

Introducing



Matt Gitzendanner *magitz@ufl.edu*

UF | Information Technology

OneIT for the #GatorGood


HiPerGator





#GATORGOOD

UF | Information Technology


OneIT for the #GatorGood



HiPerGator
The University of Florida Supercomputer

- 30,000 cores
- #115 in Nov 2015 Top 500 
 - #2 among US public universities
 - HPL RMAX 738 TFlops
- Dell PowerEdge
- Intel Xeon E5-2698v3 2.3GHz CPUs 

UF | Information Technology OneIT for the #GatorGood



CHANGE AHEAD

- GatorLink Authentication
- SLURM Scheduler
- /ufrc unifies /scratch/lfs, /lts, /rlts

UF | Information Technology OneIT for the #GatorGood

Time line

Mar 1	Early access with reduced SLEs •Must convert username to GatorLink •/ufrc may be reformatted
Apr 15 – Jun 15	All users migrate
June 15	GatorLink authentication starts
July 1	Full production of HiPerGator with 51,000 cores and 3PB /ufrc storage

Early Access SLEs

- Service Level Expectations
 - System Stability
 - Jobs may be terminated prematurely
 - Interactive sessions may be interrupted
 - Maintenance any time with little/no warning
 - File systems
 - /ufrc may be reconfigured, erasing data
 - No Samba, ownCloud, or TSM backup

Early Access SLEs

- Service Level Expectations
 - Applications
 - Most will run faster
 - Cannot rebuild all 600
 - MPI applications likely problematic
 - We will deal with problems in order:
 - Existing application does not run
 - Evidence of performance increase by rebuild
 - Application used by large fraction of community
 - As time permits

Early Access

- Must change RC username to GatorLink username
 - If they are the same
 - No change until June 15
 - June 15, GatorLink authentication goes live
 - If they are different:
 - Username must be changed
 - User logged out with no running jobs on HPG1
 - Mar 1 – Jun 15, username will be GatorLink, password will be current password
 - June 15, GatorLink authentication goes live

Storage

- Transition from /scratch/lfs to /ufrc
- Unified primary storage for HiPerGator
 - Merging /lts, /rlts, /scratch/lfs
 - 3 PB final size
- Early access
 - Copy data to /ufrc
 - Globus Endpoint: ufrc#hpg2
 - SLE: /ufrc may be reformatted

Do not store the only copy of
critical data on /ufrc

UF | Information Technology

OneIT for the #GatorGood

Globus.org



- Fast transfer from /scratch/lfs/ to /ufrc
- **Do not transfer everything!**
 - Only have 1PB in /ufrc until merged file system
- ufrc#hpg2 and ufrc#go

The screenshot shows the Globus.org web interface for file transfers. At the top, there are navigation links: Manage Data, Publish, Groups, Support, and Account. Below that, there are tabs for Transfer Files, Activity, Endpoints, Bookmarks, and Console. The main content area is titled "Transfer Files" and includes a sub-header "Get Globus Connect Personal Turn your computer into an endpoint." and "RECENT ACTIVITY" with a filter icon.

Two transfer windows are visible:

- Left Window:** Endpoint: ufrc#hpg2, Path: //ufrc/ufhpc/magitz/
- Right Window:** Endpoint: ufrc#go, Path: //scratch/lfs/magitz/

Below the transfer windows, there are two file lists:

Endpoint: ufrc#hpg2	Endpoint: ufrc#go
<ul style="list-style-type: none"> select all up one folder refresh list <ul style="list-style-type: none"> CodingFun Folder boylan Folder cpFeatures Folder genefam Folder miRNA Folder 	<ul style="list-style-type: none"> select none up one folder refresh list <ul style="list-style-type: none"> 1kp Folder CodingFun Folder DoseResponse Folder GatorVault Folder Magitz Folder

Storage

- User folders organized by group:
 - /ufrc/primary_group/username
- Same group-based quotas as before
- Reduce storage space on /scratch/lfs
 - Reduce /scratch/lfs use to below 1PB
 - Migrate data to /ufrc
 - Upgrade and grow /ufrc to 3PB

SLURM

- Why change?
 - Functionality
 - Flexibility
 - Cost



Moab to SLURM

- Documentation
 - [PBS2Slurm Command Reference](#) wiki page
 - Other Wiki pages being developed

Basic SLURM job script

```
#!/bin/bash
#SBATCH --job-name=test      #A name for your job
#SBATCH -o job_%j.out       #Name output file
#SBATCH --mail-type=ALL     #What emails you want
#SBTACH --mail-user=<Email address> #Where
#SBATCH --ntasks=1         #Processors per node
#SBATCH --mem-per-cpu=100mb #Per processor memory
#SBATCH -t 00:01:00        #Walltime in hh:mm:ss
                           #or d-hh:mm:ss

# Change to this job's submit directory
cd $SLURM_SUBMIT_DIR

hostname
module load python
python -v
```

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
 - **--ntasks=1**
- For parallel jobs on a single node:
 - **--ntasks=8**

SLURM CPU Requests

- For MPI jobs
 - **--nodes=4**
 - **--ntasks=32**
 - Gets 32 cores on 4 nodes, but may be unbalanced, e.g.: 16, 8, 4 and 4
- For MPI jobs
 - **--nodes=4**
 - **--ntasks=4**
 - **--cpus-per-task=8**
 - Gets 32 cores on 4 nodes, 8 on each node

SLURM Memory Requests

- Memory: **--mem-per-cpu=1gb**
 - Can use mb or gb
 - Like Moab, no decimal values

HiPerGator 2.0 Compute Servers:

- 128 GB total RAM
- Diskless servers: OS takes ~8GB RAM

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)

SLURM output/error files

- `#SBATCH -o output.file`
- `#SBATCH -e error.file`
- `#SBATCH -o output.file` # Without -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 Task Arrays

- `#SBATCH --array=1-200%10`
- Similar to Moab: range with % to limit number of jobs at a time
- `$SLURM_ARRAY_TASK_ID`
- Output file naming:
 - %A: job id
 - %a: task id
 - `Output.%A_%a.out`

End of free usage

- HiPerGator
 - Up to 8-cores for free
- HiPerGator 2.0 and beyond
 - Research Computing has been told we can no longer offer any free access
 - Try-and-buy loans
 - 1-3 month loan of resources
 - Test the system
 - Verify needs are met
 - Become an investor

Support

- Please open support requests
- Check Wiki HPG2 section

