Introduction to GACRC Teaching Cluster

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

• GACRC
• Overview
• Working Environment
  ➢ Three Folders
  ➢ Three Computational Partitions
  ➢ Software on Cluster
• Submit a Computational Batch Job
• Run Interactive Jobs
• GACRC Wiki and Support
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)
Support: [https://wiki.gacrc.uga.edu/wiki/Getting_Help](https://wiki.gacrc.uga.edu/wiki/Getting_Help)
Web Site: [http://gacrc.uga.edu](http://gacrc.uga.edu)
Kaltura Channel: [https://kaltura.uga.edu/channel/GACRC/176125031](https://kaltura.uga.edu/channel/GACRC/176125031)
Note: You need to connect to the UGA VPN at first when accessing from outside of the UGA main campus.
Working Environment

https://wiki.gacrc.uga.edu/wiki/Systems#Teaching_cluster

- Two nodes, your "username" is your MyID for both of them:
  1. For batch job workflow, the host to log into is teach.gacrc.uga.edu
  2. For file transfers, the host to log into is txfer.gacrc.uga.edu

- Three Directories:
  1. /home/MyID: directory for static data (e.g., scripts, software, etc...)
  2. /scratch/MyID: working space for running computational jobs
  3. /work/CourseID: directory for course data
     a. /work/CourseID/MyID: data storage space for individual user in a class (e.g., /work/binf8211/MyID)
     b. /work/CourseID/instructor_data: data shared with class by the instructors

- Three Partitions:
  1. batch: for running regular computational jobs
  2. highmem: for running high-memory jobs
  3. gpu: for running GPU jobs
Working Environment (cont.)

- **Software**
  1. Software names are long and have an Easybuild toolchain name associated to it.
  2. Complete module name: Name/Version-toolchain, e.g., Python/3.8.2-GCCcore-8.3.0.
  3. Software names are case-sensitive!
     - `module spider pattern`: Search modules using a name pattern (case-insensitive).
     - `module load/unload moduleName`: Load/remove a module.
     - `module avail`: List all available modules on the cluster.
     - `module list`: List modules currently loaded.
     - `module purge`: Remove all modules from working environment.
Submit a Batch Job

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

1. Log on to Login node using MyID and password, and two-factor authentication with Archpass Duo:
   ```bash
   ssh MyID@teach.gacrc.uga.edu
   ```
2. Change directory to /scratch directory:
   ```bash
   cd /scratch/MyID
   ```
3. Create a working subdirectory for a job:
   ```bash
   mkdir workDir
   ```
4. Change directory to workDir:
   ```bash
   cd workDir
   ```
5. Transfer data from local computer to workDir:
   Use `scp` or `WinSCP` to connect Transfer node Transfer data on cluster to workDir:
   Log on to Transfer node and then use `cp` or `mv`
6. Make a job submission script in workDir:
   ```bash
   nano sub.sh
   ```
7. Submit a job from workDir:
   ```bash
   sbatch sub.sh
   ```
8. Check job status:
   ```bash
   sq --me or Cancel a job: scancel JobID
   ```
Step 1: Log on to Login node

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

1. Teaching cluster access requires verification using two-factor authentication with Archpass Duo. If you are not enrolled in Archpass Duo, please refer to

https://eits.uga.edu/access_and_security/infosec/tools/archpass_duo/ on how to enroll

2. If you are connecting from off-campus, please first connect to the UGA VPN and then connect to teach.gacrc.uga.edu. Information on how to use the VPN is available at https://eits.uga.edu/access_and_security/infosec/tools/vpn/
Step 1: Log on to Login node - Mac/Linux using ssh

1. Open **Terminal** utility

2. Type command line: `ssh MyID@teach.gacrc.uga.edu`

3. You will be prompted for your **UGA MyID password**

4. You will verify your login using **Archpass Duo** authentication
ssh zhuofei@teach.gacrc.uga.edu  ➔ 1. use ssh to open connection

UGA DUO authentication is required for SSH/SCP access to GACRC systems. For additional help with UGA DUO authentication or to report an issue please visit: https://eits.uga.edu/access_and_security...

Password:  ➔ 2. Enter your MyID password
When you enter password, no stars or dots will show as you are typing. Please type password carefully!

Duo two-factor login for zhuofei

Enter a passcode or select one of the following options:

1. Duo Push to XXX-XXX-5758
2. Phone call to XXX-XXX-5758
3. Phone call to XXX-XXX-1925
4 5. SMS passcodes to XXX-XXX-5758 (next code starts with: 1)

Passcode or option (1-5): 1  ➔ 3. Select Duo option
Success. Logging you in...
Last login: Mon Aug  3 11:11:58 2020 from 172.18.114.119
zhuofei@teach-sub1 ~$  ➔ 4. Logged on!

