A Quick Tutorial on Slurm

General commands

PBS Command	Slurm Command	Meaning
`qsub <job-file>`	`sbatch <job-file>`	Submit <job script> to the queue
`qsub -I`	`salloc <options>`	Request interactive job
`showstart`	`squeue --start`	Show estimated start time
`qstat <-u username>`	`squeue <-lu username>` `-l: long report`	Check jobs for a particular user in the scheduling queue
`qstat <queue>`	`squeue -p <partition>`	Display queue/partition entries
`qstat -f <job_id>`	`scontrol show job <job_id>`	Show job details
`qdel <job_id>`	`scancel <job_id>`	Delete <job_id>

Frequently used Options

PBS	Slurm	Meaning
qsub	sbatch/salloc	Submit batch/interactive job script to queue*
-l procs=<number>	--ntasks=<number>	# of processes to run
-l nodes=X:ppn=Y	--ntasks=<multiply X*Y>	# of processes to run
-l walltime=<HH:MM:SS>	--time=<HH:MM:SS>	How long the job will run
-l mem=<number>	--mem=<number>	Total memory (single node)
-l pmem=<number>	--mem-per-cpu=<number>	Memory per cpu
-l …:<attribute>	--constraint=<attribute>	Node property to request (avx, IB)
-q <queue_name>	--partition=<partition_name>	Which set of nodes to run job on

Ref: https://hpcc.usc.edu/support/documentation/pbs-to-slurm/

Examples

Interactive mode: $ salloc --ntasks-per-node=16 --nodes=2 --time=1:00:00 --constraint=IB

To submit a job: $ sbatch job.sl
where job.sl is:

```
#!/bin/bash
#SBATCH --ntasks-per-node=8
#SBATCH -N 2
#SBATCH --mem-per-cpu=2GB
#SBATCH -t 01:00:00
#SBATCH -p priya
#SBATCH -A lc_pv

# to run with pure MPI
Export OMP_NUM_THREADS=1
srun -n 16 --mpi=pmi2 ./a.out
```

To get node from a specified partition, need to specify your account as well.

```
#!/bin/bash
#SBATCH --ntasks-per-node=8
#SBATCH -N 2
#SBATCH --mem-per-cpu=2GB
#SBATCH -t 01:00:00

# to run with hybrid MPI & OMP
export OMP_NUM_THREADS=2
srun -n 8 -c 2 --mpi=pmi2 ./a.out
```

Ref: https://www.nersc.gov/users/computational-systems/cori/running-jobs/example-batch-scripts/

The srun "-c" option

- Request that *ncpus* be allocated **per process**. This may be useful if the job is multithreaded and requires more than one CPU per task for optimal performance.
- The "-c" value should be set as

$$\frac{\text{\# of logical cores(hyper thread) the node can support}}{\text{\# of MPI tasks per node}}$$

e.g. To run a program with 16 MPIs/node and 4 OMPs/MPI on KNL:

$$\frac{64 * 4\ (256\ logical\ cores\ on\ KNL)}{16} = 16$$

$ export OMP_NUM_THREADS=4

$ srun -n 16 -c 16 … ./exe

Ref: [1] https://www.nersc.gov/users/computational-systems/cori/running-jobs/example-batch-scripts-for-knl/
[2] https://www.alcf.anl.gov/files/Knight_Software_JobSubmission_2.pdf

Useful jobscript generator

- https://my.nersc.gov/script_generator.php