LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-SM-228870

# Moab Quick-Start User Guide

**Donald A. Lipari**

September 29, 2008

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory in part under Contract W-7405-Eng-48 and in part under Contract DE-AC52-07NA27344.

# User Interfaces Available

Users running jobs on Moab-scheduled machines have two main options available. You can use the native set of Moab commands or one of six, familiar LCRM commands. The LCRM commands are actually wrapper scripts that behave like LCRM even though they are interacting with the Moab scheduler. The six LCRM wrappers are *psub*, *pstat*, *palter*, *phold*, *prel*, and *prm*.

There is actually a third set of commands that are available to LC users - those provided by the underlying resource manager, SLURM. While all of SLURM's status commands are available, interactive *srun* submissions are only allowed into the pdebug partition.

Users who wish to submit legacy LCRM job command scripts containing the #PSUB directives must submit their scripts using the *psub* wrapper. Moab's native *msub* command does not understand #PSUB directives.

The following table presents a comparison of all three command sets.

| | **LCRM Emulation** | **Moab** | **SLURM** |
|---|---|---|---|
| Submit a one node job | psub -ln 1 job.cmd | msub -l nodes=1 job.cmd | srun -N1 job.cmd |
| Status the job | pstat -f <jobID> | checkjob <jobID> | scontrol show job <jobID> |
| Show the queue | pstat -A | showq | squeue |
| Remove a job | prm <jobID> | mjobctl -c <jobID> | scancel <jobID> |
| | | | |
| Submit a job to start at a particular time | psub -A <time> | msub -a [[[[CC]YY]MM]DD]hhmm[.SS] | srun --begin=time |
| Submit a job with a wall clock time limit | psub -tW <time> | msub -l walltime= [[DD:]HH:]MM:]SS | srun --time=minutes |
| Submit a job to draw from a specific bank | psub -b <bank> | msub -A <bank> | srun --account= <bank> |
| Submit a job for a non-default SLURM partition | psub -pool <name> | msub -q <name> | srun --partition=<name> |

For a more complete comparison of psub and msub, see https://computing.llnl.gov/jobs/moab/psubOptions.pdf

### Job Submission Example

Submit an 8 node job to run at 4:00pm for 2 hours in the default node partition (pbatch):

```
msub -a 1600 -l nodes=8,walltime=2:00:00 moabJob.cmd
psub -A 4pm -ln 8 -tW 2:00 lcrmJob.cmd
```

**Job Queue Status**

```
showq
mdiag -j -v
squeue
pstat -A
mjstat
```

**Queued Jobs Showing Priority Factors**

```
mdiag -p
```

**When Will a Job Run**

```
showstart <jobID>
```

**When Would a Job of size@duration Run**

```
showstart <procs@duration>
```

**Status of a Single Job**

```
checkjob -v <jobID>
scontrol show job <jobID>
mdiag -j <jobID>
pstat -f  <jobID>
```

**Status of Past Jobs**

```
showq -c
pstat -T
```

**View User Account (bank) Membership**

```
mshare
mdiag -u <user>
```

- ALIST contains account membership
- ADEF indicates default account

**Hold a Job**

```
mjobctl -h user <jobID>
phold -n <jobID>
```

**Release a Job**

```
mjobctl -u user <jobID>
prel -n <jobID>
```

**Remove a Job**

```
mjobctl -c <jobID>
canceljob <jobID>
```
(this command works even though it is deprecated)

```
prm <jobID>
```

## Job State Comparison

| Job State | LCRM | Moab | SLURM |
|---|---|---|---|
| Resource Manager Communication Problem | WHOST | NotQueued | |
| Waiting in Queue | ELIG, WCPU, WPRIO | Idle | PENDING |
| Held | | Hold | |
| Held | | BatchHold | |
| Held | HELDs | SystemHold | |
| Held | HELDu | UserHold | |
| Staging | BAT_WAIT, STAGING | Staging | |
| Starting | RUN | Starting | |
| Running | RUN | Running | RUNNING |
| Suspended | | Suspended | SUSPENDED |
| Deferred | DEFERRED | Deferred | |
| Depends on another job | DEPEND | | |
| Completed | COMPLETE | Completed | COMPLETED |
| Job reached its wall-clock limit | REMOVED | Removed | TIMEOUT |
| Cancelled | REMOVED | Removed | CANCELLED |
| Job Failed (non-zero exit code) | COMPLETE | Completed | FAILED |
| Node Failed | REMOVED | Vacated | NODE_FAIL |

