Day 2 Lectures Notes

HPC Principles

- **Using Loops**

```
for (i in 1:10)}{
    do_simulation(i)
}
```

- **Using HPC**

```
as.numeric(Sys.getenv("PBS_ARRAY_INDEX"))
do_simulation(i)
```

Uses for HPC

- Login Address
- Types of HPC at Imperial
    - cx1: High-Throughput
    - cx2: High-End
        - Massively parallel tasks
    - ax4: Big Data
        - Large memories

Steps of using HPC

1. **Get your code onto the cluster**

```
sftp username@login.cx1.hpc.ic.ac.uk
put filename.R
exit
# Alternatively
scp path/to/file.txt
username@login.cx1.hpc.ic.ac.uk:/home/username/
```

2. **Log into the cluster**

```
ssh username@login.cx1.hpc.ic.ac.uk
ls #List the files in $HOME
mkdir foldername
mv filename $HOME/foldername
cd foldername
cat filename # See your file to check it's contents
module load anaconda3/personal
anaconda-setup # Set up anaconda - One time only
conda install r # One time only
```

3. **make a file for your shell script**

> Do NOT run code on the `login node`

```
#!/bin/bash
#PBS -l walltime=12:00:00
# (walltime should be editted by yourself)
#PBS -l select=1:ncpus=1:mem=1gb
# (mem should be editted by yourself)
module load anaconda3/personal
echo "R is about to run"
R --vanilla < $HOME/Rtest/ForwardsNTC_V5.R
mv datafilename* $HOME
echo "R has finished running"
# This is a comment at the end of the file
```

4. **Submitting your job to the cluster**

- to submit your job type

```
qsub -J 1-32 filename.sh
# 32 means your code will be run 32 times in parallel
qstat # S changes from Q to B when running
```

- to delete a job

```
qstat
qdel job-id[] # [] is for array jobs only
```

5. **Check that all is well**

```
qstat # is your job running still
ls
cat filename.sh.ejob-id.index # are error files empty?
cat filename.sh.ojob-id.index # are strandard putput files as
expected?
qstat # is your job running still?
exit
```

6. **Get your results back from the cluster**

```
qstat # is your job running still?
cd $HOME
ls
cat output filename # check the content
cat filename.sh.ejob-id.index # any errors?
cat filename.sh.ojob-id.index # output expected?
tar czvf filename.tgz *
mv filename.tgz $HOME
exit

# Then, to get file to you own computer
sftp username@login.cx1.hpc.ic.ac.uk
get filename.tgz
```

```
exit
# Untar your files on your local computer
tar xzvf filename.tgz
```

Some DO NOTs

- Do NOT use the cluster without knowing `memory` and `time requirements`
- Do NOT run jobs on the `login node`
- Do NOT try to use cx2 or ax4 parts of the cluster
- Do NOT output data to the `hard disk` regularly
- Do NOT use the `same random seed` for your simulations
- Do NOT copy and paste your shell script
- Do NOT leave your results in `$TMPDIR`
- Do NOT waste too much of your own time `optimizing your code`
- Do NOT run code on the cluster that hasn't been tested locally first