Quick Start: Deploying the Moab Grid Suite on HP Clusters

1 Overview

Beyond clusters, grids of computer systems dispersed over a wide area network can provide an effective way to harness an organization's computing resources. The Moab Workload Manager® for Grids, part of the Moab Grid Suite from Cluster Resources [1], is a powerful and effective tool for coordinating and managing grids.

This document uses the following conventions:

Convention	Description
Document title	The title of a document. On the web and on the Instant Information CD, it may be a hot link to the document itself.
`UserInput`	Commands and other text that the user types.
`$`	User command prompt.
`#`	Superuser (`root`) command prompt.

2 The Cluster Resources Moab Grid Suite

The Moab Grid Suite products produced by Cluster Resources include the Moab Workload Manager (MWM), a meta-scheduler and resource manager; Moab Access Portal (MAP), a Web-based browser interface for submitting and managing user jobs; and Moab Grid Manager (MGM), a graphical interface for configuring a grid using Moab Workload Manager. The following sections describe these products in more detail.

2.1 Moab Workload Manager

The Moab Workload Manager (WMW) acts as a meta-scheduler on your system, working with any of a number of possible resource managers. See the Moab Workload Manager Administrator's Guide [2] for more information.

MWM relieves the local resource manager of the responsibility for scheduling resources, using the resource manager only to start, suspend, resume, and stop jobs, as needed.

MWM provides the following features:

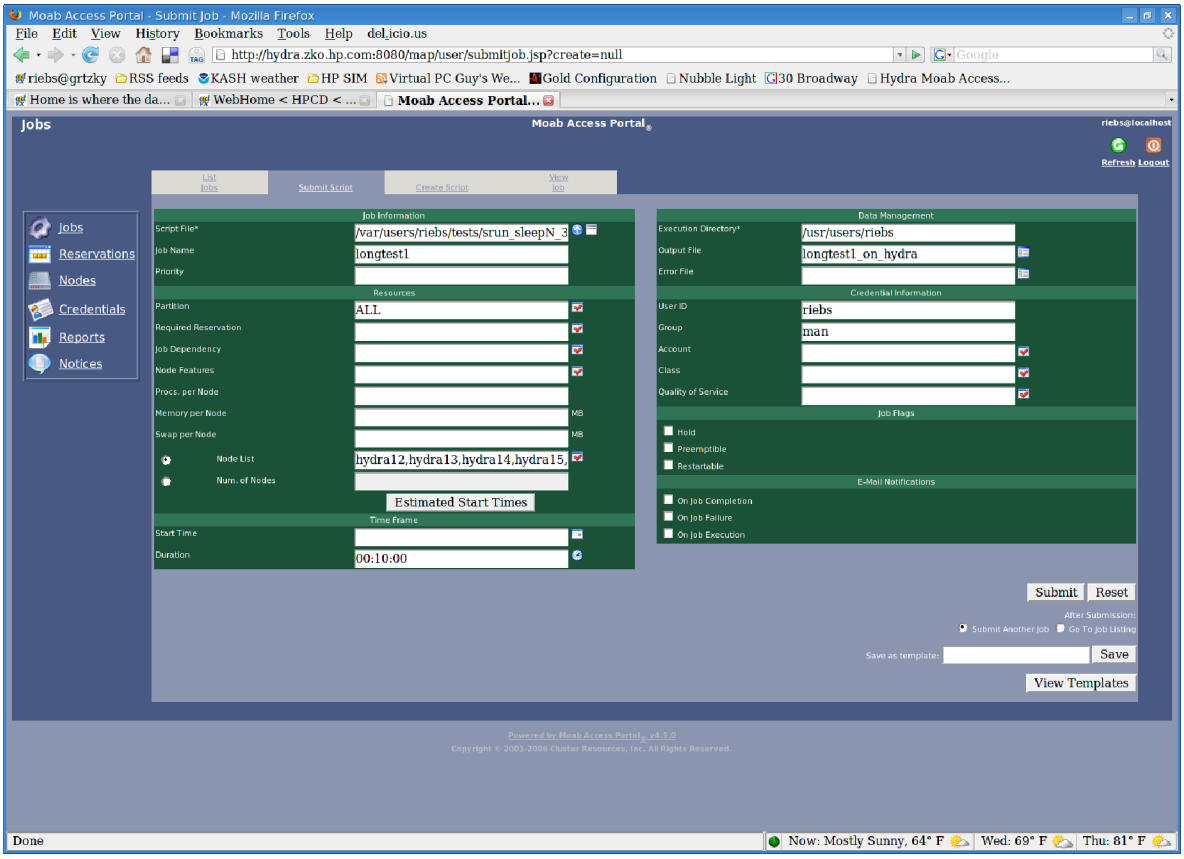
- A rich set of scheduling commands and parameters, allowing very fine-grained control of scheduling and workload management decisions.
- The ability to emulate some of the user commands of other resource managers, making it easier to migrate users to Moab.
- A choice of Master-Slave and Peer-Peer relationships between systems, allowing remote management of the workload or autonomous local control.
- The ability to integrate systems and clusters running a wide variety of software from a variety of vendors into a coherent grid, minimizing the need to change software or retrain users on the member systems.
- A view of the grid for the user that is very similar to the view of the local system, making it easy to expand the grid and make it available to users.

