UF Research Computing: An Introduction

Matt Gitzendanner: magitz@ufl.edu
Alex Moskalenko: om@hpc.ufl.edu

8/28/12

UF | Information Technology
www.it.ufl.edu

UF Research Computing

UNIVERSITY *of* FLORIDA

‣ Mission
  ◦ Improve opportunities for research and scholarship
  ◦ Improve competitiveness in securing external funding
  ◦ Provide high-performance computing resources **and support** to UF researchers

UF | Information Technology
www.it.ufl.edu

UF Research Computing

‣ Funding
  ◦ Faculty participation (i.e. grant money) provides funds for hardware purchases
    • Matching grant program!
‣ Any UF Faculty can use
  ◦ Up to 8 cores at a time
  ◦ Investors gain priority and access to additional resources
‣ Comprehensive management
  ◦ Hardware maintenance and 24x7 monitoring
  ◦ Relieve researchers of the majority of systems administration tasks

UF | Information Technology
www.it.ufl.edu

Matching Program

Consolidating Resources to Improve Efficiency and Capacity

RESEARCH FACULTY
$321k

+

RESEARCH COMPUTING
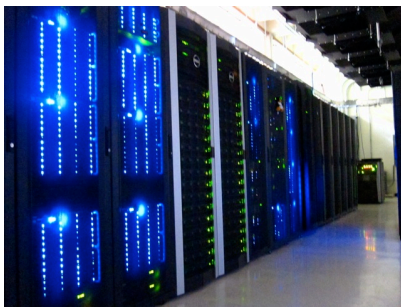$321k

=

RESEARCH INFRASTRUCTURE
$642k

The Research Computing Matching Program pooled $642k, thereby creating synergies and improving research infrastructure.

UF | Information Technology
www.it.ufl.edu
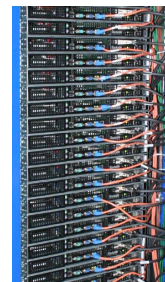

UF Research Computing

UF | Information Technology
www.it.ufl.edu

UF Research Computing

‣ Shared Hardware Resources
  ◦ Over **6K cores** AMD and Intel
  ◦ High-speed, low-latency **InfiniBand** interconnects
  ◦ **>1 PB**, high performance Lustre and Nexenta storage
  ◦ **GPUs**—CUDA code, etc.
  ◦ Several large memory (**512GB**) nodes

UF | Information Technology
www.it.ufl.edu

UF Research Computing
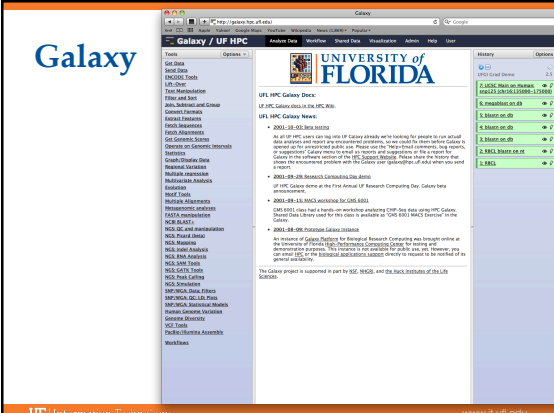

Where do you start?

UF Research Computing

- Account Policies
  - Personal activities are strictly prohibited on HPC Center systems
  - Class accounts deleted at end of semester
  - **Data are not backed up!**
  - Home directories must not be used for I/O
    - Use /scratch/hpc/$USER
  - Storage systems may not be used to archive data from other systems
  - Passwords expire every 6 months

What can you run?

