

Short Read Sequencing Analysis Workshop

Day 3 – Compute Cluster Usage and Data Transfer

Tortuga Queues

- compute – used for running all jobs on Tortuga with the exception of Jupyter (see below)
- notebook – used for running all Jupyter notebooks through Jupyterhub

Slurm Batch Script Directives

```
#SBATCH -p <queue> → specify job queue to run the job
#SBATCH --time=00:00:00 → hrs:minutes:seconds
#SBATCH --mem=<memory>
#SBATCH --ntasks= → specify how many cores/processors are needed for the job
#SBATCH --job-name=<jobname> → give the job a name
#SBATCH --mail-type=ALL
#SBATCH --mail-user=<you@email.com>
#SBATCH --error=<path/file.err> → specify name and location of error file
#SBATCH --output=<path/file.log> → specify name and location of log file
```

Slurm Command Line Utilities

```
$ sbatch <jobscript.sbatch> → submit a job to the queue
```

```
$ squeue → view the status of all jobs on cluster
```

```
$ squeue -u <username> → view status of jobs submitted by <username>
```

```
$ squeue -j <job id> → view the status of a specific job with job ID
```

```
$ scancel <job id> → delete/cancel a specific job with job ID
```

Data Transfer

scp = secure copy protocol usage: \$ scp <source> <destination>

to specify the server: <username>@<Host_Address>:/full/path

to copy recursively: scp -r

curl -O – used to download files from remote hosts Usage: \$ curl -O <url>

→ download file at <url> to current directory