Intro to High Performance Computing (HPC) and Open OnDemand (OOD)
Learning Objectives

- Why HPC?
- What is HPC?
- Supercomputer components
- FASRC clusters
  - Cannon
  - FASSE (FAS secure environment)
- Open OnDemand/VDI
  - Jupyter Notebook, RStudio, Stata
Why HPC?

• Size: problems that can’t fit on a PC (personal computer), for example 500 GB of RAM
• Speed: problems that take months on a PC may take a few hours on a supercomputer
• Amount: need 1000s of runs
What is HPC?

- HPC: biggest and fastest computing machines right now
- Supercomputers: rule of thumb - at least 100 times as powerful as a PC (personal computer)
- Jargon: other terms
  - Supercomputing
  - Cyberinfrastructure (CI)
  - Cluster computing
Node, processors, core

**Node**
- A computer in the cluster

**Core, CPU, processor**
- CPU: central processing unit
  - can have many cores
- Cores
  - basic unit of compute
  - runs a single instruction of code
Cluster

Cannon Cluster

Compute Node e.g. holy7c24604

Compute Node e.g. holy7c24605

Compute Node
Cluster

Rack

Interconnect: Infiniband Switch

Compute Node

From HPC@LSU training (http://www.hpc.lsu.edu/training/weekly-materials/2022-Fall/HPC_UserEnv1_Fall2022.pdf)
Node

GPU
- Graphics processing unit
- Accelerator
- At FASRC: NVidia A100s

From HPC@LSU training (http://www.hpc.lsu.edu/training/weekly-materials/2022-Fall/HPC_UserEnv1_Fall2022.pdf)
Cluster Architecture
FASRC Clusters

- Cannon - general purpose
- FASSE - secure environment
  - The FAS Secure Environment (FASSE) is a secure multi-tenant cluster environment to provide Harvard researchers access to a secure enclave for analysis of sensitive datasets with DUA's and IRB’s classified as Level 3.
Login and access

Cannon

IQSS Cannon Quickstart Guide

What is Cannon?
Cannon is the Faculty of Arts and Sciences research computing cluster for users with Data Security Level 2 data. This guide explains how to begin using Cannon. If you have Data Security Level 3 data, you must use the FAS Secure Environment (FASSE) cluster.

Fun fact: Cannon is named after the early 20th century Harvard astronomer Annie Jump Cannon.

Pre-requisite steps
Get set up on FASRC:
1. Get a FASRC account
   1. Important: Be sure to request “FASRC Cluster Access” on the “Services” page
2. Set your FASRC password
3. Configure 2FA
4. Configure VPN

https://docs.rc.fas.harvard.edu/kb/iqss-cannon-quickstart-guide/

FASSE

IQSS FASSE Quickstart Guide

What is FASSE?
FASSE is the Faculty of Arts and Sciences Secure Environment research computing cluster. FASSE is available for users with Data Security Level 3 data. This guide explains how to begin using FASSE.

Pre-requisite steps
Get an account on FASSE: https://docs.rc.fas.harvard.edu/kb/get-a-fasse-account-and-project-group/

https://docs.rc.fas.harvard.edu/kb/iqss-fasse-sid-quickstart-guide/
What is Open OnDemand?

- Open source web portal to access clusters
- Web-based, no software needs be installed on your local machine (except for a modern browser like Google Chrome, Mozilla Firefox)
- Easy to use and simple to learn
- Very similar to desktop applications
- The easiest way to run GUI applications remotely on a cluster
- Individual applications, remote desktop, shell, file browser

Modified from https://drive.google.com/file/d/1Kc14V7z9FahM3jst1hmJ2C5ZkhEJhKhg/view
OOD on Cannon and FASSE

