HPC at FIU

Introducing the Panther Cluster and IBM Platform LSF

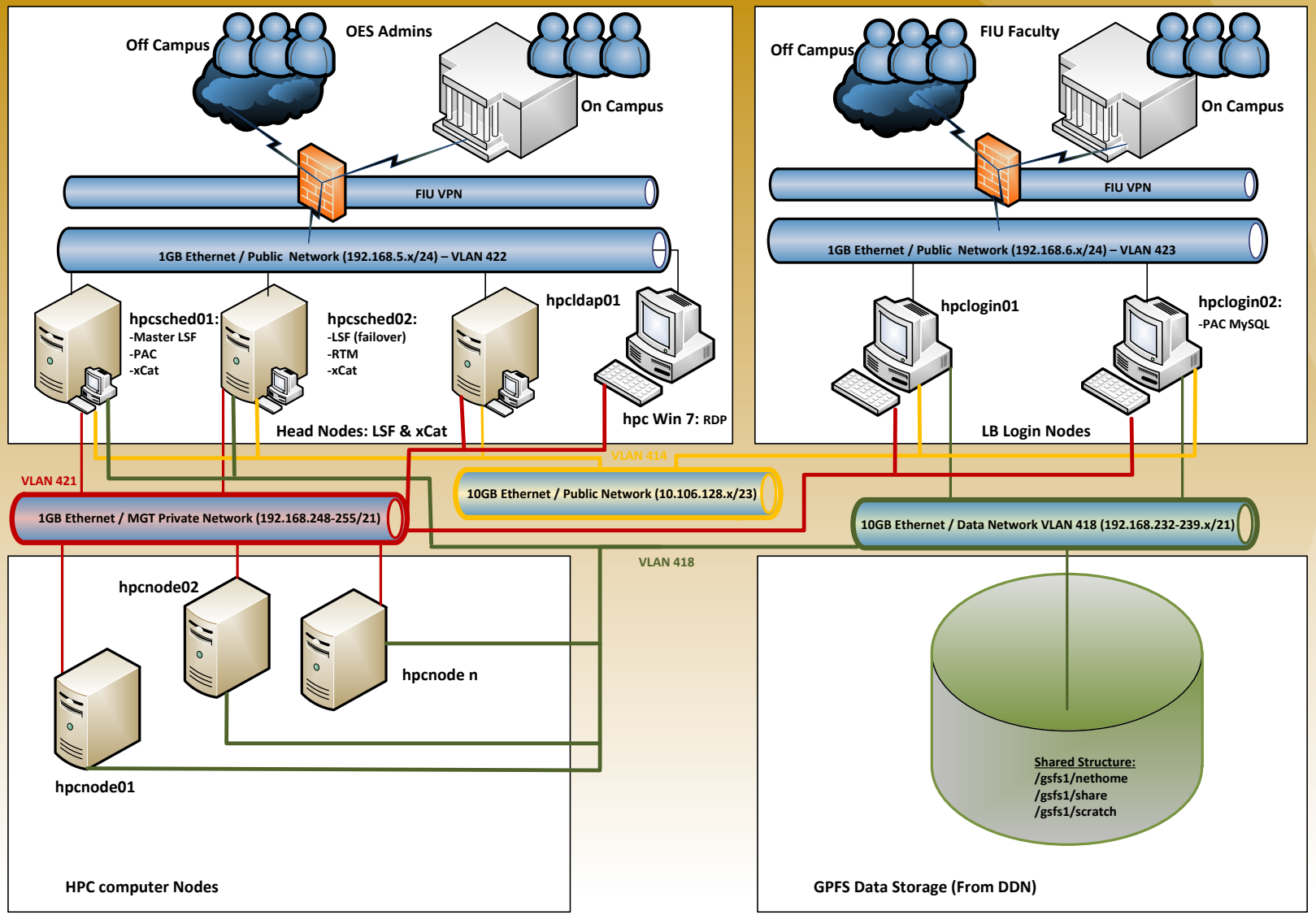
FIU | FLORIDA INTERNATIONAL UNIVERSITY

Introduction and Outline

- Why are you here?
  - Because you want to learn how to use the Panther Cluster
- What will we cover here today?
  - Introduction to the Panther Cluster
  - Basic commands of IBM Platform LSF
  - Submitting batch jobs for running on the cluster
  - Finding out information about cluster status
- What knowledge/information should you already have?
  - Basic computer skills
  - Command line skills will be useful