The HP Grid Catalyst Team carried out an extensive set of tests on MWM's commands and parameters in a peer-to-peer configuration [3].

2.2 Moab Access Portal for Grids

The Moab Access Portal (MAP) is a Web-based graphical user interface (GUI) for user job submission and control. MAP allows the user to run pre-defined scripts, or create and submit new scripts. See the Moab Workload Manager Administrator's Guide [2] and Moab Access Portal Administrator's Guide [4] for more information.

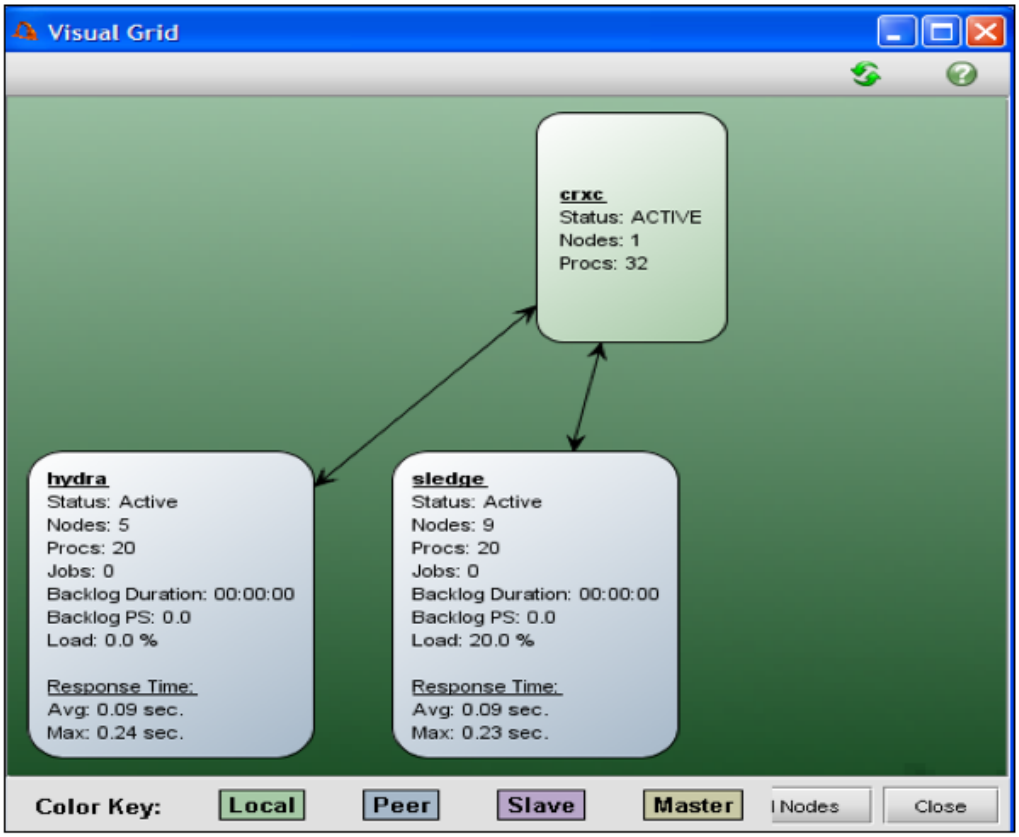
Figure 1. The Moab Access Portal (MAP).



2.3 Moab Grid Manager

The Moab Grid Manager (MGM) provides a graphical user interface to the Moab Workload Manager, simplifying the creation and management of grids. See the Moab Workload Manager Administrator's Guide [2] for more information.

Figure 2. The Moab Grid Manager (MGM).