5. Verify login using Duo
Step1 (Cont.) - Windows using PuTTY

1. Download and install PuTTY: https://www.putty.org/

2. Detailed downloading and installation instructions:
   
   https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY

3. Detailed configuring and usage instructions:
   
   https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY#Configuring_PuTTY
Step1 (Cont.) - Windows using PuTTY

The first time you connect to login node, PuTTY will give you this security alert window. Please click "Yes".
Next you will enter your UGA MyID password and initiate DUO authentication procedure:

![Command Line Interface with UGA MyID password, Select DUO option, and Logged on message]

- **UGA MyID password**
- **Select DUO option**
- **Logged on!**
Step2 - 4: cd to /scratch dir, make and cd into workDir

```
zhuofei@teach-sub1 ~$ cd /scratch/zhuofei
zhuofei@teach-sub1 zhuofei$ mkdir workDir
zhuofei@teach-sub1 zhuofei$ cd workDir/
zhuofei@teach-sub1 workDir$ ls
zhuofei@teach-sub1 workDir$
```

- `cd` command to change directory
- `mkdir` command to create a subdirectory
- `cd` command to change directory
- `ls` command to list contents of directory
- it is empty in workDir!
Step 5: Transfer data from local computer to workDir - Mac/Linux

https://wiki.gacrc.uga.edu/wiki/Transferring_Files#Using_scp_2

1. Connect to Transfer node (txfer.gacrc.uga.edu) in Terminal from your local computer
2. Use **scp** command: scp (-r) [Source] [Target]
3. Enter your MyID password, then select Duo option to verify connection

*E.g. 1:* use scp on local computer, from Local ➔ workDir on cluster

```
scp ./file zhuofei@txfer.gacrc.uga.edu:/home/zhuofei/workDir
```
```
scp -r ./folder/ zhuofei@txfer.gacrc.uga.edu:/home/zhuofei/workDir
```

*E.g. 2:* use scp on local computer, from workDir on cluster ➔ Local

```
scp zhuofei@txfer.gacrc.uga.edu:/home/zhuofei/workDir/file .
```
```
scp -r zhuofei@txfer.gacrc.uga.edu:/home/zhuofei/workDir/folder/ .
```
Step 5 (Cont.) - Windows using WinSCP

https://wiki.gacrc.uga.edu/wiki/Transferring_Files#Using_WinSCP_2

1. You need to connect to cluster’s Transfer node (txfer.gacrc.uga.edu)

2. Use WinSCP on local computer
   - WinSCP can be downloaded from https://winscp.net/eng/index.php
   - Default installation procedure is simple

Step 5 (Cont.) - Windows using WinSCP
Step 5 (Cont.) - Windows using WinSCP

Select DUO option
Step 5 (Cont.) - Windows using WinSCP

Change paths on your local computer and transfer node

Drag to transfer files or folders
Step 5 (Cont.): Transfer data on cluster to workDir

- Log on to Transfer node (txfer.gacrc.uga.edu)
  - Mac/Linux: ssh MyID@txfer.gacrc.uga.edu (page 9-10)
  - Windows: use PuTTY to log in MyID@txfer.gacrc.uga.edu (page 11-13)

- Directories you can access on transfer node:
  1. /home/MyID
  2. /scratch/MyID
  3. /work/CourseID/

- Transfer data between two folders on cluster using `cp` or `mv`, e.g.:
  ```
  mv /work/binf8211/MyID/datafile /scratch/MyID/workDir
  ```
Step 6: Make a job submission script in workDir using nano

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

$ nano sub.sh

nano is a simple text editor on Linux. You are welcome to use other editors like vim or emacs.

Ctrl-x to save file and quit from nano
Step 6 (Cont.)

1. Copy sample job to workDir:
   
   `cp -r /usr/local/training/Teach/* .`

2. Job submission script:
   
   `sub.sh`

3. Running Jobs:

   [https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster](https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster)

```bash
#!/bin/bash

#SBATCH --job-name=testBowtie2       # Job name (testBowtie2)
#SBATCH --partition=batch            # Partition name (batch, highmem_p, or gpu_p)
#SBATCH --ntasks=1                   # 1 task (process) for below commands
#SBATCH --cpus-per-task=1            # CPU core count per task, by default 1
#SBATCH --mem=4G                     # Memory per node (4GB); by default using M as unit
#SBATCH --time=1:00:00               # Time limit hrs:mins:secs or days-hrs:mins:secs
#SBATCH --output=%x_%j.out            # Standard output log, e.g., testBowtie2_12345.out
#SBATCH --mail-user=username@uga.edu # Where to send mail
#SBATCH --mail-type=END,FAIL          # Mail events (BEGIN, END, FAIL, ALL)

ml Bowtie2/2.4.1-GCC-8.3.0 # Load software module and run bowtie2 below

bowtie2 -x index/lambda_virus -U myreads.fq -S output.sam
```
Step 7: Submit a job from workDir using sbatch

https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster#How_to_submit_a_job_to_the_batch_queue

