Introduction to Research Computing and HiPerGator

Matt Gitzendanner magitz@ufl.edu

HiPerGator
The University of Florida Supercomputer for Research

UF Information Technology
OneIT for the #GatorGood
• Research Computing Strategic Goals
  • Expand HPC, data storage and research network capacity, performance and usability
  • Enhance and expand services that use HPC, data storage and network resources
  • Improve faculty awareness and access to use of Research Computing services
Research Computing: Grants Supported
Additional Services

ResVault

GatorBox

UF Information Technology

OneIT for the #GatorGood
Research Computing

Where do you start?

UF Information Technology

OneIT for the #GatorGood
Research Computing

- User Accounts
  - Need current UF faculty sponsor
  - GatorLink ID
    - We can assist in creating GatorLink IDs for collaborators

www.rc.ufl.edu
Investor Supported

- Hardware **Purchase** (5-year lifespan)
  - Compute
  - Storage
- Computation & Data **Services** (annual)

Compute investment

- **Burst capacity (9X)**
- Lower priority access to idle resources
What can you run?

- Linux-based
- Generally command line driven applications
- GUI applications through Xpra
Welcome to the University of Florida Research Computing Help and Documentation site. The information here is focused on particular applications, services, and usage examples and complements more general policies and information found on our main web site. It is used for information that changes more frequently and might become quickly dated or incorrect on the web site. This site is edited by individual UFRC staff members. If you find inaccuracies, errors, or omissions on this site please let us know.

**HiPerGator 2.0 Information**

**Getting Started**

- Getting Started
- Mailing Lists
- Events_Calendar
- Training
- SLURM Commands
- Non-Batch System Resources
- Interactive Development and Testing
- GUI Programs
- Change your Password
- Open a Support Request

**Software and Libraries**
## Installed Software

### Installed Software List

Last updated 2014-05-06.

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Website</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abyss</td>
<td>1.3.4, 1.3.7</td>
<td>[2]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>ALLPATHS-LG</td>
<td>42315, 44837</td>
<td>[3]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>Amber</td>
<td>11, 12, 14</td>
<td>[5]</td>
<td>Simulation, chemistry</td>
</tr>
<tr>
<td>AMOS</td>
<td>3.1.0</td>
<td>[6]</td>
<td>Biology, Genomics, NGS</td>
</tr>
<tr>
<td>ANNOVAR</td>
<td>20120521, 20120523, 20130321</td>
<td>[7]</td>
<td>Biology, Genetics</td>
</tr>
<tr>
<td>Apwave</td>
<td>6.0</td>
<td>[8]</td>
<td>Astrophysics</td>
</tr>
<tr>
<td>Asymmetric Power Tools</td>
<td>1.15.0</td>
<td>[9]</td>
<td>Biology, Genetics, Microarray</td>
</tr>
<tr>
<td>Asymmetry</td>
<td>3</td>
<td>[10]</td>
<td>Material Science</td>
</tr>
<tr>
<td>Augustus</td>
<td>2.6</td>
<td>[11]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>Bamtools</td>
<td>1.0</td>
<td>[12]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>Barfoo</td>
<td>2.1.1</td>
<td>[13]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>BarnJII</td>
<td>1.0.2</td>
<td>[14]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>BayesAlign</td>
<td>1.3.43</td>
<td>[15]</td>
<td>Biology, Phylogenetics</td>
</tr>
<tr>
<td>Bicoc</td>
<td>12.08.17.00.0</td>
<td>[16]</td>
<td>Network file transfer</td>
</tr>
<tr>
<td>BEAGLE</td>
<td>3.0.2</td>
<td>[17]</td>
<td>Biology, Genetics, Genotyping</td>
</tr>
<tr>
<td>BEAST</td>
<td>1.7.5, 2.0.1, 2.0.2.pr</td>
<td>[18]</td>
<td>Biology, Phylogenetics</td>
</tr>
<tr>
<td>BEDTools</td>
<td>2.16.2, 2.17.0</td>
<td>[19]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>BEST</td>
<td>2.3</td>
<td>[20]</td>
<td>Biology, Phylogenetics</td>
</tr>
<tr>
<td>BETA</td>
<td>1.0.5</td>
<td>[21]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>BFAST</td>
<td>0.7.0</td>
<td>[22]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>Bismark</td>
<td>0.7.4, 0.7.7</td>
<td>[23]</td>
<td>Biology, Motivation, NGS</td>
</tr>
<tr>
<td>BioPurl</td>
<td>1.006091</td>
<td>[25]</td>
<td>Biology, Perl</td>
</tr>
<tr>
<td>BioPython</td>
<td>1.59</td>
<td>[26]</td>
<td>Biology, Python</td>
</tr>
<tr>
<td>BLAST</td>
<td>20130815</td>
<td>[27]</td>
<td>Biology, NGS</td>
</tr>
<tr>
<td>NCBI BLAST</td>
<td>2.2.27, 2.2.28, 2.2.29</td>
<td>[28]</td>
<td>Biology, Sequence alignment</td>
</tr>
<tr>
<td>Blast2go</td>
<td>2.5.0-20140705</td>
<td>[29]</td>
<td>Biology, Annotation</td>
</tr>
</tbody>
</table>
Getting help