- Galaxy
- Linux
- Generally command line driven applications
- Graphical apps can be setup
  - SAS
  - BEAUti
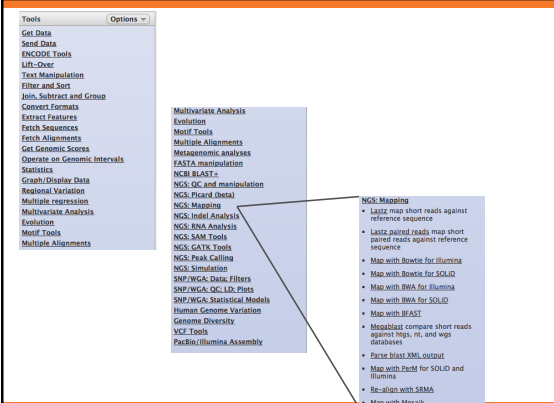
Galaxy

Galaxy: Data intensive biology *for everyone*

- Accessible, reproducible, transparent computational biology

- galaxy.hpc.ufl.edu
  - Local instance of Galaxy
    - Faster access to storage, easier upload
    - Local compute resources
    - Local control

Galaxy

Map with Bowtie for Illumina

Will you select a reference genome from your history or use a built-in index?:
Use a built-in index
Built-ins were indexed using default options

Select a reference genome:

if your genome of interest is not listed – contact Galaxy team

Is this library mate-paired?:
Single-end

FASTQ file:

Must have ASCII encoded quality scores

Bowtie settings to use:
Commonly used
For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:
☐
Bowtie produces SAM with several lines of header information by default

Execute

Galaxy

What can you run?

- Galaxy
- Linux
- Generally command line driven applications
- Graphical apps can be setup
  ◦ SAS
  ◦ BEAUti

Cluster basics

User interaction	Scheduler	Compute resources
Login node (Head node)	Tell the scheduler what you want to do	Your job runs on the cluster

Cluster login

submit.hpc.ufl.edu
ssh
submit1
submit2
/home/$USER

ssh <user>@submit.hpc.ufl.edu

Windows: PuTTY
Mac/Linux: Terminal

User interaction

Login node (Head node)

Cluster login

submit.hpc.ufl.edu
ssh
submit1
submit2
/home

ssh <user>@sub

Window
Mac/Lin

User interaction

```
Last login: Mon Jun 11 21:49:41 on ttys000
Voyager-II:~ matt$ ssh magitz@submit.hpc.ufl.edu
magitz@submit.hpc.ufl.edu's password:
Last login: Tue Jun 12 16:01:13 2012 from submit.hpc.ufl.edu

Welcome to the UF HPC Center.

Do not run interactive jobs on the login nodes.  If you need to
run an interactive job, there are interactive/test nodes for that.

UF HPC Center Account Policies can be found here:

http://www.hpc.ufl.edu/users/accounts.php

[magitz@submit1 ~]$ pwd
/home/magitz
[magitz@submit1 ~]$
```

Logging in

Linux Command Line

▸ Lots of online resources
  ◦ Google: Linux cheat sheet
▸ Training sessions
▸ User manuals for applications

Storage at HPC

submit.hpc.ufl.edu

ssh

submit1
submit2

/home/$USER

/scratch/hpc/$USER

`$ cd /scratch/hpc/$USER/`

Copy your data to submit using **scp** or a SFTP program like Cyberduck or FileZilla

UF Research Computing

▸ Storage
  ◦ **Home**: /home/$USER
    · For code compilation and user file management only, *do not write job output here*
  ◦ **Scratch space**: Lustre File System
    · /scratch/hpc/$USER
    · 500GB per lab
  ◦ Other options available for purchase and bio investors

/scratch/hpc/ must be used for all file input/output during job execution

Scheduling a job

▸ Need to tell scheduler what you want to do
  ◦ **How many CPUs** you want and how you want them grouped
  ◦ **How much RAM** your job will use
  ◦ **How long** your job will run
  ◦ The commands that will be run

Scheduler

Tell the scheduler what you want to do

UF Research Computing

▸ Ordinary Shell Script

```
#!/bin/bash

date
module load test_app
test_app –i file.txt
```

Read the manual for your application

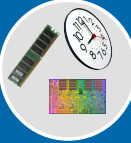Commands typed on the command line can be put in a script.

