

Moab

How do I increase the stack size for my jobs?

Issue:

The stack size for compute jobs may be limited to the system's default stack size, which is sometimes inadequate for a job. Since the job is actually forked from the pbs_mom service, a job can never increase it beyond the limit for pbs_mom.

Affected Versions: All

Solution:

The stack size can be increased for the pbs_mom, but it requires modifications to the service.

Details:

First locate the pbs_mom.service file. On CentOS 7, for example, the actual file is "/usr/lib/systemd/system/pbs_mom.service", with a symbolic link with the same name, pointing back to the real file, in "/etc/systemd/system/multi-user.target.wants".

Now modify that file and locate the line with "LimitSTACK=unlimited", or instead of unlimited, you can use your own value. By default, Torque normally sets it to 12,582,912 bytes (or 12,288k). Once that has been set you will need to restart the pbs_mom service, which will cause any jobs currently running on that node to fail.

Once this change has been made and pbs_mom has been restarted, you can run "ulimit -s" from within a job to check the maximum stack size allowed.

Since this requires restarting pbs_mom.service, thus terminating any running jobs, care must be taken to either make this change on all nodes during a maintenance window, or to change the service file at any time, and restart the pbs_mom.service only as nodes become idle. The "rolling upgrade" feature of Torque cannot be used for this, as running that only "exec"s pbs_mom which inherits the currently-running service properties. Fortunately there's at least some help for this.

The recommended way to change all nodes while a system is running is to do it a few at a time. You take a few "offline" (pbsnodes -o), which allows jobs they're running to complete but does not start any new jobs on those nodes. As each node becomes idle, use systemctl to restart the pbs_mom service, then use "pbsnodes -c" to clear the "offline" state of that node.

This whole process can potentially be scripted. If the files are all the same, some type of distribution automation (pdsh, for example, or even scp) can be used. If only a few nodes need the change, they can be done manually.

Unique solution ID: #1234

Author: Rob Greenbank

Last update: 2022-02-01 20:17