



# Technical Bulletin

No. 451 ♦ April 12, 2010

## CHAOS 4.3/TOSS 1.3 Release Notes

### Summary

CHAOS 4.3/TOSS 1.3 represents a refresh of the CHAOS software stack against Red Hat Enterprise Linux (RHEL) 5.4, along with other minor bug fixes and enhancements. The major software updates contained in CHAOS 4.3/TOSS 1.3 are:

Software	CHAOS 4.2	CHAOS 4.3
SLURM	2.0	2.1
Lustre	1.6.6	1.8.2

### Red Hat Release Notes

Red Hat's release notes for RHEL 5.4 are available from their Web site at the following URL:

[http://www.redhat.com/docs/en-US/Red\\_Hat\\_Enterprise\\_Linux/5.4/html/Release\\_Notes/](http://www.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/5.4/html/Release_Notes/)

Some highlights from Red Hat's release notes are provided below.

- **OFED 1.4:** The InfiniBand drivers in the RHEL 5.4 kernel have been updated to upstream OpenFabrics Alliance Enterprise Distribution (OFED) version 1.4-rc3.
- **GNU Compiler Collection (GCC) 4.4:** The technology preview version of GCC has been updated from GCC 4.3 to GCC 4.4.
- **getrusage(2) System Call Enhancement:** The `getrusage(2)` system call is used to examine the resource usage of a process. In RHEL 5.4, the `getrusage(2)` call has been enhanced with the ability to report resource usage for a single thread instead of the process as a whole. To query resource usage for the current thread, the `RUSAGE_THREAD` 'who' parameter has been added to the existing 'who' parameters: `RUSAGE_SELF`, which queries the process as a whole, and `RUSAGE_CHILDREN`, which reports resource usage for the process and all its children.
- **New glibc malloc Behavior:** The upstream glibc has been changed recently to enable higher scalability across many sockets and cores. This is done by assigning threads their own memory pools and by avoiding locking in some situations. The amount of additional memory used for the memory pools (if any) can be controlled using the environment variables `MALLOC_ARENA_TEST` and `MALLOC_ARENA_MAX`. `MALLOC_ARENA_TEST` specifies that a test for the number of cores is performed once the number of memory pools reaches this value. `MALLOC_ARENA_MAX` sets the maximum number of memory pools used, regardless of the number of cores.

The glibc in the RHEL 5.4 release has this functionality integrated as a technology preview of the upstream malloc. To enable the per-thread memory pools, the environment variable `MALLOC_PER_THREAD` needs to be set in the environment. This environment variable will become obsolete when this new malloc behavior becomes default in future releases. Users experiencing contention for the malloc resources could try enabling this option.

## Lustre 1.8

The version of Lustre in CHAOS 4.3/TOSS 1.3 has been updated to Lustre version 1.8. The changes in Lustre 1.8 (since version 1.6) include:

- Numerous bug fixes.
- Provides future upgrade path. Lustre 2.0 will be a new protocol.
- Version-Based Recovery (VBR).
  - Improves robustness of client recovery operations.
  - Allows Lustre recovery to continue even if multiple clients fail at the same time as the server.
- Asynchronous journal commits.
  - Improves write performance for some workloads.

## SLURM 2.1

The version of SLURM in CHAOS 4.3/TOSS 1.3 has been updated to SLURM 2.1, the latest stable release. It is not expected that any user-visible impact will result from this release except for the following notable enhancements.

- A `--detail` option to `scontrol show job` to display the CPU/memory allocation information on a node-by-node basis.
- A new job wait reason, `ReqNodeNotAvail`: Required node is not available (down or drained).
- Environment variable support to `sattach`, `salloc`, `sbatch`, and `srun` to permit user control over exit codes so application exit codes can be distinguished from those generated by SLURM. `SLURM_EXIT_ERROR` specifies the exit code when a SLURM error occurs. `SLURM_EXIT_IMMEDIATE` specifies the exit code when the `--immediate` option is specified and resources are not available. Any other non-zero exit code would be that of the application run by SLURM.

Other updates include:

- Tab completion for the `module(1)` command for bash and tcsh users. (The `module(1)` command is part of the environment-modules suite that is used to manage the addition/removal of packages in `/opt` from the user environment).
- The `io-watchdog` utility now exports a client API to get and set the current `io-watchdog` timeout. (See `io-watchdog(3)` for more information).
- An updated version of git. (git 1.6.6 is available under `/opt [ module load git/1.6.6 ]`)

**If you have any questions, please contact the LC Hotline—  
send e-mail to [lc-hotline@llnl.gov](mailto:lc-hotline@llnl.gov) or [lc-hotline@pop.llnl.gov](mailto:lc-hotline@pop.llnl.gov) (SCF)  
or phone (925) 422-4531**

Web Pages	
<a href="https://computing.llnl.gov/">https://computing.llnl.gov/</a>	User Information
<a href="https://computation.llnl.gov/icc/">https://computation.llnl.gov/icc/</a>	Department home page
<a href="https://lc.llnl.gov/computing/techbulletins/">https://lc.llnl.gov/computing/techbulletins/</a>	Technical Bulletins
<a href="http://www.llnl.gov/computing/">http://www.llnl.gov/computing/</a>	SCF only