Farm and Slurm

ECL 243, Feb-March 2018

Info about the UCDFarm:
https://github.com/RILAB/lab-docs/wiki/Farm-Usage

https://github.com/WhiteheadLab/Lab_Wiki/wiki/Using-the-farm-cluster

https://github.com/ngs-docs/2016-adv-begin-shell-genomics/blob/master/ucd-farm-intro.md

Farm Basics

What the Farm is:

- A bunch of computers working together to make a computing cluster

Getting onto the Farm:

- use a shell to help you interface with the whole system
  - there are a lot of different shells (PuTTY, GitBash, Command, Power Shell), they all can do different things
    - PuTTY is a shell that networks with other clusters
    - using the GitBash shell for this class
    - Many programs can only be run in the shell/command line
      - ex. plink is a program that has to be run into a shell
- need an SSH key

  - SSH: secure shell, "a network protocol for operating network services securely over an unsecured network. The best known example application is for remote login to computer systems by users." (wikipedia)

  - You can make SSH keys with you GitBash shell using: http://happygitwithr.com/ssh-keys.html. It will make one private and public key.

- need a farm account

o   just fill out an application, add keys, then wait

Interacting with the Farm:

Basics

- sign on with:
  ssh myname@farm.cse.ucdavis.edu
- The Farm is Linux, which is a popular Unix OS, so it uses the same language at the shells
- SLURM: communicates with the Farm OS to allocate needs and jobs
- Head Node: the top CPU in the cluster. Make sure you aren't on it, because this is what SLURM uses to do its thing
- when you start it will automatically load slurm
  ```
  module slurm/
  ```
  o   First thing, get off the head node with
  ```
  srun -t 02:00:00 -p ecl243h --pty bash
  ```
  o   here, "-t 00:02:00" means get me off the head node for 2 hours and give me a corner of the cluster (a node) to work with. This will be automatically assigned, after which you are kicked back to the head node.
  - the longer time you ask for, the lower priority it is for the cluster

Language/commands

- squeue #tells you all the jobs on the cluster right now
- grep #means find
- squeue | grep "myname" #will find only my jobs
- smap -c #slurm map clean, a cleaner version of squeue
- smap -c | grep "myname" #just like squeue | grep "myname"
- pwd #where you are in the farm. It will show you in home, which is the directory where users live
- cd #change directory
- ls #lets you see whats in the directory
- cd … #takes you up a directory (the space is important)
- ls … #tells you what's in the directory above you
  o   ls …/… #let's you see two directories up

- mkdir name #makes a directory
- rmdir name #removes it
  - Farm gives me the "myname" directory, I need to make the others
- cp #copies a file to rename in a new format
  ```
  cp polarbearSNPcleanMar8new.txt polarbearSNPclean.hmp.txt
  ```
- mv #move something from the folder I'm in to another one
  ```
  mv sortex.hmp.txt /home/myname/ecl243/data/tassel-5-source/
  ```
- module load "program" #loads the program you want
- man #to look up the manual for a command. So for example, man ls will bring up the different options for the ls command you can use. man mkdir, man cd, man sbatch, and so on
- nano
  - #nano is a text editor with two functions
  - nano hello #make a file called hello in a new screen
  - write in the file, follow the commands at the bottom to exit and save
  - nano hello #now it opens the existing file "hello"
  - rm hello #removes that file
  - **To write scipts: nano hello.sh #adding ".sh" tells slurm this is a shell file, meaning it's a script**

Transfering files to the Farm

- use the scp command, with examples shown in this article: https://research.computing.yale.edu/support/hpc/user-guide/transfer-files-or-cluster
- If I am in the directory on my computer that I want to transfer a file from, then use:
  ```
  $ scp polarbearSNPperfect2.txt
  myname@farm.cse.ucdavis.edu:/home/myname/ecl243/data
  ```
  - man scp to see how our particular cluster wants your computer's directory and your username/cluster directory formatted.
- can also use winscp, comes with GUI

Running scripts

- nano the name of what you will call your script
  - start with the blue bash script to tell slurm you are initiating a bash script in a new screen

```
#SBATCH -D /home/jri/
```

```
#SBATCH -o /home/jri/stdout-%j.txt
#SBATCH -e /home/jri/stderr-%j.txt
#SBATCH -J steve #SBATCH --mem 8000
#SBATCH -t 0:05:00
set -e
set -u
```

start with #!/bin/bash -l to tell computer we are in bash

-D #directory we are working in

-o #where we want the SLURM output to be, we have to tell slurm where to put things, as well as where the script will put things

-e #error output file

-J #job name

-t #how much time to take

–mem=44g #how much gigabytes you think you need

set -e #sets that error happens, we don't know about set -u

- module load program
    - o #then type your script
- when done running, you need to call it on slurm in the main screen
- sbatch name #putting your work order in to slurm
    - o can add variables here, don't need to do that thought bc its already in the script

Using the Farm with Friends!

- You can cd into other people's directories to copy files from or just snoop around.
    - o For instance,
      `cd /home/myname/ecl243/data/` will take one into mynames's directories where you can nano shell scripts or whatever you feel like.
    - o You can't change files around, but you *can* copy things out of my directory with the cp command. So you'd type `cp /home/myname/ecl243/ants/data/run\_salm.sh /home/yourname/ecl243/scripts/` to copy the run_salm.sh file from my directory into a hypothetical scripts directory you've made.

Running the Farm through R

- Use "system" command in r to run things in unix, check above resources for notes on running r with the farm