$ sbatch sub.sh
Submitted batch job 33791

**Tips:** sub.sh is a job submission script for

1. specifying computing resources
2. loading software using `module load`
3. running any Linux commands you want to run
4. running the bowtie2 commands
Step 8: Check job status using `sq --me`

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

```bash
$ sq --me

<table>
<thead>
<tr>
<th>JOBID</th>
<th>NAME</th>
<th>PARTITION</th>
<th>USER</th>
<th>NODES</th>
<th>CPUS</th>
<th>MIN_MEMORY</th>
<th>PRIORITY</th>
<th>TIME</th>
<th>TIME_LIMIT</th>
<th>STATE</th>
<th>NODELIST(REASON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33791</td>
<td>testBowtie2</td>
<td>batch</td>
<td>zhuofei</td>
<td>1</td>
<td>1</td>
<td>4G</td>
<td>1</td>
<td>3:00</td>
<td>1:00:00</td>
<td>RUNNING</td>
<td>tcn18</td>
</tr>
</tbody>
</table>
```
Step 8 (Cont.): Cancel job using scancel

https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster#How_to_delete_a_running_or_pending_job

$ sq --me

<table>
<thead>
<tr>
<th>JOBID</th>
<th>NAME</th>
<th>PARTITION</th>
<th>USER</th>
<th>NODES</th>
<th>CPUS</th>
<th>MIN_MEMORY</th>
<th>PRIORITY</th>
<th>TIME</th>
<th>TIME_LIMIT</th>
<th>STATE</th>
<th>NODELIST(REASON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33791</td>
<td>testBowtie2</td>
<td>batch</td>
<td>zhuofei</td>
<td>1</td>
<td>1</td>
<td>4G</td>
<td>1</td>
<td>0:23</td>
<td>1:00:00</td>
<td>RUNNING</td>
<td>tcn18</td>
</tr>
</tbody>
</table>

$ scancel 33791

$ sq --me

$ sq --me
Step 8 (Cont.): Check job details using `sacct-gacrc -X` and `seff`  
https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_the_teaching_cluster

```
$ sacct-gacrc -X

<table>
<thead>
<tr>
<th>JobID</th>
<th>JobName</th>
<th>User</th>
<th>Partition</th>
<th>NNode</th>
<th>NCPUS</th>
<th>ReqMem</th>
<th>CPUTime</th>
<th>Elapsed</th>
<th>Timelimit</th>
<th>State</th>
<th>ExitCode</th>
<th>NodeList</th>
</tr>
</thead>
<tbody>
<tr>
<td>33791</td>
<td>testBowtie2</td>
<td>zhuofei</td>
<td>batch</td>
<td>1</td>
<td>1</td>
<td>4Gn</td>
<td>00:00:29</td>
<td>00:00:29</td>
<td>01:00:00</td>
<td>CANCELLED+</td>
<td>0:0</td>
<td>tcn18</td>
</tr>
</tbody>
</table>
```

```
$ seff 33791  # Check computing resources used by a COMPLETED job

Job ID: 33791
Cluster: gacrc-teach
User/Group: zhuofei/ gacrc-instruction
State: CANCELLED (exit code 0)
Cores: 1
CPU Utilized: 00:00:00
CPU Efficiency: 0.00% of 00:00:29 core-walltime
Job Wall-clock time: 00:00:29
Memory Utilized: 1.25 MB
Memory Efficiency: 0.03% of 4.00 GB
```
Step 8 (Cont.): Check node info using sinfo

