

UF Research Computing: An Introduction and Getting Started

Matt Gitzendanner
maqitz@ufl.edu

8/29/13

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

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

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

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

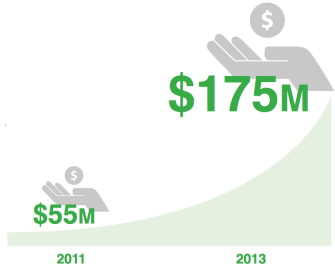


HiPerGator

The University of Florida Supercomputer for Research

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

Research Computing: Grants Supported



GRANTS SUPPORTED

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

Research Computing: Disciplines and Departments

25 **85**

2011 2013

DISCIPLINES AND DEPARTMENTS SUPPORTED

UF Information Technology www.it.ufl.edu

Internet2 Network

UF was 1st in the nation

- Internet2 Innovation Platform
 - 100 Gpbs connectivity
 - Campus Research Network now 200 Gbps

UF Information Technology www.it.ufl.edu

UF Data Center

- UF Data Center on Eastside Campus
 - 10,000 sq.ft and 1.75 MW total
 - 5,000 sq. ft. space for Research Computing

UF Information Technology www.it.ufl.edu

HiPerGator

The University of Florida Supercomputer for Research

- 16,384 cores
- Infiniband interconnect
- 2.1PB fast, high-availability, storage
- Current cluster will be integrated—Fall 2013

UF Information Technology www.it.ufl.edu

Research Computing

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

UF Information Technology www.it.ufl.edu

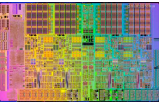
Research Computing

Where do you start?

UF Information Technology www.it.ufl.edu

Research Computing


- ▶ User Accounts
 - Qualifications:
 - Current UF faculty sponsor
 - Non-investors can run on 8 cores at a time
 - Investors gain priority access



www.hpc.ufl.edu

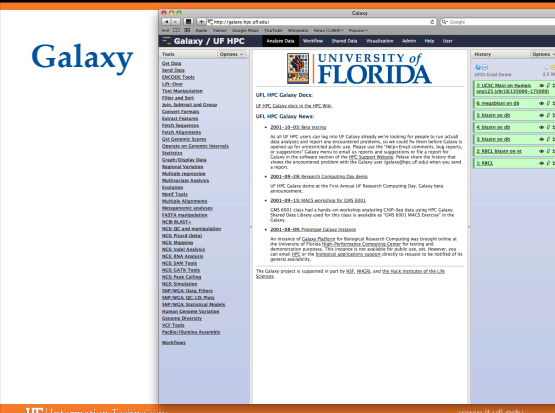
UF Information Technology www.ufl.edu

What can you run?

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



UF Information Technology www.ufl.edu



UF Information Technology www.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.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.ufl.edu



UF Information Technology www.ufl.edu

What can you run?



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

UF Information Technology www.it.ufl.edu



Wiki.hpc.ufl.edu

Magitz Talk Preferences Watchlist Contributions Log out

UF HPC Center
(Redirected from Main Page)

Welcome to the University of Florida Research Computing Wiki. The information here augments what is found on our web site with information we think is best provided in a less formal and official context. It is used for information that changes rapidly and might become quickly dated or incorrect on the web site. While there is good and helpful information here, you will find text that has not been proofed to the same standards that we try to maintain on our web site. We hope you find it helpful.

Getting Started [edit]

- Getting Started
- Most Command Summary
- Training
- Mailing Lists
- Changing your Password
- Samba Access

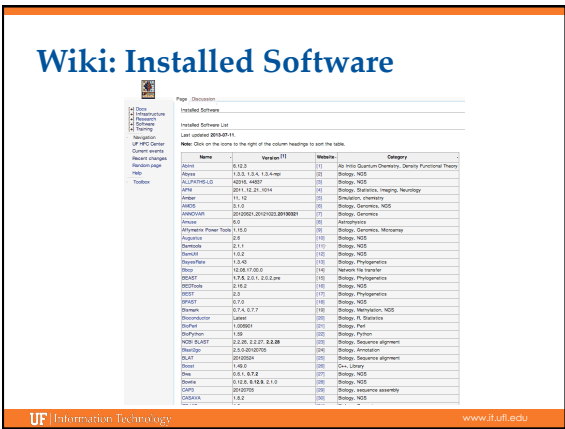
Software and Libraries [edit]

- Installed Software
- Modules
- Galaxy
- Installing Perl Modules
- GPUs and CUDA

Specific Research Areas [edit]

UF Information Technology www.it.ufl.edu

Wiki: Installed Software



Note: Click on the icon to the right of the column headings to sort the table.

