Introduction to GACRC Teaching Cluster

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

• GACRC

• Overview

• Computing Resources
  ➢ Three Folders
  ➢ Three Computational Queues
  ➢ Software

• Submit Batch Job

• GACRC Wiki and Support
We are a high-performance-computing (HPC) center at UGA.

We 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

- [http://wiki.gacrc.uga.edu](http://wiki.gacrc.uga.edu) (GACRC Wiki)
- [https://wiki.gacrc.uga.edu/wiki/Getting_Help](https://wiki.gacrc.uga.edu/wiki/Getting_Help) (GACRC Support)
- [http://gacrc.uga.edu](http://gacrc.uga.edu) (GACRC Web)
Please note:
You need to connect to the UGA VPN when accessing from outside of the UGA main campus.

1. ssh with MyID and password
2. Verify with Archpass Duo two-factor authentication

Node: Computer for a specific function on cluster, e.g., login node
Queue: Collection of compute nodes for specific computing need
Cluster: Nodes + Drives, all connected by network
Computing Resources

- **Two Nodes:**
  1. Login node (MyID@teach.gacrc.uga.edu): for submitting computational jobs
  2. Transfer node (MyID@txfer.gacrc.uga.edu): for transferring data files

- **Three Directories:**
  1. /home/MyID: working space for computational jobs
  2. /work/CourseID/MyID: data parking for individual user in the class (e.g., /work/binf8940/MyID)
  3. /work/CourseID/instructor_data: data shared with class by the instructors

- **Three Queues:**
  1. batch: for running regular computational jobs
  2. highmem: for running high-memory jobs
  3. gpu: for running GPU jobs

5 GACRC TEACHING CLUSTER NEW USER TRAINING WORKSHOP
Computing Resources (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/2.7.14-foss-2016b
3. Software names are case-sensitive!

   - **module avail**: List all available software modules installed on cluster
   - **module load moduleName**: Load a module into your working environment
   - **module list**: List modules currently loaded
   - **module unload moduleName**: Remove a module from working environment
   - **ml spider pattern**: Search module names matching a pattern (case-insensitive)
Submit Batch Job

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 SSH File Transfer 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
Step1: 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 MyID password

4. Teaching cluster access requires ID 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

More information: https://wiki.gacrc.uga.edu/wiki/Connecting#Connecting_to_the_teaching_cluster
Step 1 (Cont.) - Mac/Linux

Using ssh in Terminal!

ssh zhuofei@teach.gacrc.uga.edu  ➔ 1. Log on

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
Success. Logging you in...
Last login: Fri Aug 3 11:24:43 2018 from 172.22.72.35
[zhuofei@teach ~]$  ➔ 5. Logged on!

9/21/2018
GACRC TEACHING CLUSTER NEW USER TRAINING WORKSHOP 9
Step1 (Cont.) - Windows

1. Download and install SSH Secure Utilities: [http://eits.uga.edu/hardware_and_software/software/](http://eits.uga.edu/hardware_and_software/software/)

2. You can use PuTTY as an alternative: [https://www.putty.org/](https://www.putty.org/)
Please Note:
Authentication Method needs to be set as
Keyboard Interactive in default <profile Setting>
Step 1 (Cont.) - Windows using SSH Secure Utilities

4. Connect to Remote Host
   - Host Name: teach.gacrc.uga.edu
   - User Name: MyID
   - Port Number: 22

5. Enter your UGA MyID password and click OK
Step 1 (Cont.) - Windows using SSH Secure Utilities

9. Enter “push” and click OK

10. Verify login using Duo
Step1 (Cont.) - Windows using SSH Secure Utilities

11. Click OK

12. Logged on!
Step2 - 3: Create and change directory to workDir

[zhuofei@teach ~]$: ls

[zhuofei@teach ~]$: mkdir workDir

[zhuofei@teach ~]$: ls

workDir

[zhuofei@teach ~]$: cd workDir/

[zhuofei@teach workDir]$: ls

[zhuofei@teach workDir]$: 

← ls command to list folder’s contents

← mkdir command to create a subdirectory

← cd command to change directory

← it is empty in workDir!
Step 4: Transfer data from local computer to workDir - Mac/Linux

1. Connect to Transfer node (MyID@txfer.gacrc.uga.edu) in Terminal on local computer
2. Type scp command: `scp (-r) [Source] [Target]`
3. Once you input MyID password, scp command will send “push” to your Duo Enrolled mobile device for verification

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

```bash
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

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

[https://wiki.gacrc.uga.edu/wiki/Transferring_Files#The_File_Transfer_node_for_the_teaching_cluster](https://wiki.gacrc.uga.edu/wiki/Transferring_Files#The_File_Transfer_node_for_the_teaching_cluster)
Step 4 (Cont.) - Windows using SSH Secure Utilities

Please Note:
Authentication Method needs to be set as Keyboard Interactive in default <profile Setting>
Step 4 (Cont.) - Windows using SSH Secure Utilities

4. Host Name: txfer.gacrc.uga.edu
   User Name: MyID
   Port Number: 22

5. Connect to Remote Host

6. Enter Host Name, User Name, Port Number: 22

7. Click Connect

8. Enter your UGA MyID password and click OK

Steps 9 - 11 are the same as listed on page 13 - 14!
Step4 (Cont.) - Windows using SSH Secure Utilities

12. Logged on!

13. Click yellow button

14. Change local and remote paths
15. Drag data between local computer and remote cluster
Step4 (Cont.): Transfer data on cluster to workDir

• Log on to Transfer node (MyID@txfer.gacrc.uga.edu)
  ✓ Mac/Linux: ssh MyID@txfer.gacrc.uga.edu (page 8-9)
  ✓ Windows: use SSH Secure Client app (page 14-16)

• Directories you can access on txfer:
  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/pbio6550/MyID/datafile /home/MyID/workDir`
Step5: Make a job submission script in workDir using nano

