

# Introduction to RSPH HPC Cluster

## Cluster Structure

---

- 25 Compute nodes
  - 24 of which have 32 cores and 196 GB RAM each
  - One "large memory node" with 1.5 TB RAM
  - A private node with partition "yanglab" (32 cores, 188GB RAM) available for yang lab members
- Connected via 25 GB Ethernet network
- 1 Petabyte Panasas parallel file system
  - Extra 10TB storage for yanglab `/projects/YangLabData/`
- Support 800 concurrent running job managed by SLURM

## Login

---

- First login to Emory VPN
- Login using `ssh` by typing the following command in respective terminal `ssh <your_Emorey_NetID>@clogin01.sph.emory.edu`
  1. On Mac OS X system, open a terminal app and type `ssh`
  2. Under Linux, open a terminal and type `ssh`
  3. Under Windows 10, open up a Powershell (search for "Powershell" in the search tool next to the Start menu) and type `ssh`. (Older versions of the operating system may require the installation of a third party `ssh` client, like Putty.)
- [Login without password](#)

## Mount Remote Directory to Local Computer

---

- Use [WinSCP](#) for Windows system
- Use [sshfs](#) for MAC
  1. Install [SSHFS 2.5.0](#)
  2. Install [macFUSE 4.1.0](#)
  3. Create a mount point directory on local machine `mkdir /Users/jyang51/RHPC`
  4. Mount RHPC server directory to your local mount directory Example command to mount remote directory `/home/jyang51/` to local mount directory `/Users/jyang51/RHPC/`

```
sshfs jyang@clogin01.sph.emory.edu:/home/jyang51/ /Users/jyang51/RHPC/ -o auto_cache -ovolname=RHPC -o follow_symlinks
```

## Copying Data to the Cluster

---

- General command by using `scp` as follows: `scp <path_to_file> <username>@<remote_system_name>:<destination_path>`

## Using a Software

---

### Check if a software is available on the cluster

- `which [software_name]`
- `module spider` to list all installed software modules
- `module spider [software_name]` to find detailed information about the software

### Using a software installed on the cluster

- `module load [software_name]` to load the software for using

- For example, type command `R` to start using R after loading R by `module load R/4.0.3`

## Install a software without root access

- Install software under `~/.local/`

## Share installed software with the lab

- Make a soft link of the executable tool directory under `/projects/YangLabData/bin` by the following commands

```
cd /projects/YangLabData/bin/
ln -s [tool directory]
```

- Add `/projects/YangLabData/bin` into the `PATH` environment variable by including the following command line in your `~/.bashrc` file

```
export PATH=/projects/YangLabData/bin:$PATH
```

## Submit Jobs by SLURM

### Basic SLURM commands

- `sbatch` to submit jobs (like `qsub` in Grid Engine)
  - Set arguments to `sbatch` in a wrapper shell (job submission) script, for example, you may use `sbatch normal.sh` with the following `normal.sh` script:

```
#!/bin/bash
#SBATCH --job-name=normal.R
#SBATCH --partition=yanglab
## This puts all output files in a separate directory.
#SBATCH --output=Out/normal.%A_%a.out
#SBATCH --error=Err/normal.%A_%a.err
## Submitting 100 instances of srun commands listed below
#SBATCH --array=0-100
## For notification purposes. Use your Emory email address only!
#SBATCH --mail-user=<your_email_address>@emory.edu
#SBATCH --mail-type=END,FAIL
module purge
module load R
srun /home/<user>/normal.R
```

- `squeue` to display information about the run queue
- `scancel [jobid]` used to cancel or kill a job
- `scontrol` used to show information about running or pending jobs
  - `scontrol show job [jobid]` to show system details of a submitted job
  - `scontrol show partition day-long-cpu` to show partition configuration
- `srun` used to run an interactive instance
  - Open an interactive session to test your *R* code:

```
module load R
srun -p interactive-cpu --pty bash
R
```

- Simply exit the interactive session by `exit`
  - **Do not test running *R* scripts on the login node *clogin01***
- `sinfo` used to report the state of the cluster partition and nodes

## Storage on the Cluster

## Sub volumns all mounted across the cluster

- `/home` where user home direcotry is located
  - Each user have a **25GB** quota for home directory, suitable for running computations
  - Retains two snapshots of data at `/home/.snapshot` . If a user accidentally deletes a file, one of two possible copies of the file can be restored from the snapshot directory
- `/scratch` where user may load data for immediate computation
  - Usable for large computations with the **two-week maximum retention policy**
  - **100GB/user**
- `/projects` where groups may store project data and run computations. Requestable in increments of **1TB**, \$75 per year
  - **10TB** storage space is available for yang lab under `/projects/YangLabData/`
  - You may creat a directory under this lab storage space for your work
- `/apps` and `/projects/YangLabData/bin` where all softwares can be found

## Monitoring storage usage

- With command `pan_quota`

## All files are not backed up on the cluster

## Additional Resources

---

- Request cluster assistance or an account: Email [help@sph.emory.edu](mailto:help@sph.emory.edu)
- The [Emory VPN](#)
- Software [Modules](#)
- SLURM [Quickstart](#)
- Linux Commands Video training module accessible to all Emory affiliates persons via [LinkedIn learning](#)