3 Interoperating with Various Resource Managers

HP Cluster systems are often configured to use the open source SLURM package as their resource manager, so we will discuss that configuration here. Moab Workload Manager can work with a number of other resource managers; see [2] for more information.

4 Setting Up Moab Workload Manager

This section assumes that you have SLURM installed as your resource manager. This section will guide you through the process of installing MWM and configuring SLURM to work with MWM.

4.1 Prerequisite Software Versions

For the following configuration you will need:

- MWM version 5.1.0-p7 or later.
- SLURM version 1.2.4 or later.

4.2 Installation Notes

HP recommends that you:

- Install MWM in `/opt/moab`.
- Install MWM programs in `/usr/local/bin and /usr/local/sbin`.
- Install MWM emulation tools in `/usr/local/tools`, with symbolic links from `/usr/local/bin`.

- Use MWM version 5.1.0-p7 or later.

The rest of this document will guide you through a peer-to-peer Moab grid installation that implements these recommendations.

For more detailed information, see the Moab Workload Administrator's Guide [5] and [6].

4.3 Installation

To begin the installation, acquire the MWM kit and a license file from Cluster Resources [1] (the kit will have a name such as `moab-5.1.0-x86_64-slurm-p7.tar.gz`), and store them temporarily in `/tmp` on the head node of your system.

1. Log in as root:

   ```
   $ su -
   Password:
   #
   ```

2. Create the base directory for MWM:

   ```
   # mkdir -p /opt/moab
   ```

3. Change your current directory to `/opt/moab`:

   ```
   # cd /opt/moab
   ```

4. Unpack the Moab kit:

   ```
   # tar zxf /tmp/moab-*.tar.gz
   ```

5. Move to the newly-created MWM kit directory:

   ```
   # cd moab-*
   ```

6. Configure MWM to use `/opt/moab` as its base directory:

   ```
   # ./configure --homedir=/opt/moab
   ```

7. Build and install your new configuration:

   ```
   # make
   # make install
   ```

8. Copy your MWM license to `/opt/moab`:

   ```
   # cp /tmp/moab.lic /opt/moab
   ```

4.4 Configuration

Note: This discussion assumes that the SLURM resource manager is already installed and configured on your system.

4.4.1 Configure SLURM to work with MWM

The SLURM configuration file resides in `/hptc_cluster/slurm/etc` on an HP XC system; otherwise, it probably resides in `/etc` on your system. Use your favorite editor to modify `slurm.conf` as follows:

1. Comment out any existing values for `SchedulerType` or `SchedulerPort` by prefixing them with a # character, such as:

   ```
   # SchedulerType=builtin
   # SchedulerPort=8132
   ```

2. Add new values for `SchedulerType` or `SchedulerPort` as follows:

   ```
   SchedulerType=sched/wiki2
   SchedulerPort=7321
   ```

3. Modify partition definitions, as necessary, to specify `RootOnly=NO` and `Shared=YES`. For example, change:

```
PartitionName=part2 RootOnly=YES Shared=FORCE Nodes=sled[128-250]
```

to:

```
PartitionName=part2 RootOnly=NO Shared=YES Nodes=sled[128-250]
```

4. Note: on an HP XC system, ensure that no partition is named "lsf" if you will be using MWM directly with SLURM.

In the same directory as `slurm.conf`, create a file named `wiki.conf` with the secret key for this system. This example uses `1234567890` as the authorization key:

```
# touch wiki.conf
# chmod 700 wiki.conf
# echo "AuthKey=1234567890" >>wiki.conf
```

Now restart SLURM:

```
# /etc/init.d/slurm restart
```

4.4.2 Configure Moab Workload Manager

Create or modify `/opt/moab/moab.cfg` to include parameters such as the following:

```
SCHEDCFG[tag]      MODE=normal SERVER=cluster:42559
ADMINCFG[1]        USERS=root
RMCFG[tag]         TYPE=WIKI:SLURM AUTHTYPE=CHECKSUM
```

The following table describes these parameters:

Where...	Is...
Tag	Used to match parameters to the same object. In this example, "tag" is used to specify that the RMCFG and SCHEDCFG parameters belong to the same system.
cluster	The IP name or address to which MWM should address requests for the workload manager; it is typically the external Ethernet address of the host on which MWM has been installed.

Now create or modify `/opt/moab/moab-private.cfg` as follows:

```
# touch /opt/moab/moab-private.cfg
# chmod 700 /opt/moab/moab-private.cfg
# echo "CLIENTCFG[RM:tag] KEY=1234567890" >> /opt/moab/moab-private.cfg
```

In the last line ***tag*** is the same tag used in `moab.cfg`. If you are working directly with SLURM, the KEY value must be the same value that you used in `wiki.conf`.

