|  |  |  |
| --- | --- | --- |
| **User Commands** | **Slurm** | **SGE** |
| Job submission | sbatch [script\_file] | qsub [script\_file] |
| Job deletion | scancel [job\_id] | qdel [job\_id] |
| Job status (by job) | squeue --job [job\_id] | qstat -u \\* [-j job\_id] |
| Job status (by user) | squeue -u [user\_name] | qstat [-u user\_name] |
| Detailed job status | scontrol show JobId=[job\_id] |  |
| Update job | scontrol update JobId=[jobid] | qalter [jobid] |
| Job hold | scontrol hold [job\_id] | qhold [job\_id] |
| Job release | scontrol release [job\_id] | qrls [job\_id] |
| Queue list | squeue | qconf -sql |
| Node list | sinfo -N OR scontrol show nodes | qhost |
| Cluster status | sinfo | qhost -q |
| GUI | sview | qmon |
| Interactive session | srun -N 1 -n 1 --pty /bin/bash | qlogin |
|  |  |  |
| **Environment** | **Slurm** | **SGE** |
| Job ID | $SLURM\_JOBID | $JOB\_ID |
| Submit Directory | $SLURM\_SUBMIT\_DIR | $SGE\_O\_WORKDIR |
| Submit Host | $SLURM\_SUBMIT\_HOST | $SGE\_O\_HOST |
| Node List | $SLURM\_JOB\_NODELIST | $PE\_HOSTFILE |
| Job Array Index | $SLURM\_ARRAY\_TASK\_ID | $SGE\_TASK\_ID |
| Number of tasks | $SLURM\_NTASKS | $NSLOTS |
| Scratch directory | N/A | $TMPDIR |
|  |  |  |
| **Job Specification** | **Slurm** | **SGE** |
| Script directive | #SBATCH | #$ |
| Queue | -p [queue] | -q [queue] |
| Wall Clock Limit | -t [min] OR -t [days-hh:mm:ss] | -l h\_rt=[seconds] |
| Standard Output File | -o [file\_name]; SGE: -o "name.o%A.%a" | -o [file\_name] |
| Standard Error File | -e [file\_name] | -e [file\_name] |
| Combine stdout/err | (use -o without -e) | -j yes |
| Copy Environment | --export=[ALL | NONE | variables] | -V |
| Event Notification | --mail-type=[events] | -m abe |
| Email Address | --mail-user=[address] | -M [address] |
| Job Name | -J [name] | -N [name] |
| Job Restart | --requeue OR --no-requeue  (NOTE:configurable default) | -r [yes|no] |
| Working Directory | --workdir=[dir\_name] | -wd [directory] |
| Resource Sharing | --exclusive OR--shared | -l exclusive |
| Memory Size | --mem=[mem][M|G|T] OR --mem-per-cpu=mem][M|G|T] | -l mem\_free=[memory][K|M|G] |
| Account to charge | --account=[account] | -A [account] |
| Node Count | --nodes [min[-max]] | N/A |
| Scheduling priority | --nice=[nice value] | -p [priority] |
| Tasks Per Node | --tasks-per-node=[count] | (Fixed allocation\_rule in PE) |
| CPUs Per Task | --cpus-per-task=[count] | -pe [PE] [count] |
| Job Dependency | --depend=[state:job\_id] | -hold\_jid [job\_id | job\_name] |
| Job Project | --wckey=[name] | -P [name] |
| Job host preference | --nodelist=[nodes] AND/OR --exclude=nodes] | -q [queue]@[node] OR -q |
| Quality Of Service | --qos=[name] | [queue]@@[hostgroup] |
| Job Arrays | --array=[array\_spec]%[num tasks] | -t [array\_spec] -tc [num tasks] |
| Generic Resources | --gres=[resource\_spec] | -l [resource]=[value] |
| Licenses | --licenses=[license\_spec] | -l [license]=[count] |
| Submit now or fail | --immediate | -now |
| Requeue | --requeue | --no-requeue | -r [yes|no] |
| Begin Time | --begin=YYYY-MM-DD[THH:MM[:SS]] | -a [YYMMDDhhmm] |
| Wrap script with shell directive | --wrap [script] | -S /bin/bash |
| Output only jobid | --parsable | -terse |
| Email upon failure | --mail-type=FAIL --mail-user me@nyumc.org |  |

Adapted from http://slurm.schedmd.com/rosetta.pdf

See also https://github.com/mauranolab/sge2slurm

**Notes**

**Config file**: /etc/slurm/slurm.conf -- see https://computing.llnl.gov/linux/slurm/configurator.html

**Utilization**

sacct -o JobID,state,start,end,jobname,ncpus,cputime,MaxRSS,elapsed

sreport cluster AccountUtilizationByUser start=01/01/16

**Job control**

Change multiple jobs

for id in `squeue -o '%9F %.3p %45j %.8u %2t %19S %5P %.3C %.10K %R' -S 'P,-t,B,-p' | grep 562 | awk '$5!="R" {print $1}'`; do

scontrol update JobId=$id partition=defq

scontrol update JobId=$id Dependency=singleton

scontrol update JobId=$id qos=low

done

estimated start time: squeue -start

configured priority weights: sprio -w

priority scores per job: sprio -j

Bring down cluster. Setting node to down will requeue running jobs, without any failure code. Set to resume when done

scontrol update NodeName=isgnode001,isgnode002,isgnode003,isgnode004,isgnode005,isgnode006,isglcdctap001 State=drain Reason="Downtime"

**Cluster status**

counts per node:

squeue -o '%9F %.3p %45j %.8u %2t %19S %5P %.3C %.10K %R' -S 'P,-t,B,-p' | awk 'NR>1 && $5=="R"' | sort -k10,10 | awk '{for(i=1; i<=$8; i++) {print $10}}' | hist

counts per user:

squeue -o '%9F %.3p %45j %.8u %2t %19S %5P %.3C %.10K %R' -S 'P,-t,B,-p' | awk 'NR>1 && $5=="R"' | sort -k4,4 | awk '{for(i=1; i<=$8; i++) {print $4}}' | hist

**Node resource info**:

sinfo -o "%15N %10c %10m %25f %10G"

sinfo -o "%P %.5a %.10l %.10s %.4r %.8h %.10g %.6D %.11T %N %B %C %z %m"

scontrol show node