4

UF Research Computing

▸ Submission Script

```
#!/bin/bash
#
#PBS -N My_Job_Name
#PBS -M Joe_Shmoe@ufl.edu
#PBS -m abe
#PBS -o My_Job_Name.log
#PBS -j oe
#PBS -l nodes=1:ppn=1
#PBS -l walltime=00:05:00
#PBS -l pmem=900mb

cd $PBS_O_WORKDIR
date
module load test_app
test_app -i file.txt
```

Compute resources

Your job runs on the cluster

Scheduler

Tell the scheduler what you want to do
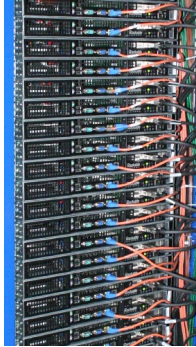
Nodes and processors

#PBS -l nodes=1:ppn=4
#PBS -l nodes=2:ppn=8

RAM

#PBS -l pmem=900mb

▸ Lots to consider, but do your best at estimating RAM needed for job
▸ Over about 2GB of RAM, "costs" toward CPU allocation

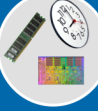Wasted RAM lead to idle CPUs and low job throughput

Walltime

#PBS -l walltime=00:50:00

• Fairly straight forward

• As with all resource requests, accuracy helps ensure *your* jobs and all other jobs will run sooner

Scheduler

Tell the scheduler what you want to do

UF Research Computing

▸ Job Management
  ◦ qsub <file_name>: job submission
  ◦ qstat -u <user>: check queue status
  ◦ qdel <JOB_ID>: job deletion

UF Research Computing

▸ Job Scheduling and Usage
  ◦ Test nodes (test01-06) available for interactive use, testing and short jobs (**EL5**: 01, 04, 05; **EL6**: 02,03,06)
    • e.g.: ssh test01
  ◦ Job scheduler selects jobs based on priority
    • Priority is determined by several components
    • Investors have higher priority
    • Non-investor jobs limited to 8 processor equivalents (PEs)
    • RAM: requests beyond a couple GB/core starts counting toward the total PE value of a job

Training Schedule

- ✓ Aug 28: Intro to UFHPC, getting started
- ▸ Sept 10: Modules, RHEL6 Transition, User Q&A
- ▸ Sept 17: The Linux/Unix Shell - An Introduction
- ▸ Sept 24: Running Jobs, Submission Scripts, Modules
- ▸ Oct 1: Galaxy Overview, The Basics
- ▸ Oct 8: NGS Data Techniques: General Methods and Tools
- ▸ Oct 15: NGS Data Techniques: Reference Based Mapping
- ▸ Oct 22: NGS Data Techniques: de Novo Assembly
- ▸ Oct 29: Phylogenetic Analyses
- ▸ Nov 5: Multiprocessing at the HPC Center
- ▸ Nov 12: Using Git and CMake to Organize and Drive Data Analysis Pipelines
- ▸ Nov 19
- ▸ Nov 29
- ▸ Dec 3

UF | Information Technology   www.it.ufl.edu

UF Research Computing

- ▸ Help and Support
  - ◦ Help Request Tickets
    - · https://support.hpc.ufl.edu
    - · For any kind of question or help requests
    - · Searchable database of solutions
  - ◦ We are here to help!
    - · support@hpc.ufl.edu

UF | Information Technology   www.it.ufl.edu

UF Research Computing

- ▸ Help and Support (Continued)
  - ◦ http://wiki.hpc.ufl.edu
    - · Documents on hardware and software resources
    - · Various user guides
    - · Many sample submission scripts
  - ◦ http://hpc.ufl.edu/support
    - · Frequently Asked Questions
    - · Account set up and maintenance

UF | Information Technology   www.it.ufl.edu