GACRC Sapelo2 Cluster New User Training

Georgia Advanced Computing Resource Center (GACRC)

Enterprise Information Technology Services(EITS)

The University of Georgia

Outline

- GACRC
- Sapelo2 Cluster
  - Cluster Diagram
  - Overview
  - Five Directories
  - Four Computational Queues
  - Software Modules
- Batch Job Submission Workflow
- GACRC Wiki and User Support

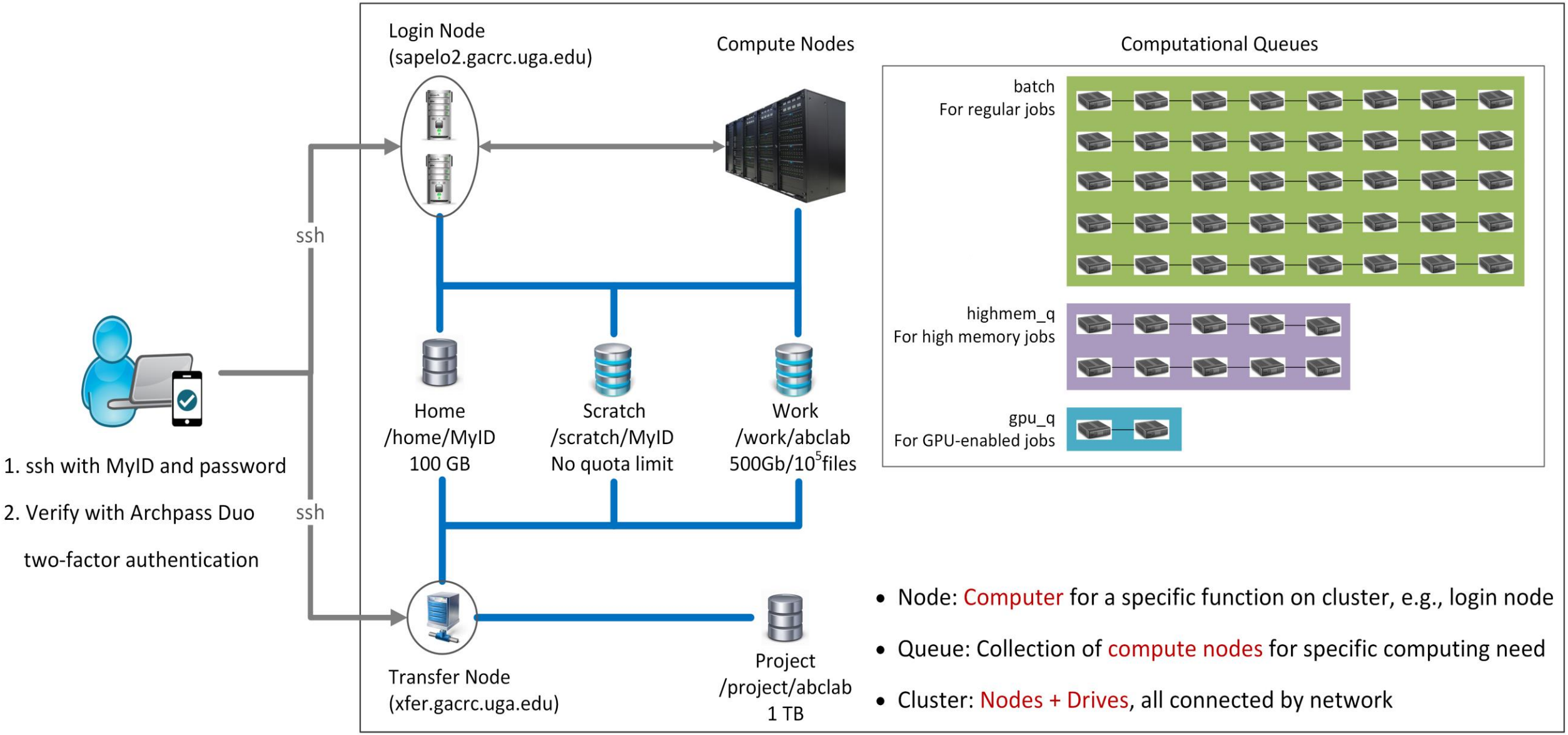
GACRC

- 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://uga.teamdynamix.com/TDClient/Requests/ServiceCatalog?CategoryID=11593

