Torque Roll: Users Guide

Version 5.1 Edition

Torque Roll: Users Guide

Version 5.1 Edition

Published Jan 2009

Table of Contents

List of Tables

Preface

The Torque Roll installs and configures the Terascale Open Resource Manager and the Maui scheduler.

Please visit the Clusterresources Inc. site[1] to learn more about their release and the individual software components.

Notes

1. http://www.clusterresources.com/

Chapter 1. Overview

Table 1-1. Summary

Name	torque
Version	5.1
Maintained By	University of Tromso
Architecture	i386, x86_64, ia64
Compatible with Rocks™	5.1

Table 1-2. Roll Compatibility

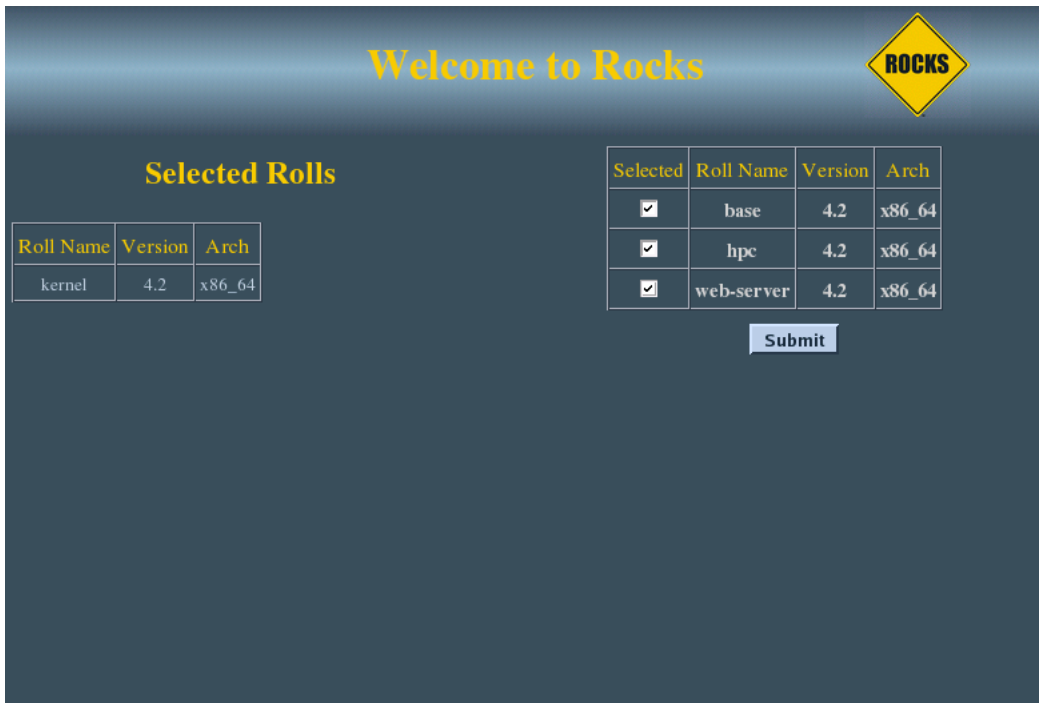
Roll	Requires [a]	Optional [b]	Conflicts
alpha		X	
area51		X	
base	X		
bio		X	
condor		X	
ganglia		X	
grid		X	
hpc	X		
java		X	
kernel	X		
os (disk 1)	X		
os (disk 2)	X		
os (disk 3)	X		
os (disk 4)	X		
os (disk 5)	X		
os (disk 6)	X		
os (disk 7)	X		
pbs		X	
service-pack		X	
sge			X
viz		X	
web-server		X	
xen		X	

Roll	Requires [a]	Optional [b]	Conflicts
Notes: a. You may also substitute your own OS CDs for the Rocks™ OS Roll CDs. In this case you must use all the CDs from your distribution and not use any of the Rocks™ OS Roll CDs. b. Only Rolls that have been verified as compatible with this Roll are listed. Other Rolls will likely work, but have not been tested by the maintainer of this Roll.			

Chapter 2. Installing

2.1. On a New Server

The torque Roll should be installed during the initial installation of your server (or cluster). This procedure is documented in section 1.2 of the Rocks™ usersguide. You should select the torque Roll from the list of available rolls when you see a screen that is similar to the one below.



2.2. On an Existing Server

The torque Roll may not be installed on an already existing server. The only supported method of installation is to install the Roll at the time of the server installation.

Chapter 3. Using

3.1. Do I need Torque?

Well, that depends. If you run a small departemental cluster where all the users can agree among themselves when and where to run their jobs, you probably don't need a queueing system at all. But, if you have too few machines and users start stepping on each others feet you definitely need one.

How does Torque compare to SGE? Don't know, I've never tried SGE, go figure out yourself what you like best. Torque (or rather PBS) and Maui has been around for a long time and has a lot of features that come in handy when you try to keep your users satisfied sustaining a reasonable job throughput.

