BIOWULF
HIGH PERFORMANCE COMPUTING AT THE NIH

Partitions:

norm(default)	- default partition. Restricted to single-node jobs
multinode	- for parallel jobs. Single-node jobs are not allowed.
gpu	- GPU nodes, including several flavors of GPUs
largemem	- Reserved for jobs with memory requirements that cannot fit on norm partition
unlimited	- no walltime limits
quick	- jobs < 4 hrs long. Will run on buyin nodes when they are free.
[ccr, forgo etc]	- buyin nodes

Job Submission: Useful sbatch options

--partition=abcd	Job to be run on partition 'abcd'.
--ntasks=#	Number of tasks to be run
--cpus-per-task=#	Number of cpus required for each task (e.g. '8' for an 8-thread multithreaded job)
--ntasks-per-core=1	Do not use hypercores (typically for parallel jobs)
--mem=#g	Memory required for the job
--constraint=nodetype	Node feature desired (e.g. 'x2650' for a parallel job)
--exclusive	Allocate the node exclusively
--no-requeue | --requeue	If an allocated node hangs, whether the job should be requeued or not.
--error=/path/to/dir/filename	Location of stderr file (by default, slurm######.out in the submitting directory)
--output=/path/to/dir/filename	Location of stdout file (by default, slurm######.out in the submitting directory)
--ignore-pbs	Ignore PBS directives in batch scripts. (by default, Slurm will try to utilize all PBS directives
--license=app	Request a license for 'app'
--time=HH:MM:SS	Set a walltime limit for this job
--dependency=afterany:JobID	Run this job after JobID completes
--gres=gpu:k20x:#	Request # GPU k20x processors (set # to '1' or '2' only)

Job deletion: Useful scancel options

--user=username	Delete all jobs belonging to 'username'
--name=JobName	Delete jobs with the jobname ' JobName
--state=RUNNING|PENDING	Delete jobs in RUNNING or PENDING state
--nodelist=node1,node2…	Delete jobs belonging to this user running on nodes node1, node2….

BIOWULF
HIGH PERFORMANCE COMPUTING AT THE NIH

Command	Example	Description
Display information about running and queued jobs		
squeue	squeue -u $USER	slurm equivalent to PBS 'qstat': shows all running and queued jobs on the cluster.
sjobs	sjobs	In-house script that uses 'squeue' and 'sacct' to show all running and queued jobs for a user.
showq	showq	In-house alias to 'squeue': shows all running and queued jobs
jobload [-j -u -n]	jobload -j 21232	displays load and memory usage for running jobs
showjob	showjob 22250	In-house alias to slurm's 'scontrol show job' : will display detailed information about all running jobs.
Display information about completed jobs		
jobhist	jobhist 21423	In-house script that reports data about completed or running jobs.
Display cluster information		
sinfo	sinfo	displays information about all Slurm nodes
batchlim	batchlim	In-house script that displays current per-user and walltime limits on all partitions
freen [-n]	freen	In-house script that concisely displays 'sinfo' data

BIOWULF
HIGH PERFORMANCE COMPUTING AT THE NIH

Default CPU allocation = 1 physical core (= 2 hypercores in old notation) = 2 CPUs in Slurm notation
DefMemPerCPU= 2 GB. Therefore, default memory alloc = 4 GB

Job type	Sample Command	Allocation		Notes
		CPU	Mem	
Single-threaded	`sbatch jobscript`	2	4 GB	
	`sbatch --mem=Mg jobscript`	2	M GB	
Multi-threaded	`sbatch --cpus-per-task=C --mem=Mg jobscript`	C (rounded up to nearest 2)	M GB	Default memory alloc = C*2 GB Use $SLURM_CPUS_PER_TASK inside the script to specify # threads
Swarm of single-threaded apps	`swarm -f swarmfile`	2 per task	1.5 GB per task	(swarm is set up with a lower default to account for the older B1 nodes)
	`swarm -g G -f swarmfile`	2 per task	G GB per task	
Swarm of multi-threaded apps	`swarm -t T -g G -f swarmfile`	T per task (rounded up to nearest 2)	G GB per task	Use $SLURM_CPUS_PER_TASK inside swarmfile to specify # threads

Interactive Job	`sinteractive`
	`sinteractive --mem=8g`
	`sinteractive --cpus-per-task=4`
	`sinteractive --ntasks=4 --ntasks-per-core=1`
		Default 2 CPUs	Default 4 GB memory	

BIOWULF
HIGH PERFORMANCE COMPUTING AT THE NIH

Job type	Command	Allocation		Notes
		CPU	Mem	
Parallel (MPI)	`sbatch [--partition=ibfdr] --ntasks=C [--constraint=nodetype] --exclusive --ntasks-per-core=1 jobscript`	2*C (C MPI procs on 2*C CPUs)	4 GB per MPI process	Use $SLURM_NTASKS inside script to specify # MPI processes. OpenMPI on B2 is built with slurm support, so in most cases $NP and $SLURM_NTASKS will not be needed.
	`sbatch [--partition=ibfdr] --ntasks=C [--constraint=nodetype] --exclusive --mem-per-cpu=Gg --ntasks-per-core=1 jobscript`	2*C (C MPI procs on 2*C CPUs)	2*G GB per MPI process	
Auto-threaded apps	`sbatch --exclusive jobscript`	At least 16 cores, 32 CPUs	C*2 GB	
	`sbatch --exclusive --constraint=cpu32 jobscript`	32	32*2 = 64 GB	
Swarm of auto-threaded apps	`swarm -t auto -f swarmfile`	32	1.5 GB	
	`swarm -t auto -g G -f swarmfile`	32	G GB	
Job array	`sbatch --array=0-10 jobscript`	2 per process	2 GB per process	Use $SLURM_ARRAY_TASK_ID inside script to specify array subjob
	`sbatch --array=0-10 --mem=Gg jobscript`	2 per process	G GB per process	