Web Site: http://gacrc.uga.edu



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/

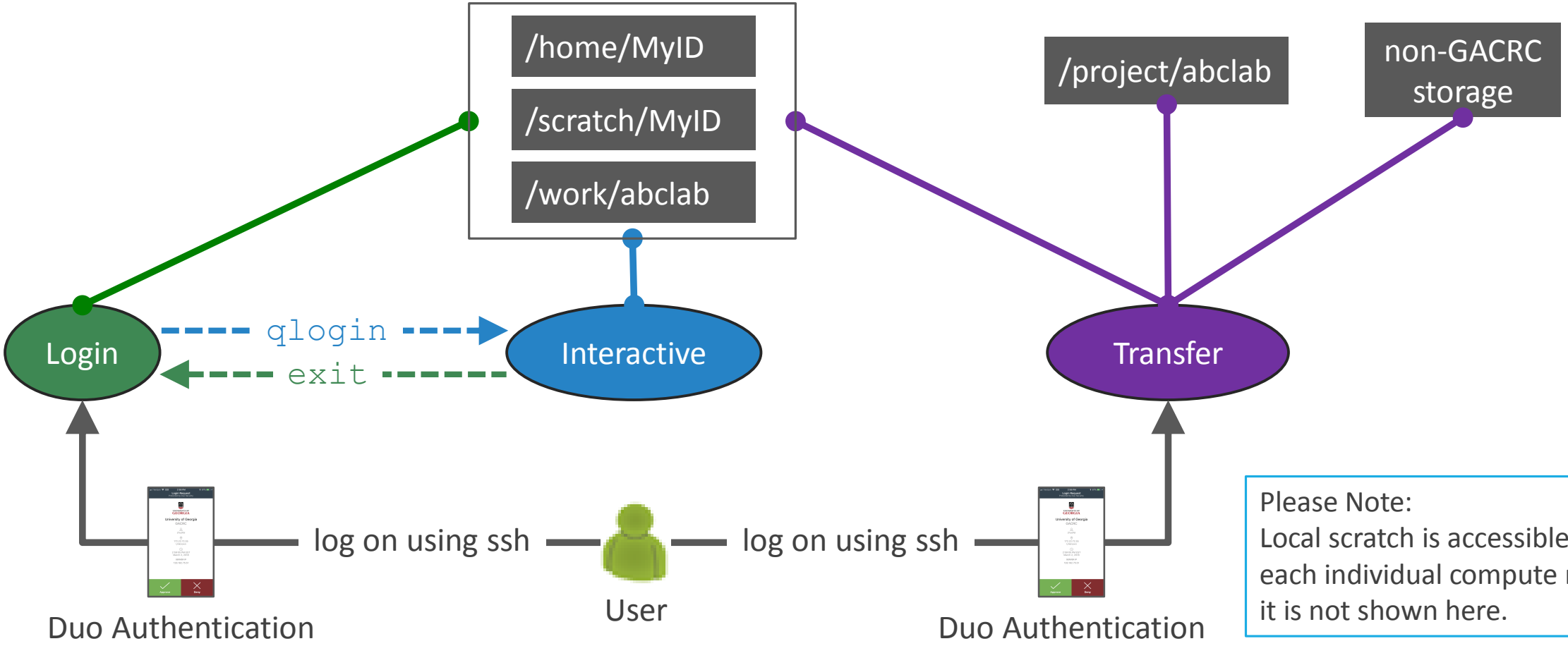
Overview https://wiki.gacrc.uga.edu/wiki/Systems#Sapelo2

- 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

Five Directories https://wiki.gacrc.uga.edu/wiki/Disk_Storage

Directory	Name	Quota	Accessible from	Intended Use	Backed-up	Important Notes
/home/MyID	Home	100GB	Login Transfer Compute	Static data, e.g. 1. Scripts, source codes 2. Local software	Yes	Not for storing data of your jobs!
/scratch/MyID	Scratch	No Limit		Temporary files needed for current running jobs	No	Clean up when your job finished! Subject to "30-day purge" policy
/work/abclab	Work	500GB 10^5 files		Input files needed for repeated jobs	No	Clean up any old data! Group sharing is possible
/project/abclab	Project	1TB (initial)	Transfer	Temporary data parking	Yes	Group sharing is possible
/lscratch	Local Scratch	200GB	Compute	Jobs with heavy disk I/O operations	No	Clean up when job exits from node! Data are persistent

Accessing Directories from Nodes



Please Note:
Local scratch is accessible on each individual compute node; it is not shown here.