Name	Version/URL	Website	Category
Abita	8.12.3		Biolog. NGS
Abseq	1.8.0, 1.9.0, 1.9.1, 2.0		Biolog. NGS
ALZRA/RELO	4200, 4480		Biolog. NGS
APRIS	2011, 12, 21, 1014		Biolog. Statistics, Imaging, Neurology
Armit	1.1, 1.2		Education, Metadata
ARCOS	3.1.0		Biolog. Semantics, NGS
ARVISE	201209-20121023, 20120928		Biolog. Semantics
Artemis	16.0		Microarrays
ArrayExpress Power Tools	1.0.0		Biolog. Semantics, Microarray
Argonavis	2.4		Biolog. NGS
Bamtools	2.4.1		Biolog. NGS
Barista	1.6.2		Biolog. NGS
Bayesian	1.4.0.0		Biolog. Pylogenetics
BioCr	14.08.17.02.0		Network file transfer
BRISAT	1.4.0, 2.0.1, 2.0.2, 2.0.3		Biolog. Pylogenetics
BRISATv2	3.0.0		Biolog. NGS
BSP	0.1.0		Biolog. Pylogenetics
BSPAT	0.1.2		Biolog. NGS
Bracken	0.15.0.1.7		Biolog. Metagenom, NGS
BrackenViewer	1.0000		Biolog. R, Statistics
Burpbat	1.03600		Biolog. Perl
BufPython	1.8		Biolog. Python
NS2-0.4.01	2.2.36, 2.2.37, 2.2.38		Biolog. Security agreement
BurpType	2.0.0, 20120706		Biolog. Annotation
BUST	201209		Biolog. Security agreement
Boost	1.49.0		C++ Library
Bow	0.12.0, 0.12		Biolog. NGS
Bowtie	0.12.2, 0.12.2, 2.1.0		Biolog. NGS
CATS	201209		Biolog. response assembly
CCSNAK	1.0.2		Biolog. NGS

UF Information Technology www.it.ufl.edu

Getting help: support.hpc.ufl.edu

HPC Support - Enter Request: HPC Cluster

Home | New | Browse | Search | Search | Reports | Preferences | Administration | Help

Before reporting a request, please read the [request writing guidelines](#), please look at the list of most frequently reported issues

Show Advanced Fields (* = Required Field)

Product: HPC Cluster Reporter: magitz@ufl.edu

Components:

- 01- Software
- 02- Hardware
- 03- Access
- 04- Performance
- 05- Batch System
- 06- Documentation
- 07- Question

Component Description: Select a component to read its description.

Version: All CMG-T2 HPC-Center Merz Not Applicable

Severity: Normal

Hardware: Compute Node OS: RHEL



Summary:

Description:

UF Information Technology www.it.ufl.edu

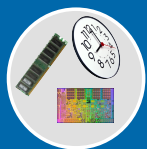
Cluster Basics

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

Cluster login


hipergator.hpc.ufl.edu


ssh

gator1
gator2


/home/\$USER

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

Windows: PuTTY 

Mac/Linux: Terminal 

User interaction



Login node (Head node)

UF Information Technology www.it.ufl.edu

Cluster login

User interaction

```

ssh hipergator.hpc.ufl.edu
gator1
Last login: Wed Aug 28 09:28:30 on tty001
it-rc-matt:~ matt$ ssh magitz@hipergator.hpc.ufl.edu
magitz@hipergator.hpc.ufl.edu's password:
Last login: Wed Aug 28 07:14:42 2013 from ip68-101-68-42.ga.at.cox.net

Welcome to the UF HPC Center.

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

http://wiki.hpc.ufl.edu/doc/Test_Nodes
UF HPC Center Account Policies can be found here:
http://www.hpc.ufl.edu/about/policies/account
[magitz@gator1 ~]$
    
```

Logging in

The screenshot shows the PuTTY Configuration window with the following settings:

- Category: Session
- Host Name (or IP address): submit.hpc.ufl.edu
- Port: 22
- Connection type: SSH
- Default Settings: Fisher
- Close window on exit: Only on clean exit

Linux Command Line

- ▶ Lots of online resources
 - Google: Linux cheat sheet
 - ▶ Training sessions
 - Sept 5: The Linux/Unix Command Line - An Introduction
 - ▶ User manuals for applications

Cluster Storage

- /home/\$USER: 10GB limit
 - Do not use for input or output for jobs
- /scratch/hpc/\$USER: 500GB limit
 - Older cluster, do not use for HiPerGator jobs
- /scratch/lfs/\$USER: 1TB limit
 - Primary storage for all HiPerGator jobs

