High Performance Computing (HPC) on GACRC Sapelo2 Cluster

Georgia Advanced Computing Resource Center (GACRC)
Enterprise Information Technology Services (EITS)
The University of Georgia
Outline

• GACRC
• What is High Performance Computing (HPC)
• What is a Cluster
• 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
Help and Support: http://help.gacrc.uga.edu
Web Site: http://gacrc.uga.edu
Kaltura Channel: https://kaltura.uga.edu/channel/GACRC/176125031
What is High Performance Computing?

1 worker

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

6 months later......
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
Sapelo2 Cluster

1. ssh with MyID and password
2. Authenticate with ArchPass Duo

- **Cluster**: Nodes + Storage, all connected by network
- **Node**: Computer with a particular purpose within the cluster

**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/](https://eits.uga.edu/access_and_security/infosec/tools/vpn/)
Compute nodes are divided into groups called partitions. A partition is a collection of compute nodes for a particular computing need.
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 folder; 200GB quota; Backed-up
2. **Scratch**: Job working space; storing temp files needed for running jobs; NO quota; NOT backed-up
3. **Work**: Group sharing space, sharing files needed for running jobs; per group quota of 500GB; max 100,000 single files; NOT backed-up
4. **Project**: Temporary data (of active projects) parking; per group quota of 1TB; Backed-up
5. **Local Scratch**: Local storage in individual compute node; ~800GB quota; NOT backed-up

Four Computational Queues: batch, highmem_q, gpu_q, groupBuyin_q
Connect

➢ To connect to cluster, you need to get a Sapelo2 user account:

1. Your professor request a user account for you: https://uga.teamdynamix.com/TDClient/2060/Portal/Requests/ServiceDet?ID=25839
2. Pass necessary 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, you can access to Sapelo2 from your local computer:

https://wiki.gacrc.uga.edu/wiki/Connecting

➢ VPN is needed to connect to Sapelo2 from off-campus (outside the UGA main campus):

https://eits.uga.edu/access_and_security/infosec/tools/vpn/
Your job working space is your scratch folder /scratch/MyID/

- Why?
  High-performance filesystem with fast I/O!
- How to access?
  From login nodes or all compute nodes, use cd /scratch/MyID
- What do you need to pay attention to?
  Clean up your working space after your job completed!

What do you run in a computational job?

- Software! We installed ~1300 software on Sapelo2: [https://wiki.gacrc.uga.edu/wiki/Software](https://wiki.gacrc.uga.edu/wiki/Software)
- Your own programs or scripts (Python, Perl, Java, Fortran, C/C++, etc.)
1. Log on to Login node using MyID and password, and two-factor authentication with Archpass Duo: 
   
   `ssh MyID@sapelo2.gacrc.uga.edu`

2. On Login node, change directory to your `scratch` space: `cd /scratch/MyID`

3. Create a working subdirectory for a job: `mkdir workDir`

4. Change directory to `workDir`: `cd workDir`

5. Transfer data from local computer to `workDir`: use `Globus` to transfer data to the cluster  
   Transfer data on cluster to `workDir`: use `Globus` or log on to Transfer node and then use `cp` or `mv`

6. Make a job submission script in `workDir`: `nano sub.sh`

7. Submit a job from `workDir`: `sbatch sub.sh`

8. Check job status: `squeue --me` or Cancel a job: `scancel jobID`
GACRC Wiki http://wiki.gacrc.uga.edu
Kaltura channel https://kaltura.uga.edu/channel/GACRC/176125031

System: https://wiki.gacrc.uga.edu/wiki/Systems#Sapelo2
Connection: https://wiki.gacrc.uga.edu/wiki/Connecting#Connecting_to_Sapelo2
Software: https://wiki.gacrc.uga.edu/wiki/Software_on_Sapelo2
Running Jobs: https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_Sapelo2
Monitoring Jobs: https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_Sapelo2
Sample scripts: https://wiki.gacrc.uga.edu/wiki/Sample_batch_job_submission_scripts_on_Sapelo2
Transferring Files: https://wiki.gacrc.uga.edu/wiki/Globus
Linux Commands: https://wiki.gacrc.uga.edu/wiki/Command_List
Open OnDemand: https://wiki.gacrc.uga.edu/wiki/OnDemand
Training: https://wiki.gacrc.uga.edu/wiki/Training
Thank You!

Georgia Advanced Computing Resource Center
101-108 Computing Services building
University of Georgia
Athens, GA 30602
GACRC Help: http://help.gacrc.uga.edu