

HP XC How To



Using PBS Professional™ on HP XC Clusters

Version 1.0 October, 2005

© 2005 Hewlett-Packard Development Company, L.P.

The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Linux is a U.S. registered trademark of Linus Torvalds.

Altair® is a registered trademark of Altair Grid Technologies, LLC.

PBS Professional™ is a trade mark of Altair Grid Technologies, LLC.

Contents

Introduction

Procedures

Installing PBS Professional™ under HP XC	6
Planning the installation.....	6
Performing installation actions specific to HP XC.....	6
Configuring PBS Professional™ under HP XC.....	8
Configuring the OpenSSH scp utility	8
Removing nodes from the SLURM or LSF configuration.....	8
Adding nodes to the PBS Professional™ configuration.....	8
Replicating execution nodes	9
Starting the service daemons.....	10
Setting up PBS Professional™ at the user level.....	10
Running HP MPI tasks	11

Revision history

Revision table

Table 1 Revisions

Date	Edition	Revision
Oct 2005	V1.0	First Edition

Introduction

PBS Professional™, distributed by Altair®, is an implementation of NASA's Portable Batch System (PBS). It provides a workload management solution for high-performance computing systems and Linux clusters.

For more information, and to download kits and documentation, refer to the distributor's URL:
http://www.altair.com/software/pbs_abo.htm

Procedures

The following topics are covered in this XC How To:

- Installation planning.
- Installation actions specific to HP XC.
- Configuration of PBS Professional™.
- Replicating execution nodes and starting services.
- User-level setup.
- Running HP MPI tasks

Before you install PBS Professional™, read the HP XC-specific instructions in this XC How To, and then follow the installation instructions from the *PBS Pro Quick Start Guide*

Installing PBS Professional™ under HP XC

The information in this section supplements the vendor's installation guide.

Planning the installation

Determine the type of installation that you require for each node of the cluster. Choose one of three following installation types for each node in the cluster, based on how you intend to use the node:

- Designate one node of the cluster (possibly the HP XC head node) as the PBS server. This node runs the server and scheduler daemons. Other documentation might refer to the PBS server as the **front-end node**.
- You can designate a number of nodes to run user tasks, referred to as **PBS execution nodes**. Execution nodes run the PBS Professional™ *MOM* (Machine Oriented Miniserver) daemon. On a HP XC cluster, you designate the compute nodes to function as PBS execution nodes. You can choose to use all the compute nodes, or a subset of the compute nodes. The PBS server node (front-end-node) is automatically designated as an execution node. See [Adding nodes to the PBS Professional™ configuration](#) for information.
- Optionally, you can designate a number of **PBS client nodes**. Client nodes do not run any daemons and are not used to execute user tasks. The PBS Professional™ software is installed on all client nodes, enabling these nodes to interact with the PBS server (front-end-node).

You can submit user tasks and monitor the tasks from a client node. In an HP XC cluster, there are typically one or more nodes (other than the head node) that are designated for interactive logins. Because these servers are not used as compute nodes, you can configure them as PBS clients.

Performing installation actions specific to HP XC.

Use the following installation sequence:

1. Install the PBS server node (front-end-node) first, using the installation script provided by the software vendor, and specifying the following values:

- a. You can accept the default value offered for the PBS_HOME directory, which is /var/spool/PBS).
- b. When prompted for the type of PBS installation, select: option 1 ("Server, execution and commands").
- c. If available, you may enter your license key(s) during the interactive installation. Otherwise, you can execute the script named /usr/pbs/etc/pbs_setlicense on the PBS server node after the installation is complete. (See the section titled [Replicating execution nodes](#))
- d. When the following installation script prompt is displayed:

```
Would you like to start PBS now? n
```

Enter n (no). There are additional installation and configuration steps that you must perform first.

