

High performance computing

Due to the snow day please go through these instructions in class. Some of you may want to take advantage of the additional computing power available on campus beyond your personal computer. Read chapter 4 of your textbook for more information.

- If you are using Windows download and install PuTTY (<http://www.putty.org/>)
- Open your Terminal (Mac) or PuTTY (Windows)
- Connect to URI's seawulf cluster using your ssh (**secure shell**) and your URI username. For example

```
ssh rsschwartz@seawulf.uri.edu
```

Your password is set to your URI ID. You will need to change this on your first login.

There are two ways to run jobs on a cluster. First, you can run them interactively just like you do on your own computer. The advantage of this is you can work easily and directly. The disadvantage is that you might want to run a job that takes a long time and you'd like to take advantage of the computing power of a the cluster (that's why you're using it in the first place).

You are now logged in on on the "head node". A computing cluster is just that - multiple computers attached together. Obviously it would be inefficient for many people to use the same computer on the cluster. It could even crash the cluster!

- Access your own "node" of the cluster to work on by submitting an interactive job.

```
qsub -I
```

This gives you 1 CPU for a period of time. If you need more CPUs or more time you can specify that. For example to get 4 CPUs for 2 hours use `-l` then specify the number of nodes, processors per node, and the maximum time you will need them.

```
qsub -I -l nodes=1:ppn=4 -lwalltime=2:00:00
```

You can now run almost all the same commands you do on your own computer. Because the cluster is running Linux (CentOS) there may be a few that are slightly different. For example, both `man ls` and `ls --help` work on Linux. Additionally, while you have access to many installed programs you need to load them before you can use them.

- List available software on seawolf

```
module avail
```

- Because that list can be long limit your list to software starting with known values (e.g. P)

```
module avail P
```

- Load software so you can use it (for example Python)

```
module load Python/3.5.2-foss-2016b
```

You can use tab complete to get the full title of the module.

- List what you have loaded already

```
module list
```

- Run python (this is a programming language we will use later in the course) to show it has been loaded

```
python
```

- Use secure copy (scp) to move files between your computer and the cluster. Copy the script from the homework assignment to the cluster in a folder called homework (you should make this directory on the cluster first).

```
scp 01-run.sh rsshwartz@seawulf.uri.edu:homework
```

- Copy your whole data-shell folder to the cluster.

```
scp -r data-shell rsshwartz@seawulf.uri.edu:
```

- Make a change to a file in the data-shell folder. Rather than copying the whole folder again just update with these recent changes.

```
rsync -r data-shell rsshwartz@seawulf.uri.edu:
```

If you are working with analyses that will take some time you should not use an interactive job. Instead you will write a script to submit a job. Your script might look like

```
#!/bin/bash
#PBS -l walltime=1:00:00
#PBS -l nodes=1:ppn=4

bash ~/homework/01-run.sh ~/homework/saccharomyces_cerevisiae.gff
```

You are already familiar with the shebang line. All lines starting with `#PBS` indicate parameters related to job submission. `PBS -l walltime=1:00:00` allows the job to run for up to 1 hour. `PBS -l nodes=4:ppn=16` indicates you need 1 node with 4 processors per node. There are many other possible PBS parameters not included here.

- Exit your interactive session. Submit this batch file as a job by running `qsub job_example.sh`
- Check the status of your job (and others) using `qstat`.
- Your output and any errors have been sent to files rather than printed on the screen. Use `ls` and `cat` to see the files and your output.

If you submit a job that you need to cancel first run `qstat` to get the job id then run `qdel <jobid>`

Note: If you have any large or important data you are working you should not store it in your home directory. Additional storage is connected to the cluster and access can be arranged by the HPC manager.