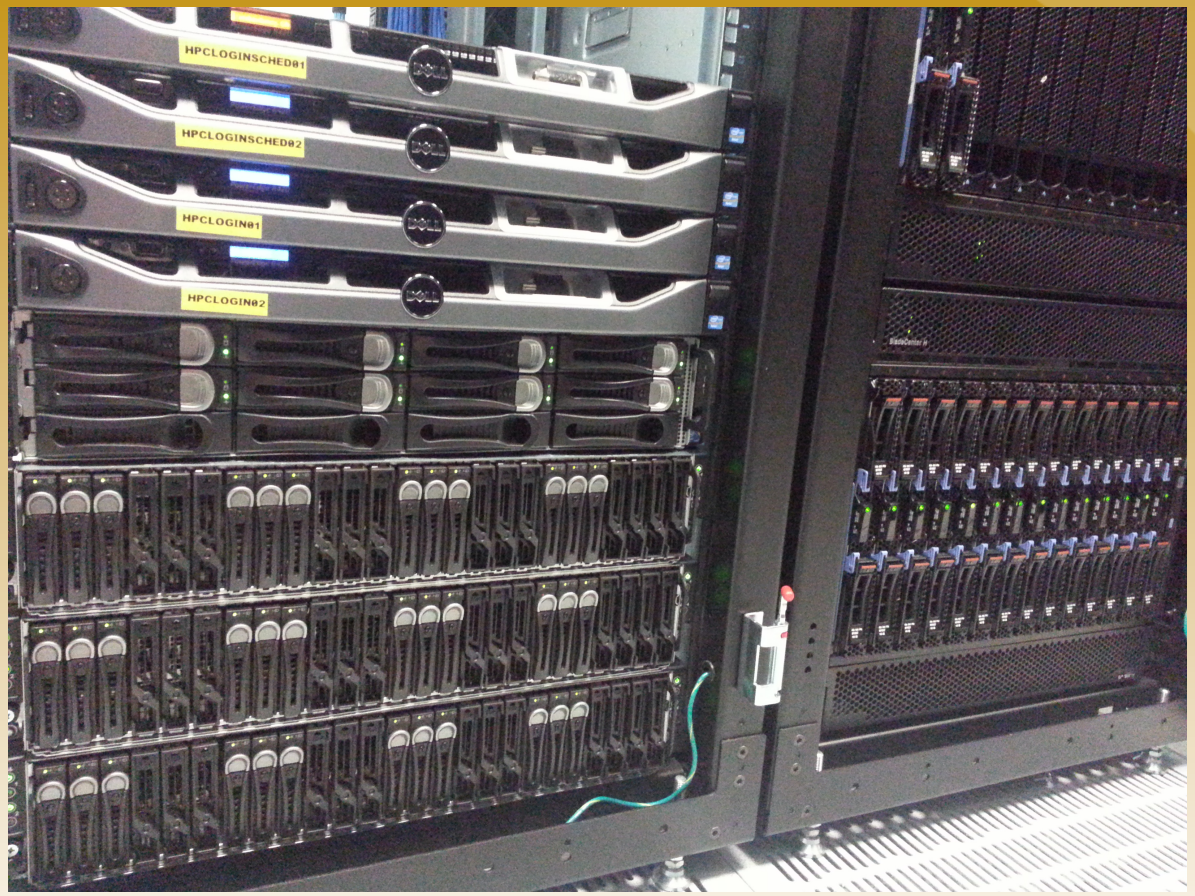
FIU | FLORIDA INTERNATIONAL UNIVERSITY

Panther Cluster Diagram



Panther Cluster Photograph

Coming soon:
224 cores
448 GB RAM



192 cores
1536 GB RAM

168 cores
896 GB RAM

Total: 584 cores, 2.8 TB RAM

FIU FLORIDA INTERNATIONAL UNIVERSITY

Logging In and Setting Up

- You must `ssh` to one of the login nodes
- There are two – hpclogin01 and hpclogin02
  - Which one should you choose?
- Linux or Mac terminal: `ssh username@hpclogin01`
- In Windows, use PuTTY or similar
- To login off-campus, you must first access the FIU VPN
  - This requires Cisco AnyConnect VPN client
  - http://vpninfo.fiu.edu/anyconnect/