4.4.3. Useful Moab Workload Manager Commands

The following table lists and describes some commonly used Moab Workload Manager (MWM) commands.

Command	Description
Moab	Starts MWM.
mschedctl -R	Restarts MWM, possibly with new configuration parameters.
mschedctl -k	Stops MWM.
showq -v	Displays the current queues and shows where migrated jobs have landed.
mdiag -n	Lists all nodes on the grid.
mdiag -R	Lists all clusters on the grid.

5 Setting up Moab Access Portal for Grids

The following sections describe how to set up the Moab Access Portal (MAP) for grids.

5.1 Installation

Keep the following installation notes in mind:

- By default, MAP does not encrypt or otherwise protect user names and passwords when they are solicited from the user through a browser; see the Moab Access Portal Administrator's Guide [7] if you plan to use MAP in environments where network security is a concern.
- MAP is typically accessed from http://your-server:8080/map .
- MAP does not need to be installed on the same system as MWM, but was for this document.
- The MAP server executes from the directory where you install it.

To begin the installation, acquire the MAP kit from the Cluster Resources Web site [1] and store it in `/tmp` on the head node of your system.

1. Log in as root:

```
$ su -
Password:
#
```

2. Select a location to install MAP, such as `/opt/moab`:

   ```
   # cd /opt/moab
   ```

3. Unpack the MAP kit:

   ```
   # tar zxf /tmp/map-*.tar.gz
   ```

5.2 Configuration

To move into the MAP directory, enter:

```
# cd /opt/moab/map-*
```

Configure MAP for your system. In this case, accept the defaults to run on the local machine using port 22 for ssh access:

```
# ./configure

Moab Access Portal(R) Automated Configuration
-----------------------------------------

This configure script will now setup Moab Access Portal (MAP) and its related dependencies.
This version of MAP runs fully encapsulated in its extracted directory.
No files will be installed outside of this directory.

Hit ENTER to proceed with the configuration...

Creating configuration files:
Found current '/opt/moab/map-4.5.0/etc/log4j.properties' - leaving alone.
Found current '/opt/moab/map-4.5.0/etc/map.properties' - leaving alone.
Found current '/opt/moab/map-4.5.0/etc/nav.properties' - leaving alone.
Uninstalling existing map.war...Done

Extracting 'map.war' to './tomcat/webapps/'...Done

Enter the address or hostname of the head node Moab Access Portal will communicate with:
(For the default, 'localhost', simply hit ENTER) hydra.zko.hp.com

Enter the port used by SSH to access the head node 'hydra.zko.hp.com':
(For the default, '22', simply hit ENTER)

To activate Moab Access Portal use the newly created start-up script: 'mapctl.sh'

Setup is complete.


================================================================================
REMEMBER: The full path to your map.properties file is '/opt/moab/map-
4.5.0/etc/map.properties'!
This value will be passed into the servlet engine via the CONFIG-FILE system property so MAP
can locate it.

#
```

5.3 Running the MAP Server

The MAP server must be run from the directory in which you installed it, so start by logging in as root:

```
$ su -
Password:
#
```

Now switch to the MAP directory and start it up:

```
# cd /opt/moab/map-*
# ./mapctl.sh start
```

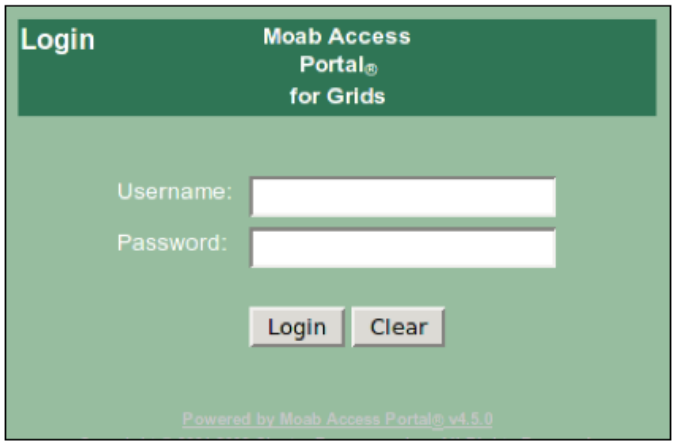
5.5 Using MAP

From any system on your network, browse to the MAP server; for example:

```
$ mozilla-firefox \
 http://your-server:8080/map
```

You will then be presented with a login screen; simply log in and follow the directions on the screen.

