

Node Partitions

Nodes on Livermore Computing (LC) production clusters are typically segregated into login nodes and two node partitions (a.k.a. node pools) under the control of a resource manager—a batch partition and an interactive partition. The batch partition is generally named “pbatch,” and the interactive partition is usually named “pdebug.” (The interactive partition is named “viz” on Purple.) A generic node configuration is shown in the figure below.



These node partitions serve different purposes. The batch partitions are dedicated to running production jobs, so the node size limits are higher and the run-time limits are longer than those of the interactive partitions. The interactive partitions are intended for short debugging sessions, so the run time and node size limits are smaller in order to foster quick turnarounds and much shorter queue wait times. There is a news item on every machine that summarizes the node and time limits for each partition. Read `news job.lim.<cluster_name>`.

The batch system for LC clusters has two layers. Each cluster has its own resource manager (SLURM) that is responsible for launching jobs on that cluster. The workload manager sits on a layer above SLURM and connects all the clusters into a single grid. There are two workload managers currently in use, Moab and LCRM. Moab will eventually become the sole workload manager as LCRM is retired.

To run jobs in the batch partitions, users must submit their jobs to the workload manager (either Moab or LCRM). To run in the interactive partitions, users have the option of submitting jobs to Moab (`msub -q pdebug jobscript.cmd`) or LCRM (`psub -pool pdebug jobscript.cmd`) or directly to SLURM (`srun`, `sbatch`, or `salloc`) or `poe` on AIX machines. See the man pages for `srun`, `sbatch`, and `salloc` for more info.

We recommend that users running jobs in the interactive partitions submit their jobs directly to SLURM. These jobs will run sooner because they will run without Moab (or LCRM) intervention. Many more jobs can be run per hour when submitted directly to SLURM rather than through Moab. Jobs submitted to Moab or LCRM requesting an interactive partition will actually be at a disadvantage. SLURM will run jobs submitted by `srun`, `sbatch`, or `salloc` first before accepting jobs submitted by `msub` or `psub`.

SLURM-submitted job IDs begin at 1,000,001 and go up. Moab-submitted job IDs range from 1 to 999,999. SLURM-submitted jobs are invisible to Moab and do not appear in Moab’s `showq` output. All jobs will appear in SLURM’s `squeue` output.

SLURM-submitted jobs are placed in a simple FIFO queue. There is no fair-share scheduling policy imposed on jobs in queues for the interactive partitions.