section for HT link and channel type errors; the DIMM and CPU type errors have the same output.

o xtdaemonconfig changes:

  Added three ER parameters: min_mem_footprint, max_mem_throttle_sec, and hi_priority_events

o xtbootsys changes:

  - New built-in function crms_affect_daemons
  - Changed auto.xtshutdown so that it only kills the daemons associated with the boot session

o xtnmi now requires either an id-list or a --p partition option as a command-line argument.

o The rtr(8) man page was corrected; common options, options that are for Cray XE systems only, and options that are for Cray XT systems only are now documented.

o The xtdumpsys(8) man page was clarified. Cray provides an example plug-in file to collect additional information from service nodes; we strongly recommend installing this file after customizing it for your site.

o The xtcli_part(8) man page was corrected; to activate a partition that has been deactivated, partitions must have at least one member, one boot node, one SDB node, and a cpio path.

- 5.1.UP01B

  o This update improves reliability and resolves several critical and urgent problems.

- 5.1.UP01A

  o This update improves reliability and resolves several critical and urgent problems.

- 5.1.UP01 - The following new features have been added for this update.

  o Cray Gemini Network Resiliency and Congestion Management
  For information about this feature see Cray Linux Environment (CLE) 3.1 Software Release Overview Supplement, section 1.1.1,
Cray Gemini Network Resiliency and Congestion Management.

- Application Completion Reporting (ACR)

  Application Completion Reporting (ACR) provides a way to track, manage, and report application data, which gives system administrators an effortless way to see how system resources are being allocated.

  Data provided by ACR reports can make it easier to see where system resources are being used so that recommendations can be made for application improvement, different allocation schemes, or use of different time slices.

  ACR extends the data persistence schema provided by Cray Management Services (CMS) and provides three commands to examine and report ALPS application data.

    - Use the mzjob command to examine reservation data; you can specify the format of the search output so that you can use it in other programs or scripts.

    - Use the mzreport command to examine application completion status information; the command supports claims (applications), jobs (reservations), or user completion status.

    - Use mz2attr to read and display node attributes.

ACR requires both CLE 3.1.UP01 or later and SMW 5.1.UP01 or later; support for this feature was deferred in earlier SMW and CLE releases.

The following API library functions were planned for the ACR feature but have not been implemented. Man pages for these functions, previously deferred, have been deleted:

mz_create_label.3
mz_delete_label.3
mz_free_label.3
mz_free_labellist.3
mz_free_labelnamelist.3
mz_free_labelnodes.3
mz_get_all_labelnames.3
mz_get_all_labels.3
mz_get_labelnodes.3
For more information refer to Using Cray Management Services (CMS) (S-2484-5101) and the mzjob(8), mzreport(8) and mz2attr(8) man pages.

- The updated version of "Installing Cray System Management Workstation (SMW) Software," S-2480-5101, is provided with the SMW 5.1.UP01 release.

- Start and End Actions of Additional Commands Automatically Logged: Each time one of the commands listed below starts and ends, an entry is now automatically logged into the daily /opt/craylog/commands/log.YYYYMMDD file. The directory and file are owned by crayadm (group and user). These actions are logged for the following commands:

  xtcli                  xtmmr
  xtdaemonconfig        xtnmi
  xtfileio              xtwarmswap
  xtmem2file            xtrsh
  xtmemio

This logging feature was added to assist in problem diagnosis because the actions of these commands change the state of the system or other commands depend on the actions of these commands. A log entry includes:

- A standard date/time as time stamp

- The PID of the process that is being logged

- The directive (START or END) specifying start or end of the command

- The name of the command

The START entry includes parameters specified in the command line; the END entry returns the status of the command. For example:

  2010-08-11 14:37:02 23889 START command parameter(s)
  2010-08-11 14:37:02 23889 END exit status 0

When the command being logged takes parameters or is invoked with a configuration or initialization file, the parameters or contents of the file are also logged.