## Terminology Translation Table

| SLURM | LCRM | Moab | (Simplified) Definition |
|---|---|---|---|
| node | node | node or host | One or more CPUs on single motherboard |
| partition | pool | class or queue | A subset of nodes on a host |
| host | host | host | One or more nodes comprising a single system |
| | partition | partition | One or more hosts managed as a single entity (At LLNL, we have configured each Moab partition with only one host) |
| | bank | account | The project to charge for resource usage |
| user | user | user | User |
| group | | group | Unix group |
| | job class | Quality of Service (QoS) | normal, standby, and expedite |
| interactive job | interactive job | non-msub job | a job submitted directly to SLURM via srun |
| | | interactive job | a job submitted to Moab via msub that provides a shell prompt when the job runs |

## Environment Variables

All of the standard SLURM environment variables (see ENVIRONMENT VARIABLES section of srun man page) will be present for Moab-scheduled jobs. In addition, when the psub wrapper is used to submit jobs to Moab, the following environment variables will be present:

| Variables Defined When Job Is Submitted with psub | Description | Variables Defined When Job Is Submitted with msub |
|---|---|---|
| ENVIRONMENT | Set to BATCH for batch jobs | ENVIRONMENT |
| PSUB_DEP_JOBID | Dependent Job ID | SLURM_DEPENDENCY |
| PSUB_HOME | Set to $HOME on the submission host | |
| PSUB_HOST | Submission node name | |
| PSUB_JOBID | Job ID | SLURM_JOB_ID |
| PSUB_LOGNAME | Set to $LOGNAME on the submission host | |
| PSUB_PATH | set to $PATH on the submission host | |
| PSUB_REQNAME | Job name (psub -r / msub -N value) | SLURM_JOB_NAME |
| PSUB_SHELL | Job's shell interpreter | |
| PSUB_SUBDIR | Directory from which you invoke psub on the submitting host | |
| PSUB_TZ_ENV | set to $TZ on the submission host when defined | |
| PSUB_USER | set to $USER on the submission host | |
| PSUB_WORKDIR | Home directory on the execution host | PSUB_WORKDIR*<br>QSUB_WORKDIR* |
| SESSARGS | Arguments to psub | |

*These variables are inserted by Moab.

The psub wrapper is not present at LANL and SNL. Users are advised to adopt the native SLURM variables in place of the old PSUB_ variables.

No deprecated PCS_ environment variables are supported (see User Enhancement Item 15 in the LCRM 6.14.3 Release Notes https://lc.llnl.gov/computing/techbulletins/bulletin401.pdf).

## Notes

1. All Moab commands and LCRM wrappers are located in /usr/bin on Linux machines and /opt/freeware/bin on AIX.
2. Most Moab commands offer a "-v" option that provides more verbose output.
3. Descriptions of all Moab commands are found in: http://www.clusterresources.com/products/mwm/moabdocs/a.gcommandoverview.shtml. The recommended commands are listed in the first table. The second table contains deprecated commands. LC provides man pages for native Moab commands.
4. Time in Moab is represented in HH:MM:SS format while in LCRM it is only HH:MM.
5. The *pstat* wrapper does not accept combined options. For example, "pstat -fT" needs to be entered as "pstat -f -T".
6. Moab's job priority ranges from -1,000,000,000 to 1,000,000,000. Rather than attempt to translate this in *pstat* and map priority values into LCRM's zero to one floating point representation, we decided to reflect Moab's job priority in the *pstat* output. The higher the number, the sooner your job will run. The *psub* wrapper's -p priority option is not supported.
7. *pstat*'s EXEHOST output will show ALL initially. This again reflects Moab's architecture where it initially considers your job for all machines. As your job is committed to run on, say atlas, this display will update to indicate "atlas."
8. Moab will sometimes defer a job if nodes are down. The job is still eligible. As that job bubbles to the top of the queue, Moab will run the job. The *pstat* wrapper will translate a deferred state to indicate ELIG.
9. There are three new scripts to facilitate the transition from LCRM to Moab
    a. *mjstat* replicates the output of the *spjstat* command.
    b. *lcrm2moab* converts a legacy LCRM job command script to a Moab job command script.
    c. *mshare* provides a simplified display of the user's account memberships and the assigned shares of those accounts.

    These commands will be found in /usr/bin (or /opt/freeware/bin).
10. There are no Moab analogs to LCRM's *plim* and *brlim* commands. Instead, you will find job and host limits by invoking *mdiag -t -v* or reading *news job.lim.<machine>*.
11. Jobs submitted to Moab should be cancelled by the Moab *mjobctl -c* command or by the *prm* wrapper. Do not use SLURM's *scancel* to cancel Moab jobs.
12. Information on alerting a job when its time is about to expire can be found on the Moab Web site, https://computing.llnl.gov/jobs/moab/GetRemainingTime.pdf
13. More Moab information can be found on the Moab Web sites: https://computing.llnl.gov/jobs/moab/ https://computing-int.llnl.gov/moab/MoabIssues.pdf