Cannon

FASSE

https://vdi.rc.fas.harvard.edu/pun/sys/dashboard

https://fasseood.rc.fas.harvard.edu/pun/sys/dashboard
Jupyter Notebook

You can create your own kernels using conda AND command line

1. Connect to cluster via ssh using command line:
   https://docs.rc.fas.harvard.edu/kb/terminal-access/

2. Create conda environment with packages that you need:
   https://docs.rc.fas.harvard.edu/kb/python/

3. Install package ipykernel so you can see the conda environment in Jupyter Notebook:
   https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/
   #optional_Creating_and_loading_a_conda_environment

4. Launch new Jupyter Notebook session

5. Select the newly create conda environment (kernel)
RStudio Server vs. RStudio Desktop

RStudio Server

- Go-to RStudio application with many precompiled R packages
- Cannot set `R_LIBS_USER`
- `R_LIBS_USER` is set to `~/R/ifxrstudio/<IMAGE_TAG>`
- Cannot use `module load`
- Cannot use slurm commands (e.g. `sbatch`)

RStudio Desktop

- Highly customized environment
- Can set `R_LIBS_USER`
- Can use `module load` → you can set specific compilers (e.g. `openmpi`, `gcc`)
- Can use slurm commands (e.g. `sbatch`)

Documentation: [https://docs.rc.fas.harvard.edu/kb/rstudio-server-vs-rstudio-desktop/](https://docs.rc.fas.harvard.edu/kb/rstudio-server-vs-rstudio-desktop/)
Stata

- Stata can be run from the Open on Demand/VDI interface
- Documentation: [https://docs.rc.fas.harvard.edu/kb/stata-on-cluster/](https://docs.rc.fas.harvard.edu/kb/stata-on-cluster/)
- On FASSE may need to have proxy settings changed for loading certain HTTP-only libraries
  - if you set httpproxy on, make sure to set httpproxy off before you end the session
Remote Desktop app

- It can be used to launch most GUI applications
- How
  - First: load module
  - Second: set environmental variables (not always needed)
  - Third: Launch software
- Documentation: [https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/](https://docs.rc.fas.harvard.edu/kb/ood-remote-desktop-how-to-open-software/)
Request Help - Resources

- [https://docs.rc.fas.harvard.edu/kb/support/](https://docs.rc.fas.harvard.edu/kb/support/)
  - Documentation
    - [https://docs.rc.fas.harvard.edu/](https://docs.rc.fas.harvard.edu/)
  - Portal
    - [http://portal.rc.fas.harvard.edu/rcrt/submit_ticket](http://portal.rc.fas.harvard.edu/rcrt/submit_ticket)
  - Email
    - rchelp@rc.fas.harvard.edu
  - Office Hours
    - Wednesday noon-3pm [https://harvard.zoom.us/j/255102481](https://harvard.zoom.us/j/255102481)
  - Consulting Calendar
    - [https://www.rc.fas.harvard.edu/consulting-calendar/](https://www.rc.fas.harvard.edu/consulting-calendar/)
  - Training
    - [https://www.rc.fas.harvard.edu/upcoming-training/](https://www.rc.fas.harvard.edu/upcoming-training/)
Login & Access - Connect to Cannon

Once you have an account you can use the Terminal to connect to Cannon

- Mac: Terminal
- Linux: Xterm or Terminal
- Windows: SSH client - Putty or Bash Emulator - Git Bash

$ ssh username@login.rc.fas.harvard.edu

- ssh stands for Secure SHell
- ssh is a protocol for data transfer that is secure, i.e. the data is encrypted as it travels between your computer and the cluster (remote computer)
- Commonly used commands that use the ssh protocol for data transfer are, scp and sftp
Command line vs. OOD

**Command line**

- **Pros**
  - Very efficient for experienced users
  - Good for large-scale job submission and data processing

- **Cons**
  - Very steep learning curve
  - No GUI

**OOD**

- **Pros**
  - Simple
  - GUI
  - Similar to desktop applications

- **Cons**
  - Not as efficient as command line
  - (Mostly) limited to single node jobs

Modified from https://drive.google.com/file/d/1Kc14V7z9FahM3jst1hmJ2C5ZkhEJhKhg/view