- 2. Install one of the PBS execution nodes. Replication of this installation to the remaining execution nodes is described later in the section titled [Replicating execution nodes](#). Specify the following values:

- a. When prompted for the type of PBS installation, select: option 2 ("Execution only").
- b. You are not prompted for license information during installation of an execution node.
- c. When are prompted for the name of the PBS server node; provide an internally-resolvable name (such as n16, for the head node of a 16-node cluster), using a node-naming prefix of n.
- d. When the following installation script prompt is displayed:

```
Would you like to start PBS now?
```

Respond and proceed as follows:

- If you have finished the installation (no PBS clients to install), Enter n (no) and invoke the software from the PBS server node manually, as described in its user documentation. Proceed to the next section.
- If you are installing optional PBS client nodes, Enter n (no) and proceed to step 3.

- 3. Install any optional PBS clients, specifying the following values:

- a. When prompted for the type of PBS installation, select: option 3 ("Commands only").
- b. For PBS client installations, you are not prompted for a license key.
- c. When are prompted for the name of the PBS server node; provide an internally-resolvable name (such as n16, for the head node of a 16-node cluster), using a node-naming prefix of n.
- d. For PBS client installations only, you are not prompted to start PBS at the end of the installation. Invoke the software from the PBS server node manually, as described in its user documentation.

Configuring PBS Professional™ under HP XC

Unless specified otherwise, all the following configuration commands should be entered from the PBS server (front-end node.)

Configuring the OpenSSH scp utility

By default, PBS Professional™ uses the `rsh` utility to copy files between nodes in the cluster. The default HP XC configuration disables `rsh` in favor of the more secure `scp` command provided by OpenSSH. To use PBS on XC, configure it to default to `scp` as follows:

1. Using a text editor, open the file `/etc/pbs.conf` on the server node.
2. Search for the configuration variable `PBS_SCP`, and assign it the value `/usr/sbin/scp` as follows:

```
PBS_SCP=/usr/bin/scp
```

3. Repeat this operation on the PBS execution node, prior to performing the steps in the section titled "*Replicating execution nodes*".

Removing nodes from the SLURM or LSF configuration

Prevent SLURM or LSF from allocating jobs to PBS execution nodes as follows:

1. Remove the PBS execution nodes from all SLURM partitions specified in the file `/hptc_cluster/slurm/etc/slurm.conf`. See the *HP XC System Software Administration Guide* for details on configuring SLURM partitions.
2. Enter the following reconfiguration commands to implement the changes:

```
# scontrol reconfig
# badadmin reconfig
```

Adding nodes to the PBS Professional™ configuration

1. Create a list of nodes to manage in a file named `PBS_HOME/server_priv/nodes`, using the following syntax:

```
<node_name>[:ts] pcpus=<number_of_cpus>
```

Where:

- a. One node is specified per line.
- b. `[:ts]` - Optionally identifies the node as time sharing. (Time-shared nodes might be over-subscribed (number of jobs > number of CPUs) if the local policy permits, and are not exclusively allocated to a single job.)
- c. `<node_name>` - Specifies the node's cluster name, such as `n12`.
- d. `Pcpus` - Specifies a numerical attribute, `<number_of_cpus>` equivalent to the physical CPUs in the server.

2. The following example shows typical entries in the file `PBS_HOME/server_priv/nodes`:

```
n9 pcpus=2
n8 pcpus=2
n7 pcpus=2
n6 pcpus=2
```

3. The PBS server node is automatically configured as an execution host. To prevent jobs from running on the server node:
 - a. Do not list the server node in the file `PBS_HOME/server_priv/nodes` file
 - b. Edit the file `/etc/pbs.conf` file on the server, changing the value of variable `PBS_START_MOM` to 0 (zero).

Replicating execution nodes

During installation, you configured one PBS execution node. You can now replicate the PBS execution node installation to the remaining execution nodes by running the following commands from the already-installed PBS execution node:

```
# pdcp -rp -w "x[n-n]" /usr/pbs /usr
# pdcp -rp -w "x[n-n]" /var/spool/PBS /var/spool
# pdcp -p -w "x[n-n]" /etc/pbs.conf /etc
# pdcp -p -w. "x[n-n]" /etc/init.d/pbs /etc/init.d
```

Where `"x[n-n]"` is a nodelist expression representing all execution nodes excluding the single execution node that is configured already

For example:

- You have installed the PBS server on node `n100`.
- The first PBS execution node is node `n49`.
- You want to replicate the execution environment to nodes `n1` through `n48`.