FIU | FLORIDA INTERNATIONAL UNIVERSITY

File Transfers

- We send data to the cluster using scp (or rsync, sftp)
- In Windows, you can download and use WinSCP
- The command in a Linux/Mac terminal is

```
scp local_file_path username@hpclogin01:~/
```

- Does this file go to your home directory on hpclogin01?
- No: /home is in the parallel file system
- All nodes (compute and login) can now access this file

FIU | FLORIDA INTERNATIONAL UNIVERSITY

GPFS File System

- The compute nodes don’t necessarily have hard drives!
- Where is my data?
- Stored in /home mounted on logical volume in DDN
- Every node can see /home and /scratch through GPFS
- User directories in /home
- No user directories in /scratch – care required

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Modules

- xCAT installs only basic operating system on nodes
- Other programs installed in shared folder
- This includes simple programs like make, gcc, etc.
- Load these with `module` command

```
module load openmpi

module avail

module list
```

- You can add modules to your ~/.bashrc or your submission script

FIU | FLORIDA INTERNATIONAL UNIVERSITY

What is IBM Platform LSF?

- Time to submit a job with LSF. What is that?
- A suite of distributed resource management products
- LSF is an example of a “Scheduler”
- As the user, you submit a job to LSF
- LSF figures out how and when to run the job, based on the resources available to it
- LSF also gives you access to important system information, such as machine capabilities, other jobs being run, etc.

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Part 2: Basic Job Submission

- LSF enables “Batch Processing” on a cluster
- Users submit jobs to a queue
- LSF decides when and where to run each job
- To submit a job, use **`bsub`**
- Example: **`bsub sleep 10`**
- This runs the command **`sleep 10`** on a compute node
- How do we know it’s running? Use **`bjobs`**

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Input and Output in Batch

- Lets try using the command “hostname”
- Where is the output?
- Regular batch jobs are non-interactive
- There is no output to screen, input from keyboard, etc
- Use “-o” (-oo), “-e” (-eo) and “-i” `bsub` flags to get around this
- Examples: `bsub -o "out.txt" hostname`

```
bsub -o "outf" -i "inf" -e "errf" factor
```

FIU FLORIDA INTERNATIONAL UNIVERSITY

Submission Scripts

```
bsub -o "outf" -i "inf" -e "errf" factor
```

- This takes a long time to type, and we make mistakes
- We also forget things over time
- Solution: submission scripts
- Create a file – let's call it "subscript_factor"
- Option line looks like: `#BSUB -o "out.txt"`
- Final line is binary (executable, program): `factor`
- Submit: `bsub < subscript_factor`

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Array Jobs

- Scenario: You want to run a code that operates on a dataset and produces output. You have 500 datasets!
- Array jobs are designed for this situation. Script file:

```
#BSUB -J "jobname[1-500]%4"

#BSUB -o "output%I.txt"

executable $LSB_JOBINDEX
```