Figure 3. MAP server login screen.



6 Setting up the Moab Grid Manager

The following sections describe how to set up the Moab Grid Manager (MGM).

6.1 Installation

Keep the following installation notes in mind:

- MGM does not need to be installed on a system that is running MWM; HP recommends installing it on your personal desktop system.
- Try to avoid running MGM from a remote X display; the performance can be disappointing.

To begin the installation, acquire the MGM kit from Cluster Resources Web site [1] and store it temporarily in `/tmp`.

1. Create a convenient directory:

   ```
   $ mkdir ~/mgm
   ```

2. Change your current directory to the mgm directory:

```
$ cd ~/mgm
```

3. Unpack the MGM kit:

```
$ tar zxf /tmp/mgm-*.tar.gz
```

You have now completed the installation of MGM.

6.2 Configuration

No configuration is necessary for MGM.

6.3 Using MGM

To switch to the MGM directory, enter:

```
$ cd ~/mgm/mgm-*
```

To run MGM, enter:

```
# ./mgm
```

7 Integrating this System with Other Moab Systems

In this example, MWM is set up on a SLURM-based cluster named hydra that is then joined as a peer in a grid to a cluster named sledge. The files on sledge have been set up as follows:

```
# moab.cfg on sledge

SCHEDCFG[sledge]  MODE=NORMAL SERVER=sledge16:42559
ADMINCFG[1]       USERS=root

RMCFG[sledge]     TYPE=WIKI:SLURM AUTHTYPE=CHECKSUM
RMCFG[hydra]      TYPE=MOAB SERVER=hydra:42559


# moab-private.cfg on sledge

CLIENTCFG[RM:sledge]    KEY=19234813
CLIENTCFG[RM:hydra]     KEY=92348131 AUTH=admin1
```

The following log shows the steps necessary to run MWM on the head node of the hydra cluster and join it in a grid with the sledge cluster. MWM is first set up to run only on hydra to ensure that the local cluster is configured correctly, and then it is connected to sledge.

```
$ su -
Password:
# vi /hptc_cluster/slurm/etc/wiki.conf
# cat /hptc_cluster/slurm/etc/wiki.conf
AuthKey=19234813
# cd /opt/moab
# cat moab.cfg
SCHEDCFG[hydra]         MODE=normal SERVER=hydra16:42559
ADMINCFG[1]             USERS=root
RMCFG[hydra]            TYPE=WIKI:SLURM AUTHTYPE=CHECKSUM

# cat moab-private.cfg
CLIENTCFG[RM:hydra]       KEY=19234813

# moab
Moab Workload Manager Version '5.1.0p5' License Information:
  Current License:  Max Procs   = 2400
```

```
  Current License:  Valid Until - Mon May 18 21:00:58 2009
# mdiag -R
diagnosing resource managers

RM[hydra]  State: Active
  Type:              WIKI:SLURM
  Server:            localhost:7321

# mdiag -n
compute node summary
Name                    State   Procs      Memory         Opsys

hydra12                  Idle    4:4      2007:2007           -
hydra13                  Idle    4:4      2007:2007           -
hydra14                  Idle    4:4      2007:2007           -
hydra15                  Idle    4:4      2007:2007           -
hydra16                  Idle    4:4      2008:2008           -
-----                     ---   20:20    10036:10036      -----

Total Nodes: 5  (Active: 0  Idle: 5  Down: 0)

# mschedctl -k

moab will be shutdown immediately

# vi moab.cfg
# cat moab.cfg
SCHEDCFG[hydra]         MODE=normal SERVER=hydra16:42559
ADMINCFG[1]             USERS=root
RMCFG[hydra]            TYPE=WIKI:SLURM AUTHTYPE=CHECKSUM
RMCFG[sledge]           TYPE=MOAB SERVER=sledge-nh:42559
# vi moab-private.cfg
# cat moab-private.cfg
CLIENTCFG[RM:hydra]  KEY=19234813
CLIENTCFG[RM:sledge] KEY=92348131 AUTH=admin1
# moab
Moab Workload Manager Version '5.1.0p5' License Information:
  Current License:  Max Procs   = 2400
  Current License:  Valid Until - Mon May 18 21:00:58 2009
# showq -v

active jobs------------------------
JOBID              USERNAME      STATE PROCS   REMAINING            STARTTIME


0 active jobs              0 of 36 processors in use by local jobs (0.00%)
                           0 of 12 nodes active      (0.00%)

eligible jobs----------------------
JOBID              USERNAME      STATE PROCS     WCLIMIT            QUEUETIME

0 eligible jobs

blocked jobs-----------------------
JOBID              USERNAME      STATE PROCS     WCLIMIT            QUEUETIME

0 blocked jobs

Total jobs:  0

# mdiag -R
diagnosing resource managers

RM[hydra]  State: Active
  Type:              WIKI:SLURM
  Server:            localhost:7321
RM[sledge]  State: Active
  Type:              MOAB
  Server:            sledge-nh:42559
```

```
RM[sledge.INBOUND]  State: Active
  Type:               MOAB
# mdiag -n
compute node summary
Name                    State   Procs      Memory         Opsys

hydra12                  Idle    4:4      2007:2007           -
hydra13                  Idle    4:4      2007:2007           -
hydra14                  Idle    4:4      2007:2007           -
hydra15                  Idle    4:4      2007:2007           -
hydra16                  Idle    4:4      2008:2008           -
nh                       Idle    4:4      3072:3072           -
sledge10                 Idle    2:2      3072:3072           -
sledge11                 Down    0:2      3072:3072           -
sledge12                 Idle    2:2      3072:3072           -
sledge13                 Idle    2:2      3072:3072           -
sledge14                 Idle    2:2      3072:3072           -
sledge15                 Down    0:2      3072:3072           -
sledge8                  Idle    2:2      3072:3072           -
sledge9                  Idle    2:2      3072:3072           -
-----                     ---   36:40    37684:37684      -----

Total Nodes: 14  (Active: 0  Idle: 12  Down: 2)
#
```

