

## UF Research Computing: An Introduction

Matt Gitzendanner  
[magitz@ufl.edu](mailto:magitz@ufl.edu)

5/14/13

**UF** Research Computing  
Information Technology  
Home of High-Performance Computing and *HiPerGator*

UF | Information Technology | www.it.ufl.edu

## UF Research Computing


**UF** Research Computing  
Information Technology  
Home of High-Performance Computing and *HiPerGator*

- ▶ 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
  - Matching grant program!
- ▶ Any UF Faculty can use
  - Up to 8 cores
  - 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



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

UF | Information Technology | www.it.ufl.edu

UNIVERSITY OF FLORIDA | High-Performance Computing




# HiPerGator

The University of Florida Supercomputer for Research

UF | Information Technology | www.it.ufl.edu

## UF Research Computing


- ▶ Shared Hardware Resources
  - Over **22K cores**
  - High-speed, low-latency **InfiniBand** interconnects
  - **>3 PB**, high performance Lustre and Nexenta storage
  - **GPGPUs**— 90+, new Kepler-class
  - Several large memory (**512GB to 1TB of RAM**) nodes



UF | Information Technology | www.it.ufl.edu

## Investing in the future

- ▶ Internet2 Innovation Platform
  - 100 Gbps connectivity
  - Campus Research Network now 200 Gbps
- ▶ UF Data Center on Eastside Campus
  - 10,000 sq.ft and 1.75 MW total
  - 5,000 sq. ft. space for Research Computing
- ▶ **HiPerGator**
  - 16,384 cores
  - Infiniband interconnect
  - 2.1PB fast, high-availability, storage
  - Current ~7,000 core cluster will be integrated—Aug 2013



UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## UF Research Computing



Where do you start?

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)


## UF Research Computing

- ▶ User Accounts
  - Qualifications:
    - Current UF faculty sponsor
- ▶ 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

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)


## UF Research Computing




- ▶ User Accounts
  - Qualifications:
    - Current UF faculty, UF graduate student, and researchers
  - Request at: <http://www.hpc.ufl.edu/support/>
  - Requirements:
    - GatorLink Authentication
    - Faculty sponsorship for graduate students and researchers



UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

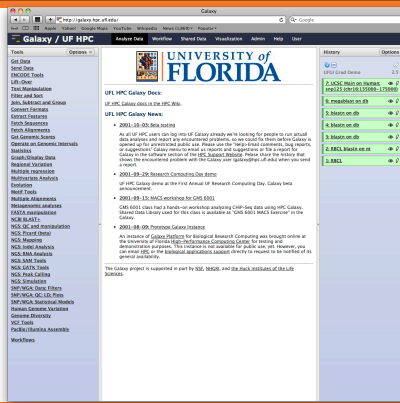
## What can you run?



- ▶ Linux-based
- ▶ Generally command line driven applications
- ▶ Galaxy 
- ▶ Graphical apps can be setup
  - SAS 
  - BEAUTi 

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## Galaxy



UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## 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

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Left-Click
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analysis
- FASTA manipulation
- NGS-BLAST
- NGS-DC and manipulation
- NGS- Picard (beta)
- NGS-Mapping
- NGS-Indel Analysis
- NGS-RNA Analysis
- NGS-SAM Tools
- NGS-GATK Tools
- NGS- Peak Calling
- NGS-Simulation
- SNP/WGA Data Filters
- SNP/WGA-DC, LD, Plots
- SNP/WGA-Statistical Models
- Human Genome Variation
- Genome Diversity
- VCF Tools
- PacBio/Illumina Assembly

**NGS Mapping**

- Lazy map short reads against reference sequence
- Lazy paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with BWA for Illumina
- Map with BWA for SOLiD
- Map with BEAT
- Mapblast compare short reads against Hg18, hg, and wgs databases
- Parse blast XML output
- Map with Pindel for SOLiD and Illumina
- Re-align with SBMA
- Map with Mosik

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## Galaxy

Map with Bowtie for Illumina

Will you select a reference genome from your history or 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?:

FASTQ file:  
 Must have ASCII encoded quality scores

Bowtie settings to use:  
 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

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## Galaxy

Galaxy / UF HPC

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Left-Click
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analysis
- FASTA manipulation
- NGS-BLAST
- NGS-DC and manipulation
- NGS- Picard (beta)
- NGS-Mapping
- NGS-Indel Analysis
- NGS-RNA Analysis
- NGS-SAM Tools
- NGS-GATK Tools
- NGS- Peak Calling
- NGS-Simulation
- SNP/WGA Data Filters
- SNP/WGA-DC, LD, Plots
- SNP/WGA-Statistical Models
- Human Genome Variation
- Genome Diversity
- VCF Tools
- PacBio/Illumina Assembly

Workflow: 1. NGS Mapping (Commonly used) → 2. NGS Mapping (Commonly used) → 3. NGS Mapping (Commonly used)

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## What can you run?

