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

• 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
Support: https://wiki.gacrc.uga.edu/wiki/Getting_Help
Web Site: http://gacrc.uga.edu
Kaltura Channel: 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`: working space for running computational jobs
  2. `/work/CourseID/MyID`: data storing space for individual user in a class (e.g., `/work/binf8940/MyID`)
  3. `/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 a 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 module names matching a pattern (case-insensitive)
- `module load moduleName`: Load a module into your working environment
- `module avail`: List all available software modules installed on cluster
- `module list`: List modules currently loaded
- `module unload moduleName`: Remove a module 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:
   
   ssh MyID@teach.gacrc.uga.edu

2. Create a working subdirectory for a job: `mkdir ./workDir`

3. Change directory to `workDir`: `cd ./workDir`

4. 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`

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

6. Submit a job from `workDir`: `sbatch ./sub.sh`

7. Check job status: `squeue` 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. 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!
Step1 (Cont.) - Windows using PuTTY

1. Download and install PuTTY: [https://www.putty.org/](https://www.putty.org/)

2. Detailed downloading and installation instructions:
   
   [https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY](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](https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY#Configuring_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:
Step 2 - 3: Create and change directory to workDir

```bash
zhuofei@teach-sub1 ~$ ls

zhuofei@teach-sub1 ~$ mkdir workDir

zhuofei@teach-sub1 ~$ ls

workDir

zhuofei@teach-sub1 ~$ cd workDir/

zhuofei@teach-sub1 ~$ ls

zhuofei@teach-sub1 ~$
```

- `ls` command to list folder’s contents
- `mkdir` command to create a subdirectory
- `cd` command to change directory
- It is empty in workDir!
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

Step5 (Cont.) - Windows using WinSCP
https://wiki.gacrc.uga.edu/wiki/Transferring_Files#Using_WinSCP_2
Step 5 (Cont.) - Windows using WinSCP

Select DUO option
Step5 (Cont.) - Windows using WinSCP

Change paths on your local computer and transfer node

Drag to transfer files or folders
Step 4 (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 (Landing home)
  2. /work/CourseID/MyID
  3. /work/CourseID/instructor_data

- Transfer data between two folders on cluster using `cp` or `mv`, e.g.:
  
  ```
  mv /work/binf8940/MyID/datafile /home/MyID/workDir
  ```
Step 5: 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 5 (Cont.)

Copy
1. sample input data
c2. job submission script
to your current working folder:
cp /usr/local/training/sample.fasta .
cp /usr/local/training/sub_blast.sh .

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

```bash
#!/bin/bash
#SBATCH --job-name=testBLAST  # Job name
#SBATCH --partition=batch        # Partition (queue) name
#SBATCH --ntasks=1               # Single task job
#SBATCH --cpus-per-task=4        # Number of cores per task
#SBATCH --mem=20gb               # Total memory for job
#SBATCH --time=2:00:00            # Time limit hrs:min:sec
#SBATCH --output=log.%j          # Standard output and error log
#SBATCH --mail-user=MyID@uga.edu  # Where to send mail
#SBATCH --mail-type=END,FAIL      # Mail events (BEGIN, END, FAIL, ALL)

cd $SLURM_SUBMIT_DIR
module load BLAST+/2.9.0-gompi-2019b

time blastn -num_threads 4 -query sample.fasta -db /db/ncbiblast/nt/06042020/nt \
   -out results.${SLURM_JOB_ID} -outfmt 6 -max_target_seqs=2
```
Step 6: 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_blast.sh
Submitted batch job 139
```

**Tips:** `sub_blast.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 blast commands
Step 7: Check job status using `squeue`

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

```
$ squeue -l
Wed Aug  8 13:40:02 2018
JOBID PARTITION   NAME      USER   STATE   TIME   TIME_LIMI NODES NODELIST
162   batch   testBLAS  zhuofei PENDING  0:00   2:00:00       1 (None)
160   batch   testBLAS  zhuofei  RUNNING  0:02   2:00:00       1 c2-11
161   batch   testBLAS  zhuofei  RUNNING  0:02   2:00:00       1 c2-11
```