3.2. Launching serial jobs

The easiest way to run jobs is to create a job script that contains your job setup.

Simple job script example:

```
$ cat simple-jobscript.sh
#!/bin/bash

#PBS -lwalltime=0:10:0

echo starting
sleep 10
echo ending
```

Comments beginning with `#PBS` is meaningful to the queueing system, here we ask it to reserve 10 minutes for the job, if the job runs longer than that it will be killed. You submit the job to the queue with `qsub`:

```
$ qsub simple-jobscript.sh
553.snowstorm.public
$
```

`qsub` answers with the job id it has assigned to the job, and you can use this to get information about it later:

```
$ showq
ACTIVE JOBS--------------------
JOBNAME            USERNAME      STATE  PROC   REMAINING            STARTTIME

553                   usera    Running     1    00:10:00  Fri Feb 13 10:39:25
549                   userb    Running     1     3:37:10  Fri Feb 13 09:16:35
541                   userc    Running     1     7:47:17  Thu Feb 12 12:26:42
551                   userd    Running     1     9:51:43  Fri Feb 13 10:31:08
552                   userd    Running     1     9:54:00  Fri Feb 13 10:33:25
534                   userd    Running     1  2:09:48:33  Wed Feb 11 16:27:58
547                   userb    Running     1  9:22:38:52  Thu Feb 12 23:18:17
536                   usere    Running     8 19:01:04:56  Thu Feb 12 09:44:21
```

```
     8 Active Jobs      15 of   18 Processors Active (83.33%)
                         8 of    9 Nodes Active      (88.89%)

IDLE JOBS----------------------
JOBNAME            USERNAME      STATE  PROC     WCLIMIT            QUEUETIME


0 Idle Jobs

BLOCKED JOBS----------------
JOBNAME            USERNAME      STATE  PROC     WCLIMIT            QUEUETIME


Total Jobs: 8   Active Jobs: 8   Idle Jobs: 0   Blocked Jobs: 0
```

`showq` is a maui command, you can also use the pbs command `qstat` to get similar information:

```
$ qstat
Job id           Name             User             Time Use S Queue
---------------- ---------------- ---------------- -------- - -----
534.snowstorm    TpaFeOOH-s1_opt  userd            42:10:20 R default
536.snowstorm    md7              usere            00:00:00 R default
541.snowstorm    mitgcm.run       userc            22:11:20 R default
547.snowstorm    hfdaQ            userb            11:17:10 R default
549.snowstorm    hfaQ             userb            01:22:14 R default
551.snowstorm    FePh-s1_B1       userd            00:07:46 R default
552.snowstorm    FePh-s1_A1       userd            00:04:59 R default
553.snowstorm    simple-jobscrip  usera                   0 R default
```

3.3. Launching parallel jobs

The preferred way of launching mpi jobs is to use the `mpiexec` utility, it uses the tm-interface of PBS to start the jobs, which is a lot faster than mpirun that uses ssh to start jobs on the nodes. Applications must be linked with the mpi library in `/opt/mpich/gnu/` for mpiexec to be able to start them. Here is an example:

```
$ cat parallel-jobscript.sh
#!/bin/bash

#PBS -lwalltime=0:10:0
#PBS -lnodes=4

echo starting
mpiexec some-mpi-app
echo ending
```

Here we ask for 4 nodes and mpiexec will pick up the nodelist from the queueing system and start the job on the correct nodes.

Appendix A. Rocks Copyright

Rocks(r)
www.rocksclusters.org
version 5.1 (VI)

Appendix B. Third Party Copyrights and Licenses

This section enumerates the licenses from all the third party software components of this Roll. A "best effort" attempt has been made to insure the complete and current licenses are listed. In the case of errors or ommisions please contact the maintainer of this Roll. For more information on the licenses of any components please consult with the original author(s) or see the Rocks™ CVS repository[1].

B.1. Torque

Torque, Terascale Open Resource Manager, is derived from the open source OpenPBS project and incorporates hundreds of community scalability, fault tolerance, and feature extension patches provided by NCSA, OSC, Sandia, PNNL, U.S. Dept of Energy, TeraGrid, the University of Buffalo, and other leading edge HPC organizations along side the enhancements provided by Cluster Resources, Inc. This version is based upon OpenPBS version 2.3.12 and may be freely redistributed. It not neither affiliated with nor endorsed by Altair Grid Technologies. It is currently in use at hundreds of leading government, academic, and commercial sites throughout the world. Support and other services for this product may be requested from Cluster Resources, Inc., or Altair as mentioned in the following notice.

* This product includes software developed by NASA Ames Research Center, Lawrence Livermore National Laboratory, and Veridian Information Solutions, Inc. Visit www.OpenPBS.org for OpenPBS software support, products, and information. OpenPBS is a trademark of Altair Grid Technologies/Veridian Information Solution

Notes

1. http://cvs.rocksclusters.org