NOTE: For a command listed above that runs on both the SMW and on
the L1/L0 controllers, only when the command is run on the
SMW will the command's actions be logged in the
/opt/craylog/commands/log.YYYYMMDD file.

o SMWinstall --media=directory Option:

The new SMWinstall --media=directory option specifies the directory
where the software media has been placed. The default is
/media/cdrom. If the media is not in the default location, the
--media option is required.

o xtbootsys Command Enhancements:

- The new xtbootsys --use-logger option sends xtnetwatch output
to Cray Management Services (CMS) in addition to the default
log file.

Note: With 5.1.CS00A, this option was "Deferred implementation".

- The global Tcl variable data: fn,config was added; it provides
the fully-qualified path of the file containing system
configuration information. The value should not be modified by
an automation file. The Default is {}.

o xtdimminfo Command Output Enhancement:

In xtdimminfo command output, socket and core numbers were added in
the comment section for HT link and channel type errors. This
additional information will help point to the core that caused a
sync flood event.

- 5.1.CS00A.PS02 - Required patch set for Cray XE6 systems running
5.1.CS00A. This patch set consists of a HSS RPM for SMW.
This is a cumulative patch set, meaning patch set
5.1.CS00A.PS01 fix is included in this PS02 patch set and
does not have to be applied separately.

- 5.1.CS00A - Update to SMW 5.1 pre-release for important reliability
based fixes

- 5.1.CS00 - Base SMW 5.1 pre-release package.

For install instructions see the Installing Cray System
<table>
<thead>
<tr>
<th>Bug #</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>766425</td>
<td>c17-2c1s4 Had an EPO and reroute did not recover rest of system</td>
</tr>
<tr>
<td>765910</td>
<td>If reroute fails system should be shutdown</td>
</tr>
<tr>
<td>765616</td>
<td>rtr glibc memory corruption causes rtr -P to fail</td>
</tr>
<tr>
<td>765363</td>
<td>Deadlock timeouts, Lustre connectivity errors, System unresponsive.</td>
</tr>
<tr>
<td>764978</td>
<td>xtbounce clears Gemini alerts, resulting in potential link resiliency disaster</td>
</tr>
<tr>
<td>764943</td>
<td>L0 reboot causes system interrupt</td>
</tr>
<tr>
<td>764925</td>
<td>l0 timeout problems at boot time can significantly add to the time it takes to boot</td>
</tr>
<tr>
<td>764904</td>
<td>state_manager segfault when trying to activate a partition</td>
</tr>
<tr>
<td>764890</td>
<td>a failing link resulted in 70+ nodes crashing for various reasons;</td>
</tr>
<tr>
<td></td>
<td>network congestion</td>
</tr>
<tr>
<td>764764</td>
<td>state_manager segfaults when activating partition with no boot/sdb nodes</td>
</tr>
<tr>
<td>764754</td>
<td>L0 timeout after a warm swap of a blade</td>
</tr>
<tr>
<td>764748</td>
<td>Warm swap remove may not mark links down, preventing warm swap add from working</td>
</tr>
<tr>
<td>763748</td>
<td>Need to downclock 2.2Ghz B3+ opteron to 2.1Ghz to match B2 opteron parts</td>
</tr>
<tr>
<td>761946</td>
<td>xtnlrd should do something more visible when reroute fails</td>
</tr>
</tbody>
</table>

RELEASE INFORMATION

Distribution for this release is restricted to Cray XE6, Cray XE6m, Cray XT6, Cray XT6m, and Cray XE5 systems. Cray XT5 and Cray XT4 support is DEFERRED at time of this release. The SMW 5.1.UP02 release package must be ordered through the Cray Software Distribution Center in any of the following ways:

E-mail: orderdsk@cray.com
CrayPort (for subscribers): crayport.cray.com
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

NOTE: Refer to the README file for entire set of information related
to this release package.