High Performance Computation (HPC) Basics

FAIR Data Practices for Omics Analysis Workshop

University of Delaware

April 19 (Day 2)

Starting with BASH

- BASH is used by issuing commands at the prompt

- The prompt is always at a location in the file system
  - When you login you should be in your "home folder": /home/username
  - Your home directory is where you can store your own files (and install software)

- You can run a command by typing its name at the prompt and hitting [enter/return]

- Some commands usually have options, called flags or parameters
  - Flags can stand by themselves.
    "`ls -l`": *ls* is a command and *-l* is a flag
  - Or they can require arguments
    "`tar -c -f file.tar input-dir`":
    - *-c and -f are* flags
    - *file.tar* is the argument for –f.
    - -c does not need an argument.
    - input-dir is a standalone argument. Some commands expect these and they almost always indicated what the input file (or directory) for a command is. They are usually the last thing listed in a command.

BIOINFORMATICS
CORE FACILITY

BASH Essentials

(90% of what you need to know in 1 slide)

- Command prompt is used to navigate the file system and is always set in a directory [*pwd*]
- Directory can be created [*mkdir*], changed [*cd*], or removed [*rmdir*]. Or its contents can be listed [*ls*]
- Files and directories have read, write, and execute permissions that can be set at the owner, group, and system levels [*chmod, chown, chgrp*]
- Files can be found [*find*], displayed [*cat, head, tail, less*], searched [*grep*], or edited/created [*nano, vi, emacs*]
- Scripts and binary applications with the execute permission can be run by typing their name
- Help about any command can be displayed [*man*] to learn about the syntax and options (parameters & flags)
- Standard output often goes to the screen by default, but can be saved (redirected) into a file using the > (replace) or >> (append) operators, or sent as input to another command using the | (pipe) operator
- Remember to use [*man*] to get help about almost any system command
- TIP: Spend some time learning to use the very powerful commands *grep*, *sed,* and *awk*

What is HPC?
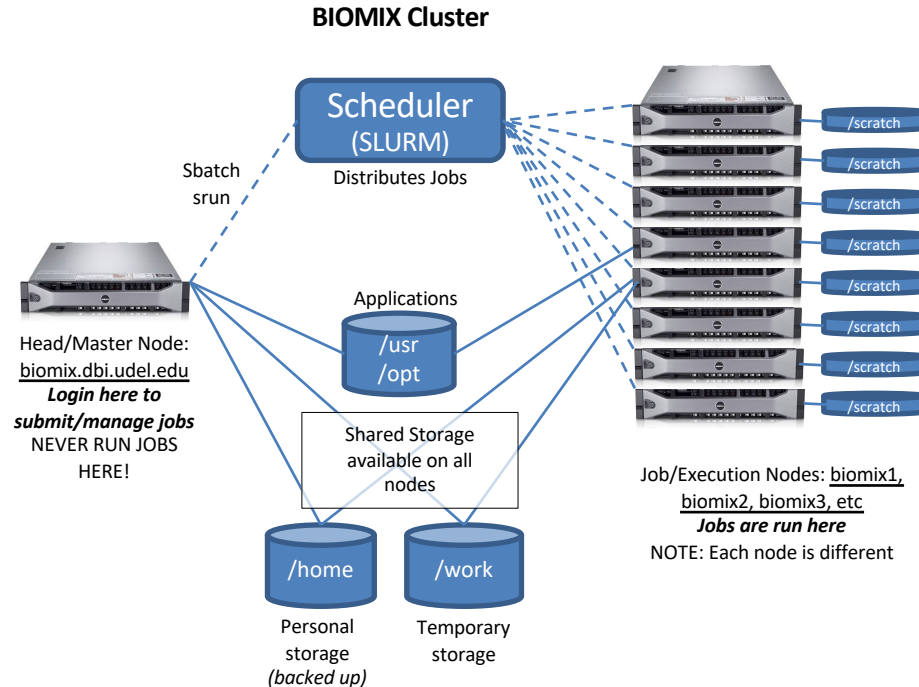
HPC = High Performance Computing

BIOMIX Cluster

- Co-hosted by CBCB and DBI

- Heterogeneous cluster with >2750 cpu cores

- Each node has 12-96 CPU cores and 128-2048GB of RAM and some amount of /scratch storage

- Each user has a home directory (storage quota)

- Username and password (change password with this: *passwd*)

- https://bioit.dbi.udel.edu (BIOMIX manual and software)

- https://bioit.dbi.udel.edu/BIOMIX/SLURM-intro.html (slurm info)

Biomix Nodes

High Performance Computing (HPC) Cluster

BIOMIX Cluster

Scheduler (SLURM)
Distributes Jobs

Sbatch srun

Head/Master Node: biomix.dbi.udel.edu
Login here to submit/manage jobs
NEVER RUN JOBS HERE!

Applications
/usr /opt

Shared Storage available on all nodes

/home

/work

Personal storage *(backed up)*

Temporary storage

/scratch
/scratch
/scratch
/scratch
/scratch
/scratch
/scratch
/scratch

Job/Execution Nodes: biomix1, biomix2, biomix3, etc
Jobs are run here
NOTE: Each node is different

BIOINFORMATICS CORE FACILITY

Interactive Submission

- Never run an application on the head node
- Use interactive submission when you want to experiment or manually run something
- srun -c 1 --mem=16000 --pty bash
  - -c 1 = reserve 1 cpu core
  - --mem = reserve 16000MB of memory (16GB)
  - --pty = run a pseudo-terminal
  - bash = run bash on the remote node
- IMPORTANT – When Finished Type: *exit*

Batch Submission

- Allows one or more commands to be submitted for automatic execution

- This is the typical way things are run on the cluster

- *sbatch script.slurm*

- By default STDOUT and STDERR goes to slurm-[jobid].out (we will change this a bit in our jobs)

Batch Submission
(Example slurm submission script)

```
#!/bin/bash
#SBATCH --job-name=blast
#SBATCH --mem=64000
#SBATCH -c 4          ← Number of cpu cores


#insert software requirements here

cd ~/exercises/blast-exercise          ← Tells the command what directory to run in…

echo "Start:"
date
hostname


# PUT THE COMMAND(s) YOU WANT TO RUN NEXT
blastp -query /home/spolson/exercises/blast-exercise/sequence.fasta -db /usr/local/blast_db/nr
-out /home/spolson/exercises/blast-exercise/blast-result.txt -num_threads 4 -outfmt 6
                                                                          ↑
                                                                   Number of cpu cores

echo "Complete:"
date
```

NOTE: A slurm software update eliminated the "#SBATCH --workdir" option that appears in the lecture video. This is replaced by the "cd" command below.

S.W. Polson, University of Delaware, Center for Bioinformatics and Computational Biology

BIOINFORMATICS
CORE FACILITY

Monitoring Cluster, Queue, Jobs

- List nodes and their capabilities:
  - *scontrol show nodes*
  - *cat /etc/dbi/NodesDB*

- List current jobs running:
  - *squeue*

- Details on a job 2992
  - *scontrol show job 2992*