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

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## Cluster basics

User interaction

Login node (Head node)

Scheduler

Tell the scheduler what you want to do

Compute resources

Your job runs on the cluster

UF Information Technology [www.it.ufl.edu](http://www.it.ufl.edu)

## 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)

UF Information Technology www.fl.ufl.edu

## Cluster login

submit.hpc.ufl.edu  
ssh  
magitz@submit1 ~\$ ssh - bash - 67x17  
Last login: Mon Jun 11 21:49:41 on ttys000  
Voyager-11: 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 ~]\$

User interaction

Window Mac/Linux

UF Information Technology www.fl.ufl.edu

## Logging in

PuTTY Configuration

Category: Session, Logging, Terminal, Keyboard, Bell, Features, Window, Appearance, Behaviour, Translation, Selection, Colours, Connection, Data, Proxy, Telnet, Rlogin, SSH, Serial

Basic options for your PuTTY session

Specify the destination you want to connect to

Host Name (or IP address) submit.hpc.ufl.edu Port 22

Connection type:  Raw  Telnet  Rlogin  SSH  Serial

Load, save or delete a stored session

Saved Sessions

Default Settings Fisher

Close window on exit:  Always  Never  Only on clean exit

UF Information Technology www.fl.ufl.edu

## Linux Command Line

> \_

All the Best Linux Cheat Sheets

1. Linux Command Line

- Linux Reference Card - Great reference published on FOSSwire website
- OSLinuxLinuxManual - Great one page reference for the most popular Linux commands
- Linux Cheat Sheet - An incredibly exhaustive reference for all things Linux.
- Terminal's Linux Cheat Sheet - A great reference with live comparisons
- Terminal's Shortcuts - Cheat sheet for the most common terminal shortcuts
- More Terminal Shortcuts - More shortcuts for history and X

- Lots of online resources
  - Google: Linux cheat sheet
- Training sessions
  - May 21: The Linux/Unix Command Line - An Introduction
- User manuals for applications

UF Information Technology www.fl.ufl.edu

## 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 Information Technology www.fl.ufl.edu

## UF Research Computing

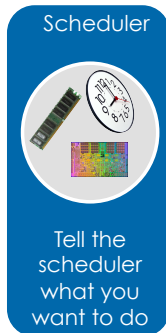
- Storage
  - Home:** /home/\$USER
    - For code compilation and user file management only
    - Do not use for job input/output!**
    - Include `cd $PBS_O_WORKDIR` or similar in scripts
  - Scratch space:** Lustre File System
    - /scratch/hpc/\$USER
    - 500GB per lab

Other storage options available for purchase

UF Information Technology www.fl.ufl.edu

## 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



## UF Research Computing

- ▶ Ordinary Shell Script

```
#!/bin/bash
date
module load test_app
test_app -i file.txt
```

Read the manual for your application

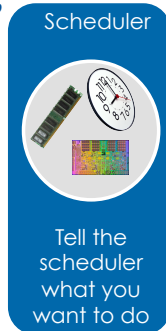
Commands typed on the command line can be put in a script

## UF Research Computing

- ▶ Submission Script

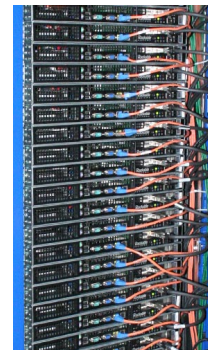
```
#!/bin/bash
#
#PBS -N My_Job_Name
#PBS -M Joe_Shmo@ufl.edu
#PBS -m abe
#PBS -o My_Job_Name.log
#PBS -j oe
#PBS -l nodes=1:ppn=1
#PBS -l walltime=00:05:00
#PBS -l pmem=900mb

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



## Nodes and processors

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



## RAM

```
#PBS -l pmem=900mb
```

- ▶ **Per-processor** RAM request
- ▶ Lots to consider, but do your best at estimating RAM needed for job
- ▶ Over about 3GB of RAM, "costs" toward CPU allocation

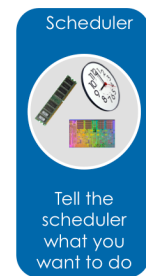
Wasted RAM leads to idle CPUs and low job throughput



## 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



## UF Research Computing

- ▶ Job Management
  - `qsub <file_name>`: job submission
  - `qstat -u <user>`: check queue status
  - `qdel <JOB_ID>`: job deletion
  - `qdelmine`: delete ALL of your current jobs

## UF Research Computing

- ▶ Job Scheduling and Usage
  - 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 few GB/core start counting toward the total PE value of a job
  - Test nodes (test01-06) available for interactive use, testing and short jobs
    - Connect from submit node:
 

```
[magitz@submit1 ~]$ssh test01
```

## Training Schedule

- ✓ Jan 14: Intro to UFHPC, getting started
- ▶ Jan 28: The Linux/Unix Shell - An Introduction
- ▶ Feb 4: Running Jobs, Submission Scripts, Modules
- ▶ Feb 11: Dr. Dhruva Chakravorty: Amber
- ▶ Feb 18: Galaxy Overview, The Basics
- ▶ Feb 25: Dr. David Ostrov: Molecular Docking
- ▶ Mar 11: NGS Data Techniques: General Methods and Tools
- ▶ Mar 18: NGS: Reference Based Mapping & de Novo Assembly
- ▶ Mar 25: Phylogenetic Analyses
- ▶ Apr 1: Multiprocessing at the HPC Center
- ▶ Apr 8: Introduction to GPU nodes
- ▶ Apr 15:
- ▶ Apr 22:

## 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](mailto:support@hpc.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