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Job Management

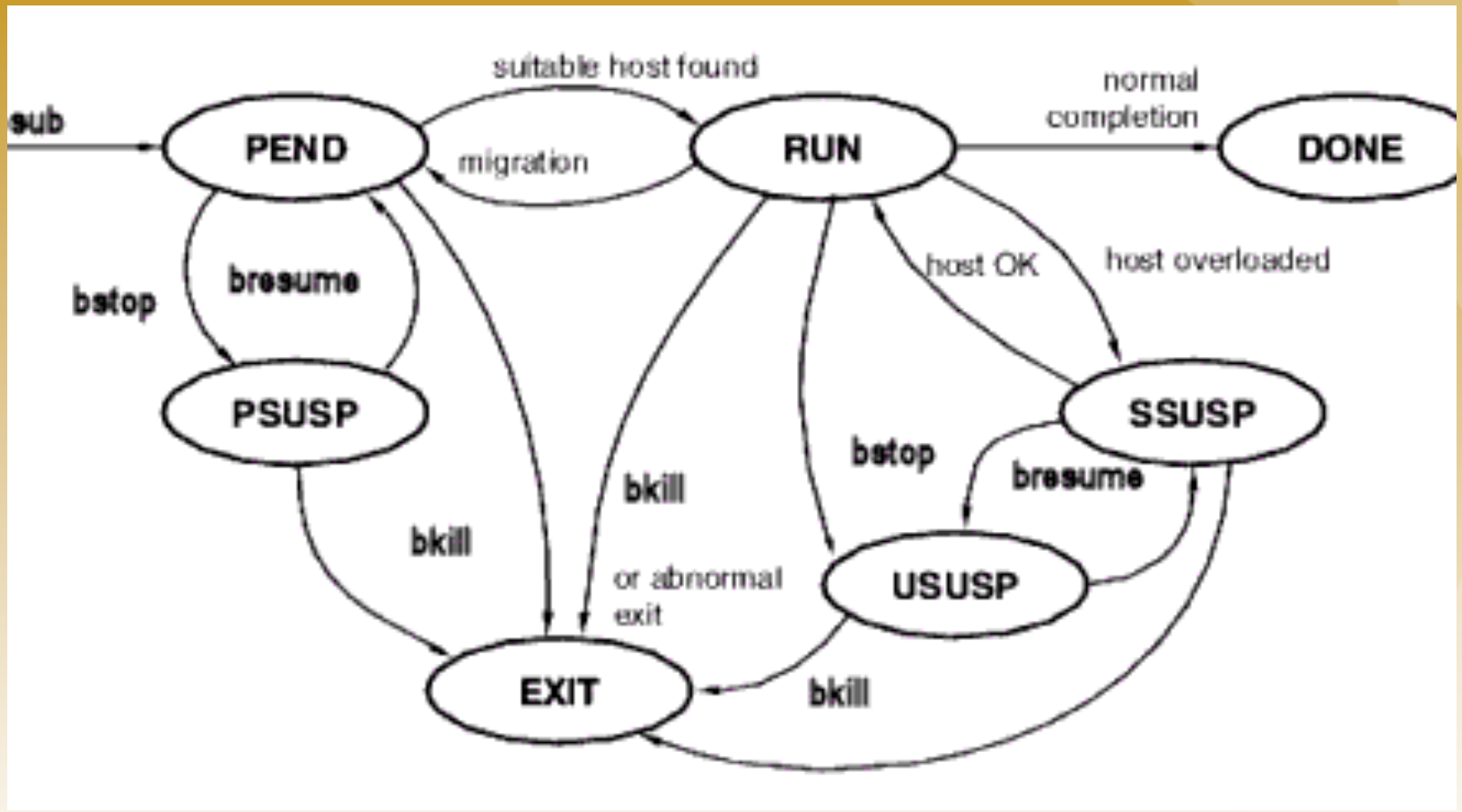
- You need to know how to see what jobs are running
- Also, how to kill, suspend and resume your jobs

Command	Function
`bjobs`	Show the state of all of your current jobs
`bkill job_ID`	Kill job with job number job_ID
`bstop job_ID`	Suspend job with job number job_ID
`bresume job_ID`	Resume job with job number job_ID

- Set job_ID = 0 to apply command to all of your jobs

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Job Workflow



FIU | FLORIDA INTERNATIONAL UNIVERSITY

Useful Options of `bjobs`

- Normally, `bjobs` displays only your running jobs
- Using flags makes this command more useful

Flag	Function
`-u all`	Display current job info for all users
`-u username`	Display current job info for user “username”
`-p job_ID`	Why is the job pending?
`-s job_ID`	Why is the job suspended?
`-a`	Display all jobs from past hour
`-l`	Display details job information

- `bhist` similar, but looks at older jobs

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Part 3: Advanced Submission

- This section is mostly about choosing resources
- Things like host, memory, number of processors
- Most of these are chosen through `bsub` flags
- Important exception: disk space
- Two types: home directory and scratch space

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Home Directory

- Landing point when logging into hpclogin01/02
- /home/your_username
- Every compute node can see this directory
- You will have a quota of ~100 GB, subject to change
- This data is not backed up – backing up is up to you
- (Service available for purchase from UTS)