support.rc.ufl.edu

Before reporting a request, please read the request writing guidelines, please look at the list of most frequently reported Requests, and please search for the request.

Show Advanced Fields

(* = Required Field)

* Summary:

Description:

Comment

Preview

Submit Request
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
Tools

ssh client to connect to hpg2.rc.ufl.edu

SFTP client to move files to/from your computer

Text editor to prepare files
Especially on Windows, be sure to convert DOS line breaks to Unix, and don’t use Word
Both have SFTP built in
**SSH Clients**

Mac/Linux: Terminal

Windows: MobaXterm or PuTTY

```
ssh user@hpg2.rc.ufl.edu
```
MobaXterm
Linux Command Line

• Lots of online resources
  • Google: Linux cheat sheet
• Training sessions
  • Jan 19: An Introduction to the Linux/Unix Command Line
• User manuals for applications
FileZilla

Host:
sftp.rc.ufl.edu

Do not use hpg2 login server for data transfer
Storage

- `/home/<user>`
  - 20GB limit
  - scripts, code, small data
  - Do NOT use for job input/output

- `/ufrc/<group>/ <user>`
  - 2TB limit per group
  - ALL input/output from jobs should go here

UF Information Technology

OneIT for the #GatorGood
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

OneIT for the #GatorGood
Cluster basics

Scheduler

Tell the scheduler what you want to do

Development servers

GUI servers

Compute servers

UF Information Technology

OneIT for the #GatorGood
Development servers

• Do not run applications on the login servers
  • Account will be suspended

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.

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@gator3 ~]$ Do not run interactive jobs on the login nodes.

• Request job in development partition, hpg2-dev
  • module load ufrc
  • srundev -t 120

UF Information Technology

OneIT for the #GatorGood
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
Basic SLURM job script

```bash
#!/bin/sh
#SBATCH --job-name=serial_job_test  # Job name
#SBATCH --mail-type=ALL             # Mail events
#SBATCH --mail-user=<email_address> # Where to send mail
#SBATCH --ntasks=1                  # Run on a single CPU
#SBATCH --mem=1gb                   # Memory limit
#SBATCH --time=00:05:00             # Time: hrs:min:sec
#SBATCH --output=serial_test_%j.out # Output and error log

pwd; hostname; date

module load python

echo "Running plot script on a single CPU core"
python /ufrc/data/training/SLURM/plot_template.py
date
```
SLURM CPU Requests

- Nodes: `--nodes` or `-N`
  - Request a certain number of physical servers
- Tasks: `--ntasks` or `-n`
  - Total number of tasks job will use
- CPUs per task: `--cpus-per-task` or `-c`
  - Number of CPUs per task

HiPerGator 2.0 Compute Servers:
- 32 cores (2 x 16-core Intel Xeon CPUs)
SLURM CPU Requests

- For single processor jobs
  
  ```bash
  #SBATCH --nodes=1
  #SBATCH --ntasks=1
  #SBATCH --cpus-per-task=1
  ```
SLURM CPU Requests

