Technical Bulletin

No. 447 (Revised) ♦ July 30, 2009

CHAOS 4.2/TOSS 1.2 Release Notes

Summary

CHAOS 4.2/TOSS 1.2 represents a refresh of the CHAOS software stack against Red Hat Enterprise Linux (RHEL) 5.3, along with other minor bug fixes and enhancements. The major software updates contained in CHAOS 4.2/TOSS 1.2 are:

Software	CHAOS 4.1	CHAOS 4.2
SLURM	1.3	2.0

Red Hat Release Notes

Red Hat's release notes for RHEL5.3 are available from their Web site at the following URL:

http://www.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/5/html/Release_Notes/index.html

libmpich Changes

CHAOS 4.2 contains an MPI change that may cause some previously working MPI applications to fail at startup with a message similar to:

```
./a.out: error while loading shared libraries: libmpich.so.1.0: cannot open
shared object file: No such file or directory
```

Users who encounter this issue should add the MPI search path "/usr/local/tools/mvapich-gnu/lib/shared" to the environment variable LD_LIBRARY_PATH. This fix may need to be in your batch script (after any #MSUB or #PSUB commands) or in a wrapper around your executable. If you need help with this work-around, please contact the LC Hotline. This work-around should solve any problems with existing executables, but it is not recommended for use with newly generated executables (because it will mask potential MPI library problems).

A more permanent fix is to relink the application with one of LC's MPI wrappers (i.e., mpiicc, mpicc) from /usr/local/bin. If relinking does not fix the problem or it occurs in a newly compiled executable, please inform the LC Hotline for assistance (because the RPATH in your application is not getting set as expected).

More Details on the libmpich.so Issue

The official LC versions of the MVAPICH mpich libraries and the MPI wrapper scripts are installed in /usr/local, and they have not changed with this update. Using one of the LC MPI wrapper scripts (i.e., mpicc) in /usr/local/bin forces your executable to use the proper /usr/local/tools version by specifying the proper executable RPATH (e.g., mpicc adds `-Wl,-rpath,/usr/local/tools/mvapich-gnu/lib/shared` to the link line for you). LC has compiled the MPI libraries for each compiler to prevent compiler-conflict-related MPI library problems, which is why using the gnu-compiled MPI library path work-around for newly compiled executables is not recommended. For the majority of applications where the setting of this RPATH worked as expected, the CHAOS 4.2 upgrade should not cause any problem.

Prior to the CHAOS 4.2 update, even those applications with the RPATH not set properly would likely still work, but they could have subtle MPI problems due to compiler conflicts. In CHAOS 4.0 and CHAOS 4.1, there was a system default GNU-compiled mpich library that would be automatically used when the RPATH was not set properly in the executable. Because this system default caused some serious problems for those users intentionally using non-LC versions of mpich, this "default" MPI was removed by request in CHAOS 4.2.

Removing this default MPI path in CHAOS 4.2 exposes MPI applications that were not using the intended LC MPI. Analysis of the few OCF applications that encountered this problem with CHAOS 4.2 (that were reported to the LC Hotline) has uncovered several subtle and unexpected ways that may result in a user having an executable without a properly set RPATH. Some of these unexpected traps have been removed in CHAOS 4.2 (thus, relinking may correct the problem), but some application link lines may need to be changed to set the proper RPATH. Please contact the LC Hotline if relinking doesn't resolve the problem and you need assistance with how to change your link line.

Third-Party Software under /opt

Users may notice the introduction of software packages under the /opt directory on CHAOS 4.2 systems. Most notably for users, the system-provided MVAPICH and OpenMPI packages have been moved from their previous location under /usr. Users should see no impact from this move because these libraries were not the default on LC systems.

In addition to MPI, several other packages have been moved or added to /opt. Packages that are built with MPI, such as system-provided mpiBench and other MPI test utilities, were moved for consistency, and third-party compilers were added so that MPI builds using these compilers could be packaged. A growing list of other software that doesn't fit in the default file system path will also be added to /opt.

The packaging of software under /opt represents the initial steps toward a common packaging and deployment solution between LLNL, LANL, and Sandia, and as such is a work in progress. At this time, use of software under /opt is considered optional for users, though not completely discouraged.

Packages provided in /opt may be added and removed from a user's environment using the environment modules (http://modules.sourceforge.net/) package, which is provided as part of CHAOS/TOSS. To list the available packages in /opt, run

```
modules avail
```

Software is loaded into the user's environment using

```
modules load <pkg>/<version>
```

where /<version> is optional, and unloaded with

```
modules unload <pkg>
```

SLURM 2.0

The version of SLURM in CHAOS 4.2/TOSS 1.2 has been updated to the latest release, SLURM 2.0. There are several user-visible changes in this release, but it is not expected that many users will need to change how they use SLURM relative to the SLURM 1.3 release.

Most SLURM 2.0 changes pertain to scheduling enhancements that will not be seen in our current configuration. One useful SLURM enhancement is the ability to limit a job's memory use per allocated processor using the `--mem-per-cpu` option (with a configurable default value). On processor-scheduled computers, this can be used to schedule the memory with the processors. For

example, a job requiring half of the memory on a node should be allocated half of the processors on the node and be changed accordingly. A job exceeding its memory allocation would be killed.

The `sacct` command, used for viewing job accounting records, has been enhanced. The `--nodes` option can be used to filter jobs by the node on which they executed. Also the default `starttime` is midnight of the current day rather than the start of the database.

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

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