SFTP-e.g.: FileZilla or CyberDuck

The screenshot shows the FileZilla interface with the local drive on the left and the remote site (scratch/lfs/magitz) on the right. The remote site contains directories like .ssh, .vim, .viminfo, .vimrc, .vimrc_bak, .vimrc_old, .vimrc_old2, .vimrc_old3, .vimrc_old4, .vimrc_old5, .vimrc_old6, .vimrc_old7, .vimrc_old8, .vimrc_old9, .vimrc_old10, .vimrc_old11, .vimrc_old12, .vimrc_old13, .vimrc_old14, .vimrc_old15, .vimrc_old16, .vimrc_old17, .vimrc_old18, .vimrc_old19, .vimrc_old20, .vimrc_old21, .vimrc_old22, .vimrc_old23, .vimrc_old24, .vimrc_old25, .vimrc_old26, .vimrc_old27, .vimrc_old28, .vimrc_old29, .vimrc_old30, .vimrc_old31, .vimrc_old32, .vimrc_old33, .vimrc_old34, .vimrc_old35, .vimrc_old36, .vimrc_old37, .vimrc_old38, .vimrc_old39, .vimrc_old40, .vimrc_old41, .vimrc_old42, .vimrc_old43, .vimrc_old44, .vimrc_old45, .vimrc_old46, .vimrc_old47, .vimrc_old48, .vimrc_old49, .vimrc_old50.

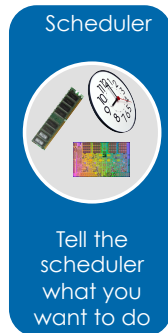
Research Computing

- ▶ 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
 - /scratch/lfs/\$USER

Other storage options available for purchase

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



Research Computing

- ▶ Ordinary Shell Script

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

Read the manual
for your application

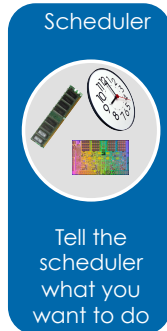
Commands typed
on the command
line can be put in a
script

Research Computing

- ▶ Submission Script

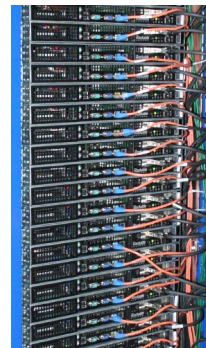
```
#!/bin/bash
#PBS -N My_Job_Name
#PBS -M Joe_Shmoefufl.edu
#PBS -m abe
#PBS -o My_Job_Name.log
#PBS -e My_job_Name.err
#PBS -l nodes=1:ppn=1
#PBS -l walltime=00:05:00
#PBS -l pmem=900mb

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=32
```



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



End-of-job emails: `#PBS -M Joe_Shmoefufl.edu` `#PBS -m abe`

```
PBS Job Id: 358634.moab.ufhpc
Job Name: NR.25.nex
Exec host: c7a-s1/60
Execution terminated
Exit_status=0
resources_used.cput=07:16:09
resources_used.mem=251348kb
resources_used.vmem=318916kb
resources_used.walltime=07:16:52
```

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

	Maximum	Short	Long
Investor	31 days	<12 hrs	7 days
Other	7 days	<12 hrs	3 days



Research Computing

- Job Management
 - qsub <file_name>: job submission
 - qstat -u <user>: check queue status
 - showq -r -u <user>: shows job efficiency
 - qdel <JOB_ID>: job deletion
 - checkjob -v <job number> (shows PE value)
 - pbs_info -f my_job.pbs (get job PE and group resources before submitting a job)

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 (dev1 and dev2) available for interactive use, testing and short jobs
 - Connect from login node:


```
[magitz@gator1 ~]$ssh dev1
```

Training Schedule

- ✓ Aug 29: Intro to UFHPC, getting started
- Sept 5: The Linux/Unix Shell - An Introduction
- Sept 12: HiPerGator: Running Jobs, Submission Scripts, Modules
- Sept 19: Galaxy Overview, The Basics
- Sept 26: NGS: Reference Based Mapping & de Novo Assembly

What topics would be most helpful for you??
Email: magitz@ufl.edu

UF Research Computing

- Help and Support
 - <https://support.hpc.ufl.edu>
 - For any kind of question or help requests
 - <http://wiki.hpc.ufl.edu>
 - Documents on hardware and software resources
 - Various user guides
 - Many sample submission scripts
 - <http://hpc.ufl.edu>
 - Frequently Asked Questions
 - Account set up and maintenance