8 Hints and Known Issues

This section provides a few helpful hints to ease the adoption of the Moab grid suite into your environment.

8.1 Firewalls

By default, the Moab grid suite components use the following ports:

Port	Usage
8080	To access the Moab Access Portal via the Web browser
42559	Used for system-to-system communication between Moab Workload Manager instances.

If you have a firewall enabled on the system that runs MWM, you will need to allow traffic through these ports. For example, on an HP XC system, the following commands will permanently open ports 8080 and 42559 for use by MAP and MWM, respectively:

```
# openipport --port 8080 --protocol tcp --interface External
# openipport --port 42559 --protocol tcp --interface External
```

8.2 Authenticating Users in Different Domains

You can use the OMAP (object map) parameter of the RMCFG directive to specify a file that remaps user names, groups, and other parameters from one domain to another. For example, to remap the group `tech` from the `exper` system to `users` on the local system, follow these steps:

1. Create `/opt/moab/exper_groups.cfg`, as follows:

   ```
   # Remap tech group from exper to group "users" here
   group:users,tech
   ```

2. Add the following line to `/opt/moab/moab.cfg`:

   ```
   RMCFG[exper] OMAP=file:///opt/moab/exper_groups.cfg
   ```

See the Moab Workload Manager Administrator's Guide [2] for more information.

8.3 Starting and Stopping the Moab Grid Suite Components

You must manually start the server components of the Moab grid suite, MWM and MAP, when your system is rebooted. You may find it convenient to create an `init.d` script appropriate for your choice of operating system.

References

[1] The Cluster Resources home page, http://www.clusterresources.com/

[2] Moab Workload Manager Administrator's Guide, http://www.clusterresources.com/products/mwm/docs/index.shtml

[3] HPCD Grid Catalyst Team, In-depth Testing of Cluster Resources' Moab 5.1, www.hp.com/go/collaboration

[4] Moab Access Portal Administrator's Guide, http://www.clusterresources.com/products/map/docs/

[5] Moab Installation Guide for Beginners, http://www.clusterresources.com/products/mwm/docs/moabinstall.shtml

[6] Moab Workload Administrator's Guide, Section 2.1, Moab Installation, http://www.clusterresources.com/products/mwm/docs/2.1installation.shtml

[7] Moab Access Portal Administrator's Guide, Section 1.6.2, Enabling a SSL Connection, http://www.clusterresources.com/products/map/docs/1.6tomcatinstall.shtml#ssl

For More Information

For questions regarding the Moab Grid Suite products, send email to support@clusterresources.com or refer to the Cluster Resources website: http://www.clusterresources.com .

For questions regarding installing Moab Grid Suite on HP Clusters, send email to pdlhpcdgrid@hp.com or refer to the section about Grid and Adaptive Infrastructure on the website: www.hp.com/go/collaboration

4AA1-6311ENW, November 2007