```
$ squeue
JOBID PARTITION   NAME      USER   ST   TIME   NODES NODELIST
162   batch   testBLAS  zhuofei  PD   0:15   1 (None)
160   batch   testBLAS  zhuofei   R   0:17   1 c2-11
161   batch   testBLAS  zhuofei   R   0:17   1 c2-11
```

Common STATE: R for Running; PD for PenDing; TO for TimedOut; S for Suspended; F for FAILED

TIME: the elapsed time used by the job, not remaining time, not CPU time.
Step7 (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

```bash
$ squeue -l
Wed Aug  8 14:03:47 2018
JOBID PARTITION   NAME     USER    STATE   TIME   TIME_LIMI  NODES NODELIST
169 batch   testBLAS  zhuofei  RUNNING  2:07    2:00:00       1 c1-38
168 batch   testBLAS  zhuofei  RUNNING  3:14    2:00:00       1 c1-39

$ scancel 169

[zhuofei@teach workDir]$ squeue -l
Wed Aug  8 14:03:47 2018
JOBID PARTITION   NAME     USER    STATE   TIME   TIME_LIMI  NODES NODELIST
169 batch   testBLAS  zhuofei  COMPLETI  2:25    2:00:00       1 c1-39
168 batch   testBLAS  zhuofei  RUNNING  3:32    2:00:00       1 c1-38

$ squeue -l
Wed Aug  8 14:04:08 2018
JOBID PARTITION   NAME     USER    STATE   TIME   TIME_LIMI  NODES NODELIST
168 batch   testBLAS  zhuofei  RUNNING  3:35    2:00:00       1 c1-38
```
Step7 (Cont.): Check job details using scontrol show job
https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_the_teaching_cluster

```
$ scontrol show job 174
JobId=174  JobName=testBLAST
    UserId=zhuofei(1772)  GroupId=gacrc-instruction(21004)  MCS_label=N/A
    JobState=RUNNING  Reason=None  Dependency=(null)
    Requeue=1  Restarts=0  BatchFlag=1  Reboot=0  ExitCode=0:0
    RunTime=00:04:28  TimeLimit=02:00:00  TimeMin=N/A
    SubmitTime=2018-08-08T14:28:44  EligibleTime=2018-08-08T14:28:44
    StartTime=2018-08-08T14:28:44  EndTime=2018-08-08T16:28:44  Deadline=N/A
...  Partition=batch  AllocNode:Sid=teach:30986
    NodeList=c1-38  NumNodes=1  NumCPUs=4  NumTasks=1  CPUs/Task=4  ReqB:S:C:T=0:0:*:*:
...  Command=/home/zhuofei/workDir/sub_blast.sh
    WorkDir=/home/zhuofei/workDir
    StdErr=/home/zhuofei/workDir/log.174
    StdOut=/home/zhuofei/workDir/log.174
```
Step 7 (Cont.): Check node info using `sinfo`

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

```
$ sinfo

PARTITION   AVAIL  TIMELIMIT  NODES  STATE      NODELIST
batch*       up     7-00:00:00  1  down*       tcn17
batch*       up     7-00:00:00 24  idle        tcn[1-16,18-25]
interactive  up     7-00:00:00  5  idle        tcn[26-30]
gpu          up     7-00:00:00  1  idle        tcgn1
highmem      up     7-00:00:00  2  idle        tchmn[1-2]
```

idle = no cores in use; mix = some cores are still free; alloc = all cores are allocated
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.
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:**

- Your user name
- Your job ID
- Your working directory
- The queue 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!
GACRC Service Catalog

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.
Services For Users

General Support
Report issues and request help with GACRC systems, except for software installation requests and account/lab creation requests.

Software Installation/Update
Request software and common application database (e.g. NCBI blast databases) installation and upgrade.

Training
Request support related to training provided by the GACRC.
Click to request
Thank You!

Telephone Support
EITS Help Desk: 706-542-3106
Monday – Thursday: 7:30 a.m. – 7:30 p.m.
Friday: 7:30 a.m. – 6 p.m.
Saturday – Sunday: 1 p.m. – 7 p.m.

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