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Scratch Space

- Another folder visible to every compute node
- /scratch
- A Logical Volume of high-speed drives is mounted here
- If your code reads and writes a lot, use /scratch
- No individual user directory – create one!
- Data persists for a week or two – no long term storage
- Again, no backup of this folder

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Queues

- All jobs submitted to the cluster go into a queue
- Add `#BSUB -q queue_name` to submission script
- Different queues have different resources available
- `bqueues` (`-l`) lists queues (with details)

Name	Properties	Priority
normal (default)	Any host, 32 slots max per job	30
short	Any host, 15 minutes	35
night	Any host, 7pm until 7am, 64 max	40

- List of queues subject to change

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Resource Flags

- Does your job have certain resource requirements?
- Specify them with the following #BSUB line:

```
#BSUB -R "command[string]"
```

command	function
`select`	Criteria for selecting hosts from the cluster
`order`	Sort hosts that meet above selection criteria
`rusage`	Specify expected resource consumption
`span`	Force parallel job to use multiple hosts
`same`	Force parallel job to run on same host type

- For possible string values, `lsinfo -r`

FIU FLORIDA INTERNATIONAL UNIVERSITY

select[string]

- Specify criteria for selecting hosts for your job
- `string` has many possible values, such as

string	value
`mem`	Amount of memory free
`ut`	Average 1-minute processor utilization
`ncpus`	Number of CPUs (cores) available
`hname`	Hostname

- Multiple resource requirements:
  - Use multiple instances of `-R`
  - Single instance of `-R` along with several commands
  - One command with `&&` or `||` operators

FIU | FLORIDA INTERNATIONAL UNIVERSITY

`select` examples and `order`

- Choose host with hostname n024

```
-R "select[hname=n024]"
```

- Choose host with >12 cores and >1 GB available memory

```
-R "select[ncpus>12 && mem>1024]"
```

- Choose 16 processor host with least cpu occupation

```
-R "select[ncpus==16] order[ut]"
```

- Reverse ordering: `order[-string]`
- `"select[string]" == "string"`

FIU | FLORIDA INTERNATIONAL UNIVERSITY

rusage[string]

- Reserve resources for your job
- Resources not reserved until your job starts
- Reservations performed on a per-slot basis
- Reserve 200 MB for each core on a host with 12 cores

```
-R "select[ncpus==12] rusage[mem=200]"
```

- Reserve an admin-created license resource for 4 cores

```
-R "rusage[soft_lic_1=1]"
```

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Resource Viewing

- We need info about resources and cluster status
- Various tools to find this:

Command	Function
lsid	Display cluster name
lshosts	Display basic info about hosts
lsload	Display host-level info about load
bhosts	Display batch-level info about load
lsinfo	Display info about various resources

Shared Memory Parallel Jobs

- Shared Memory Parallel Processing done with OpenMP
- GCC (or Intel) must be loaded before the job runs
- Submission script:

```
#BSUB -a openmp
#BSUB -n 8
. $MODULESHOME/../global/profile.modules
module load gcc/4.6.2
./executable
```

FIU FLORIDA INTERNATIONAL UNIVERSITY

Distributed Parallel Jobs

- Distributed parallel processing done with MPI (or PVM)
- Submission script:

```
#BSUB -a openmpi
#BSUB -n 32
. $MODULESHOME/../global/profile.modules
module load openmpi/1.6.3/intel
mpirun.lsf ./executable
```

- Other flavors: Intel-MPI, Platform-MPI

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Platform-MPI

- Platform-MPI is an alternative to OpenMPI
- Submission script:

```
#BSUB -n 32

. $MODULESHOME/../global/profile.modules

module load platform-mpi

mpirun -lsf ./executable
```

FIU FLORIDA INTERNATIONAL UNIVERSITY

`span` and `same`

- `span` specifies locality of parallel job
- `span[hosts=1]` - job executes on one host
- `span[ptile=2]` - no more than two cores per host
- `same` specifies that all hosts must have same resource
- `same[type]` - all hosts have same type (N/A)
- `same[ncpus]` - all hosts have same number of cpus

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Contact us !

hpcadmin@fiu.edu

FIU | FLORIDA INTERNATIONAL UNIVERSITY