- Parallel applications
  - OpenMP, Threaded, Pthreads
    - All cores on one server, shared memory
  - MPI
    - Can use multiple servers
    - See: https://wiki.rc.ufl.edu/doc/Sample_SLURM_Scripts
SLURM CPU Requests

- For threaded applications (single node):
  
  ```bash
  #SBATCH --nodes=1
  #SBATCH --ntasks=1
  #SBATCH --cpus-per-task=8
  ```
SLURM Memory Requests

- `--mem=1gb` (total memory)
- `--mem-per-cpu=1gb` (memory per core)
  - Can use mb or gb
  - No decimal values: use 1500mb, not 1.5gb

HiPerGator 2.0 Compute Servers:
- 128 GB total RAM (vs 256 GB on HPG1)
- Diskless servers: OS takes ~8GB RAM
Emails

Job ID: 94392
Cluster: hipergator
User/Group: magitz/ufhpc
State: COMPLETED (exit code 0)
Nodes: 1
Cores per node: 4
CPU Utilization: 00:00:44
CPU Efficiency: 52.38% of 00:01:24 core-walltime
Memory Utilization 1.52 MB
Memory Efficiency: 0.04% of 4.00 GB
Emails

Job ID: 5019
Cluster: hpg1
User/Group: magitz/ufhpc
State: CANCELLED (exit code 0)
Cores: 1
CPU Utilization: 00:00:00
CPU Efficiency: 0.00% of 00:00:00 core-walltime
Memory Utilization 1.26 MB
Memory Efficiency: 126.17% of 1.00 MB

Job error file:

slurmstepd: Job 5019 exceeded memory limit (1292 > 1024), being killed
slurmstepd: Exceeded job memory limit
slurmstepd: *** JOB 5019 ON dev1 CANCELLED AT 2016-05-16T15:33:27 ***
SLURM Time Request

- **Time:** `--time` or `-t`
  - 120 (minutes)
  - 2:00:00 (hh:mm:ss)
  - 7-0 (days-hours)
  - 7-00:00 (days-hh:mm)
  - 7-00:00:00 (days-hh:mm:ss)
Quality of Service (--qos)

- Each group has two QoS options
  - Investment QOS: `--qos=group`
  - Burst QOS:
    - The burst capacity, available when idle resources are available on the cluster
    - `--qos=group-b`
- Users can choose higher priority, or larger pool of resources
SLURM output/error files

- `#SBATCH -o output.file`
- `#SBATCH -e error.file`
- `#SBATCH -o output.file #W/o -e combined`
- Can also use `--output` and `--error`

- `#SBATCH -o JobFile.%j.out`
  - Use `%j` instead of `$SLURM_JOBID`
SLURM

- Note that multi-letter directives are double-dash:
  - `--mail-type`
  - `--ntasks`
  - `--mem-per-cpu`

- Do not use spaces with `=`
  - `--mail-user=magitz@ufl.edu`
  - `--mail-user magitz@ufl.edu`
  - not: `--mail-user= magitz@ufl.edu`
SLURM environment

- SLURM inherits your environment
  - This includes present working directory
    - Don’t need cd $SLURM_SUBMIT_DIR

- Modules that are loaded

- Be careful of conflicting modules
Development sessions

- `module load ufrc`
- `srundev -t 60`
  - Can add:
    - `-n` or `-c` for multiple CPUs, `--mem` for more memory, etc.
Checking on jobs

- `squeue`
- `sacct`

- See [wiki.rc.ufl.edu/doc/SLURM_Commands](http://wiki.rc.ufl.edu/doc/SLURM_Commands)

- See [http://slurm.schedmd.com/](http://slurm.schedmd.com/)
Satisfaction Survey

training.it.ufl.edu
Support

- Support: support.rc.ufl.edu

- **Web page** and **wiki**

  HiPerGator 2.0 Information

  **Getting Started**
  - Getting Started
  - Mailing Lists
  - SLURM Commands
  - Non-Batch System Resources
  - Change your Password

  Information Technology

  OneIT for the #GatorGood