High Performance Computing (HPC) on GACRC Sapelo2 Cluster

Georgia Advanced Computing Resource Center
EITS/University of Georgia
Zhuofei Hou, zhuofei@uga.edu
Outline

• GACRC

• What is High Performance Computing (HPC)?

• What is a Cluster?

• How to Work on Cluster?
GACRC

- A high-performance-computing (HPC) center at the UGA
- Provide to the UGA research and education community an advanced computing environment:
  - HPC computing and networking infrastructure located at the Boyd Data Center
  - Comprehensive collection of scientific, engineering and business applications
  - Consulting and training services

Wiki: [http://wiki.gacrc.uga.edu](http://wiki.gacrc.uga.edu)
Web Site: [http://gacrc.uga.edu](http://gacrc.uga.edu)
What is High Performance Computing?

Task Mountain

1 worker

10 years later......
What is High Performance Computing? (cont.)

2 months later......

JOB DONE
What is High Performance Computing? (cont.)

Serial Job with 1 worker

- Only 1 instruction executed at any time point on a single processor

Parallel Job with 6 workers

- 6 instructions executed *simultaneously* on 6 processors
- Shared-memory threaded parallel job or MPI distributed parallel job (Message Passing Interface)
What is a Cluster?

When you work on cluster, 3 roles are working:

- You: Who submit a job
- Queueing System: Who dispatch your job to run on cluster
- Cluster: Who run your job
Please Note: You need to connect to the UGA network using VPN when accessing from outside of the UGA main campus.

UGA VPN: https://eits.uga.edu/access_and_security/infosec/tools/vpn/
Two Nodes:

1. Login node for batch job workflow: MyID@sapelo2.gacrc.uga.edu
2. Transfer node for data transferring: MyID@xfer.gacrc.uga.edu

Five Directories:

1. **Home**: Login landing spot; 100GB quota; Backed-up
2. **Scratch**: High-speed storage for temp files needed for current jobs; NO quota; NOT backed-up
3. **Work**: High-speed storage for input files needed for repeated jobs; per group quota of 500GB and max 100,000 single files; NOT backed-up
4. **Project**: Temporary data parking; per group quota of 1TB; Backed-up (ONLY accessible from Transfer node!)
5. **Local Scratch**: Local storage on each individual compute node; 200GB quota; NOT backed-up

Four Computational Queues: batch, highmem_q, gpu_q, groupBuyin_q
Overview (cont.)

To connect to cluster, you need to have a user account:

1. Group PI request a user account for you: http://help.gacrc.uga.edu/userAcct.php
2. You need to pass new user training: https://wiki.gacrc.uga.edu/wiki/Training
3. We send you a welcome letter with detailed info about your Sapelo2 user account

Once your account is provisioned, use SSH Secure Shell program to open connection: https://wiki.gacrc.uga.edu/wiki/Connecting
How to work on cluster?

Your job working space is global scratch folder /scratch/MyID/

• Why?
  High-performance filesystem with the fastest I/O!

• How to access?
  From Login node, cd /scratch/MyID

• What do you need?
  Software! We installed ~1000 software on cluster for you. Check if the software you need is already installed at https://wiki.gacrc.uga.edu/wiki/Software

• What do you need to pay attention to?
  Clean up data you will not need after your job is finished!
1. Linux/Mac user: 
   `ssh MyID@sapelo2.gacrc.uga.edu`

2. `cd /scratch/MyID`

3. `mkdir ./workDir`

4. `cd ./workDir`

5. Linux/Mac user: 
   `scp file MyID@xfer.gacrc.uga.edu:/scratch/MyID/workDir`

6. `nano ./sub.sh`

7. `$ qsub sub.sh`

8. `$ qstat me or qdel JobID`
Useful Links

- GACRC Web: http://gacrc.uga.edu/
- GACRC Wiki: https://wiki.gacrc.uga.edu/wiki/Main_Page
- GACRC FAQ: https://wiki.gacrc.uga.edu/wiki/Sapelo2_Frequently_Asked_Questions
- GACRC Help: http://gacrc.uga.edu/help/
- GACRC Training: https://wiki.gacrc.uga.edu/wiki/Training
- GACRC User Account: https://wiki.gacrc.uga.edu/wiki/User_Accounts
- GACRC Software: https://wiki.gacrc.uga.edu/wiki/Software
Thank You!