```bash
$ nano sub.sh
```

nano is a small and friendly text editor on Linux.

Ctrl-x to save file and quit from nano
Step5 (Cont.)

Please copy

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

```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+/2.6.0-foss-2016b-Python-2.7.14

time blastn -num_threads 4 -query sample.fasta -db /db/ncbiblast/nrte/06222018/nt\ -out results.$(SLURM_JOB_ID) -outfmt 6 -max_target_seqs=2
```

More Information: https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_the_teaching_cluster
Step6: Submit a job from workDir using sbatch

$ 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
Step7: Check job status using squeue

$ squeue -l
Wed Aug 8 13:40:02 2018

<table>
<thead>
<tr>
<th>JOBID</th>
<th>PARTITION</th>
<th>NAME</th>
<th>USER</th>
<th>STATE</th>
<th>TIME</th>
<th>TIME_LIMI</th>
<th>NODES</th>
<th>NODELIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>PENDING</td>
<td>0:00</td>
<td>2:00:00</td>
<td>1</td>
<td>(None)</td>
</tr>
<tr>
<td>160</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>RUNNING</td>
<td>0:02</td>
<td>2:00:00</td>
<td>1 c2-11</td>
<td>c2-11</td>
</tr>
<tr>
<td>161</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>RUNNING</td>
<td>0:02</td>
<td>2:00:00</td>
<td>1 c2-11</td>
<td>c2-11</td>
</tr>
</tbody>
</table>

$ squeue

<table>
<thead>
<tr>
<th>JOBID</th>
<th>PARTITION</th>
<th>NAME</th>
<th>USER</th>
<th>ST</th>
<th>TIME</th>
<th>NODES</th>
<th>NODELIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>PD</td>
<td>0:15</td>
<td>1</td>
<td>(None)</td>
</tr>
<tr>
<td>160</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>R</td>
<td>0:17</td>
<td>1 c2-11</td>
<td>c2-11</td>
</tr>
<tr>
<td>161</td>
<td>batch</td>
<td>testBLAS</td>
<td>zhuofei</td>
<td>R</td>
<td>0:17</td>
<td>1 c2-11</td>
<td>c2-11</td>
</tr>
</tbody>
</table>

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.
Step 7 (Cont.): Cancel job using `scancel`.

```sh
$ 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  COMPLETED 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`

```
$ 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
```
Step7 (Cont.): Check node info using sinfo

```
$ sinfo

<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>highmem</td>
<td>up 7</td>
<td>00:00:00</td>
<td>5</td>
<td>idle</td>
<td>c1-[36-37,40],c2-[9-10]</td>
</tr>
<tr>
<td>gpu</td>
<td>up 1</td>
<td>00:00:00</td>
<td>1</td>
<td>idle</td>
<td>c2-2</td>
</tr>
<tr>
<td>interq</td>
<td>up 1</td>
<td>00:00:00</td>
<td>3</td>
<td>idle</td>
<td>c2-[4-6]</td>
</tr>
<tr>
<td>batch</td>
<td>up 3</td>
<td>00:00:00</td>
<td>3</td>
<td>mix</td>
<td>c1-38,c2-[11-12]</td>
</tr>
<tr>
<td>batch</td>
<td>up 3</td>
<td>00:00:00</td>
<td>1</td>
<td>alloc</td>
<td>c1-1</td>
</tr>
<tr>
<td>batch</td>
<td>up 3</td>
<td>00:00:00</td>
<td>36</td>
<td>idle</td>
<td>c1-[2-35,39]</td>
</tr>
</tbody>
</table>
```

idle = no cores in use; mix = some cores are still free; alloc = all cores are allocated
GACRC Wiki [http://wiki.gacrc.uga.edu](http://wiki.gacrc.uga.edu)

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)

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

Transfer File:

[https://wiki.gacrc.uga.edu/wiki/Transferring_Files#The_File_Transfer_node_for_the_teaching_cluster](https://wiki.gacrc.uga.edu/wiki/Transferring_Files#The_File_Transfer_node_for_the_teaching_cluster)

Linux Command: [https://wiki.gacrc.uga.edu/wiki/Command_List](https://wiki.gacrc.uga.edu/wiki/Command_List)

Training: [https://wiki.gacrc.uga.edu/wiki/Training](https://wiki.gacrc.uga.edu/wiki/Training)
GACRC Support [https://wiki.gacrc.uga.edu/wiki/Getting_Help](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
  ✓ Job ID
  ✓ Job submission script and command you used to submit the job
  ✓ Your working directory on cluster

- **Software Installation:**
  
  ✓ Specific name and version of the software
  ✓ Download website
  ✓ Supporting package information if have

Please note:
1. In general only software widely used by the GACRC computing community will be centrally installed.
2. Make sure of the correctness of datasets being used by your jobs!
Request Support

* indicates Required fields.

Your Name *
First & Last

MyID *
myld

E-mail *
you@example.edu

Phone Number
411-555-1212

Brief Description

Request Details *

Cluster
☐ sapelo2
☐ sapelo
☐ teach
☐ other

* For questions on cluster or software, please include the command/script used, working path and working nodes (interactive / queue name) if applicable.
* For software installation, please specify software name, version and include link to the software if applicable.
* Please review your message on the next page and then click the Submit button.
Thank You!