In this case, the value of the nodelist expression is: `"n[1-48]"`. Double quotes (") are required surrounding the expression so that the square brackets ([]) are parsed correctly by the shell. See the *HP XC System Software Administration Guide* for more information about nodelist expressions.

If you did not enter your license information when installing the PBS server, do so now by running the following command on the PBS server node:

```
# /usr/pbs/etc/pbs_setlicense
```

Starting the service daemons.

Enter the following command to start the server, scheduler, and MOM daemons:

```
# pdsh -w "x[n-n, N]" service pbs start
```

Where the nodelist "`x[n-n, N]`" specifies the range of execution nodes (`n-n`), and also the PBS server node (`N`). For example, a valid nodelist is "`n[1-49, 100]`". As for previous nodelists, the double quotation marks are required.

Enter the following command to cause PBS Professional to start automatically at boot time:

```
# pdsh -w "x[n-n, N]" chkconfig --level 345 pbs on
```

Use the same value for nodelist as specified in the previous `pdsh` command for starting the service daemons.

The preceding commands complete the minimum installation and configuration of PBS Professional on an HP XC cluster. See the *PBS Professional Administrator Guide* for information on other configuration options.

Setting up PBS Professional™ at the user level

There is some overlap in the name space of PBS Professional commands and other HP XC-provided software. Users of PBS Professional must set their `PATH` and `MANPATH` variables so that the PBS Professional versions of commands and man pages are invoked.

Users of the `csh` or `tcsh` shells must append the following the commands to an appropriate login shell script:

```
% setenv PATH /usr/pbs/bin:${PATH}
% setenv MANPATH /usr/pbs/man:${MANPATH}
```

Users of the `sh/ksh/bash` shell must append the following the commands to an appropriate login shell script:

```
% PATH=/usr/pbs/bin:${PATH} ; export PATH
% MANPATH=/usr/pbs/man:${MANPATH} ; export MANPATH
```

Users must configure OpenSSH to enable a login without a password. Use the following commands, (described in the *HP XC System Software User's Guide*), pressing the Enter key in response to all system prompts to accept the default values:

```
$ ssh-keygen -t dsa
$ cd ~/.ssh
```

```
$ cat id_dsa.pub >>authorized_keys
$ chmod go-rwx authorized_keys
```

After all above steps are complete, users are able to submit jobs as in a standard PBS Professional installation.

Running HP MPI tasks

The PBS Professional distribution contains a wrapper script named `pbs_mpihp` that is used for running HP MPI jobs. The wrapper script uses information about the current PBS Professional allocation to construct a command line and optionally, an `appfile` suitable for HP MPI. The wrapper also sets the `MPI_REMSH` environment variable to the PBS Professional remote shell utility named `pbs_tmrsh`. This remote shell allows PBS Professional to keep track of communication between MPI processes on different nodes.

To use the wrapper, replace the standard `mpirun` script, calling the `pbs_mpihp` wrapper instead in HP MPI job launch scripts. Two sample launch scripts are presented in the following examples:

```
% cat launch
#!/bin/sh
#PBS -l select=4:ncpus=2
/usr/pbs/bin/pbs_mpihp -np 8 <other MPI options> ./a.out
```

```
% cat launch
#!/bin/sh
#PBS -l select=4:ncpus=2
/usr/pbs/bin/pbs_mpihp -np 8 <other MPI options> -f appfile
```

The PBS Professional documentation for `pbs_mpihp` recommends replacing the HP MPI `mpirun` command with a symlink to `pbs_mpihp` in order to make the presence of PBS Professional completely transparent to the users of HP MPI. On clusters where PBS Pro is the only active queuing system, this transparency might be desirable. However, administrators of mixed SLURM and PBS Professional clusters must first determine if this configuration is appropriate for their needs. For more information, consult the `pbs_mpihp` reference (man) page, and the *PBS Professional Administrator's Guide*.