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

<table>
<thead>
<tr>
<th>PARTITION</th>
<th>AVAIL</th>
<th>TIMELIMIT</th>
<th>NODES</th>
<th>STATE</th>
<th>NODELIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch*</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>1</td>
<td>drain tcn17</td>
</tr>
<tr>
<td>batch*</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>4</td>
<td>mix tcn[18-21]</td>
</tr>
<tr>
<td>batch*</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>20</td>
<td>idle tcn[1-16,22-25]</td>
</tr>
<tr>
<td>interactive</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>1</td>
<td>mix tcn26</td>
</tr>
<tr>
<td>interactive</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>4</td>
<td>idle tcn[27-30]</td>
</tr>
<tr>
<td>gpu</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>1</td>
<td>idle tcgn1</td>
</tr>
<tr>
<td>highmem</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>1</td>
<td>idle tchmn2</td>
</tr>
<tr>
<td>highmem</td>
<td>up</td>
<td>7-00:00:00</td>
<td></td>
<td>1</td>
<td>down tchmn1</td>
</tr>
<tr>
<td>fsr4601</td>
<td>up</td>
<td>1:00</td>
<td></td>
<td>1</td>
<td>drain tcn17</td>
</tr>
<tr>
<td>fsr4601</td>
<td>up</td>
<td>1:00</td>
<td></td>
<td>4</td>
<td>mix tcn[18-21]</td>
</tr>
<tr>
<td>fsr4601</td>
<td>up</td>
<td>1:00</td>
<td></td>
<td>20</td>
<td>idle tcn[1-16,22-25]</td>
</tr>
<tr>
<td>fsr8602</td>
<td>up</td>
<td>10:00</td>
<td></td>
<td>1</td>
<td>drain tcn17</td>
</tr>
<tr>
<td>fsr8602</td>
<td>up</td>
<td>10:00</td>
<td></td>
<td>4</td>
<td>mix tcn[18-21]</td>
</tr>
<tr>
<td>fsr8602</td>
<td>up</td>
<td>10:00</td>
<td></td>
<td>20</td>
<td>idle tcn[1-16,22-25]</td>
</tr>
</tbody>
</table>
Obtain Job Details

https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster#How_to_check_resource_utilization_of_a_running_or_finished_job

Option 1: `squeue --me -l` for details of a running or pending jobs

Option 2: `sacct-gacrc -X` for details of computing resource usage of a running or finished job

Option 3: `seff` for details of computing resource usage of a finished job

Option 4: Email notification from finished jobs (completed, canceled, or crashed), if using:

```
#SBATCH --mail-user=username@uga.edu
```

```
#SBATCH --mail-type=ALL
```
## Run Interactive Jobs

https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_Sapelo2#How_to_open_an_interactive_session

https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_Sapelo2#How_to_run_an_interactive_job_with_Graphical_User_Interface_capabilities

<table>
<thead>
<tr>
<th>Description</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start an interactive session</td>
<td>interact</td>
</tr>
<tr>
<td>Start an interactive session with X forwarding</td>
<td>interact --x11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>interact</th>
<th>srun --pty --cpus-per-task=1 --job-name=interact --ntasks=1 --nodes=1 --partition=inter_p --time=12:00:00 --mem=2GB /bin/bash -l</th>
</tr>
</thead>
<tbody>
<tr>
<td>interact --x11</td>
<td>srun --pty --cpus-per-task=1 --job-name=interact --ntasks=1 --nodes=1 --partition=inter_p --time=12:00:00 --mem=2GB --x11 /bin/bash -l</td>
</tr>
</tbody>
</table>
Connecting: https://wiki.gacrc.uga.edu/wiki/Connecting#Connecting_to_the_teaching_cluster

Running Jobs: https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster

Monitoring Jobs: https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_the_teaching_cluster

Transfer File: https://wiki.gacrc.uga.edu/wiki/Transferring_Files#The_File_Transfer_node_for_the_teaching_cluster.28txfer.gacrc.uga.edu.29

Sample Job Scripts: https://wiki.gacrc.uga.edu/wiki/Sample_batch_job_submission_scripts_on_the_teaching_cluster

Linux Command: https://wiki.gacrc.uga.edu/wiki/Command_List
GACRC Support
https://wiki.gacrc.uga.edu/wiki/Getting_Help

➤ **Job Troubleshooting:**

Please tell us details of your question or problem, including but not limited to:

- Your user name
- Your job ID
- Your working directory
- The partition name and command you used to submit the job

➤ **Software Installation:**

- Specific name and version of the software
- Download website
- Supporting package information if have

Please note to make sure the correctness of datasets being used by your jobs!
Georgia Advanced Computing Resource Center (GACRC) service catalog.

If you would like to reach out to GACRC and do not have a UGA MyID, please send an email to gacrc-help@uga.edu, and we will respond promptly.

Categories (3)

**Services For Users**
General user support, request software installation or update, request training.

**Services for PIs**
For PIs only: Lab registration, user account creation/modification, class account requests, storage quota modifications.

**For GACRC Staff**
For GACRC’s internal use only.
Need Support?  http://help.gacrc.uga.edu