TORQUE Resource Manager 4.2.10 Release Notes

The release notes file contains the following sections:

Differences

This section contains differences in previously existing features that require a change in configuration or routine.

4.2.10

Port using the Munge API instead of forking

TORQUE now uses the Munge API, rather than forking, when configured with the `--enable-munge-auth` option.

pbsdsh -o Option Captures Sderr

The pbsdsh -o option is modified to add sderr capturing.

pbs_mom now sets environment variable for NVIDIA GPUs

A new mom config parameter, $cuda_visible_devices, was added to specify whether pbs_mom sets the CUDA_VISIBLE_DEVICES environment variable when it starts a job. The default is TRUE.

4.2.9

qmgr server parameter 'copy_on_rerun'

A new qmgr option: set server copy_on_rerun=[True|False] is available. When set to True, Torque will copy the OU, ER files over to the user-specified directory when the qrerun command is executed (i.e a job preemption). This setting requires a pbs_server restart for the new value to take in effect. Note that the MOMs and the pbs_server must be updated to this version before setting copy_on_rerun=True will behave as expected.

qmgr server parameter 'job_exclusive_on_use'

A new qmgr option: job_exclusive_on_use=[True|False] is available. When set to True, pbsnodes will report job-exclusive anytime 1 or more processors are in use. This resolves discrepancies between Moab and TORQUE node reports in cases where Moab is configured with a SINGLEJOB policy.

scan_for_terminated

Improved performance by moving scan_for_terminated to its own thread.

TORQUE accounting improvements

Two new fields were added to the accounting file for completed jobs: total_execution_slots and unique_node_count. total_execution_slots should be 20 for a job that requests nodes=2:ppn=10. unique_node_count should be the number of unique hosts the job occupied.

4.2.8

Thread pool manager and thread stack sizes modified

The thread pool manager now frees idle nodes. Additionally, the default thread stack sizes changed to a maximum of 8 MB and a minimum of 1 MB.

Exit job improvements

The way TORQUE exits jobs on moms has been improved to reduce the number of stray jobs, make epilogue run only once, and stop the occurrence of single jobs showing up on nodes they should not.

Known issues

The following are known issues in the TORQUE Resource Manager. Following each issue description is an associated issue number in parentheses.

Known issues are aggregated and grouped by the release version for which they first occurred or where reported.

4.2.10

No known issues.

4.2.9

- Some limitations exist in the way that pbsdsh can be used. Please note the following situations are not currently supported:
    - Running multiple instances of pbsdsh concurrently within a single job. (TRQ-2851)
    - Launching a large number of processes in succession (causes pbdsdsh to hang). (TRQ-2890)

4.2.8

- When you upgrade Moab and TORQUE, depending on the versions, you could encounter a problem where the core files are created frequently in `/opt/moab`. You can resolve this problem by removing the old library files from `/usr/local/lib`. (TRQ-1082)

Resolved issues

The following is a list of some key bugs fixed in TORQUE Resource Manager. Following each issue description is an associated issue number in parentheses.

Resolved issues are aggregated and grouped by the release version in which they were resolved.

4.2.10

- **Torque job can kill processes not owned by the job owner.** Fixed issue around unauthorized termination of processes. (CVE-2014-3684, TRQ-2885)
- **qstat -Q <bad_queue_name> prints queue name twice.** Fixed bug where giving a bad queue name to qstat -Q results in duplicate output. (TRQ-2025)
- **mppnodes hostlist was being re-ordered.** This bug is fixed. (TRQ-2112)
- **GPU nodes where not passed to sister nodes.** This bug is fixed. (TRQ-2837)
- **pbsdsh did not support running multiple instances concurrently.** Improved pbsdsh to better handle simultaneous use of -o and -s options. Also fixed some problems where -o output was sometimes getting truncated. (TRQ-2890)
- **Moab issue caused TORQUE crash.** This is fixed. (TRQ-2921)
- **max_user_queuable issues.** Fixed some bugs causing incorrect max_user_queuable values. (TRQ-2841)
- **Error log messages occurred when exiting jobs.** Refactored the job exit code to vastly improve the time needed to process exiting jobs; this speeds up MOM response times especially when many jobs are exiting at the same time. (TRQ-2448)
- **"Error allocating memory for add_verify_resources" with commas when using hostlist range syntax.** Fixed parsing bug when using hostlist ranges in qsub. (TRQ-2652)

4.2.9

- **XML job log error.** Corrected mismatched <Job_Id> XML tags in the job log. (TRQ-2692)
- **TORQUE accounting problems - jobs without accounting records.** Fixed bug related to accounting records on large systems. (TRQ-2367)
- **TORQUE was leaving behind error and out files when a job was preempted or requeued.** Fixed bug where OU files were being left in spool when job was preempted or requeued. (TRQ-2732)
- **momctl -q clearmsg didn't seem to clear error messages permanently.** Fixed bug where 'momctl -q clearmsg' didn't properly clear error messages. (TRQ-2828)
- **max_user_queuable limit reached; however, there were no jobs in the queue**. Fixed bug where jobs rejected due to max_user_queuable limit reached, yet no jobs in the queue. (TRQ-2795)
- **qsub did not process arguments correctly when a submitfilter is used.** Fixed bug where qsub did not process args correctly when using a submit filter. (TRQ-2646)

- **Reported cput was incorrect.** Fixed bug where reported cput was incorrect. (TRQ-2759)

- **qsub was passing arguments to submit filter all together with interactive jobs.** Fixed bug with submit filter arguments not being parsed during interactive jobs. (TRQ-2707)

- **Build bug reported with MIC libraries.** Fixed build bug related to newer Intel MIC libraries installing in different locations. (TRQ-2653)

- **TORQUE was not dividing GPUS amongst NUMA nodes.** Fixed problem where GPUs were not split between NUMA nodes. You now need to specify which gpus belong to each node board in the mom.layout file. (TRQ-2730)

  A sample mom.layout file might look like:

  ```
  nodes=0 gpu=0
  nodes=1 gpu=1
  ```

  > ⓘ This only works if you use nvml. The nvidia-smi command is not supported.

- **Crash on start up when reading empty array file.** Fixed start up bug related to empty job array (.AR) files. (TRQ-2787)

- **qstat wouldn't parse anything after a bad job ID.** Improved qstat behavior in cases where bad job IDs were referenced in the command. (TRQ-2410)

- **pbsdsh required FQDN even if other elements didn't.** pbsdsh no longer requires FQDN. (TRQ-2632)

- **Some tasks were incorrectly listed as 0 in qstat -a when requested specific nodes.** This has been resolved. (TRQ-2292)

- **Separate headers for multiple jobs IDs provided to qstat.** Fixed output format bug in cases where multiple job IDs are passed into qstat. (TRQ-2411)

4.2.8

- **TORQUE would track the wrong number of processors.** The issue where the total number of execution slots having a count off by one for every Cray compute node has been fixed. (TRQ-2501)

- **A memory leak occurred on aynchronous qrun -a commands.** The memory leak no longer occurs, and a write after free error that could lead to memory corruption was also resolved. (TRQ-2498)

- **For newer versions of NVIDIA drivers, the GPU status was only displayed at pbs_mom startup.** The GPUs now appear in the pbsnodes output. (TRQ-2647)

Product documentation

Technical documentation

The online help for TORQUE Resource Manager 4.2.10 is available in HTML and PDF format on the Adaptive Computing Documentation page.