More about scratch file system "30-day purge" policy

https://wiki.gacrc.uga.edu/wiki/Disk_Storage#Scratch_file_system

Any file that is not accessed or modified by a compute job in a time period no longer than 30 days will be automatically deleted off the /scratch file system.

Measures circumventing this policy will be monitored and actively discouraged.

- You have a list of those purgeable files located at /usr/local/var/lustre_stats/$USER.over30d.files.lst
- You are suggested to copy files from /scratch to /project or outside of GACRC
- You should first move all unnecessary files and folders to /scratch/trash/$USER
- The fastest way to save your old files is to copy them to /project area, using the fpsync utility on xfer.gacrc.uga.edu
- If you want to first create a tar archive of your /scratch area, DO NOT compress the archive when creating the archive

Queue	Total Nodes	RAM(GB) /Node	Max Mem(GB) /Single-node job	Cores /Node	Processor Type	GPU Cards /Node	InfiniBand
batch	42	192	184	32	Intel Xeon Skylake	N/A	Yes
	32	64	58	28	Intel Xeon Broadwell		
	106	128	120	48	AMD Opteron		
	16			32	AMD EPYC		
highmem_q	5	1024	990	28|48	Intel Xeon Broadwell (4) AMD Opteron (1)		
	15	512	502	32|48	Intel Xeon Nehalem (1) AMD EPYC (8)|Opteron (6)		
gpu_q	4	192	184	32	Intel Xeon Skylake	1 NVDIA P100	
	2	128	120	16	Intel Xeon	8 NVIDIA K40m	
	4	96	90	12		7 NVIDIA K20Xm	
groupBuyin_q	variable						

Software Environment

1. Software names are long and have a EasyBuild toolchain name associated to it
2. Complete module name: Name/Version-toolchain, e.g., Python/3.6.4-foss-2018a
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)

Job Submission Workflow

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

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 `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`

6. Make a job submission script in workDir : `nano ./sub.sh`
7. Submit a job from workDir : `qsub ./sub.sh`
8. Check job status : `qstat_me` or Cancel a job : `qdel JobID`

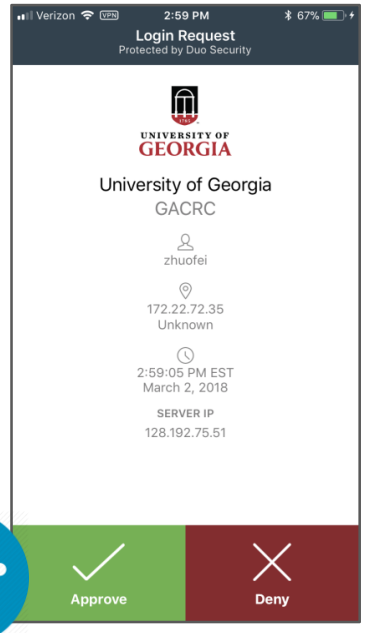
Step1: Log on to Login node - Mac/Linux using ssh

https://wiki.gacrc.uga.edu/wiki/Connecting#Connecting_to_Sapelo2

1. Open Terminal utility
2. Type command line: `ssh MyID@sapelo2.gacrc.uga.edu`
3. You will be prompted for your MyID password
4. Sapelo2 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

Step1 (Cont.) - Mac/Linux using ssh

Use Terminal utility on Mac or Linux!



zhuofeihou — zhuofei@sapelo2-sub2:~ — ssh — 80×22

```
zhuofeihou@s172-20-19-h149:~$
zhuofeihou@s172-20-19-h149:~$
zhuofeihou@s172-20-19-h149:~$ssh zhuofei@sapelo2.gacrc.uga.edu
Password:
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-1925
 3. Phone call to XXX-XXX-5758
 4. SMS passcodes to XXX-XXX-5758

Passcode or option (1-4): 1

Pushed a login request to your device...
Success. Logging you in...
zhuofei@sapelo2-sub2 ~$
zhuofei@sapelo2-sub2 ~$
```

← Log on

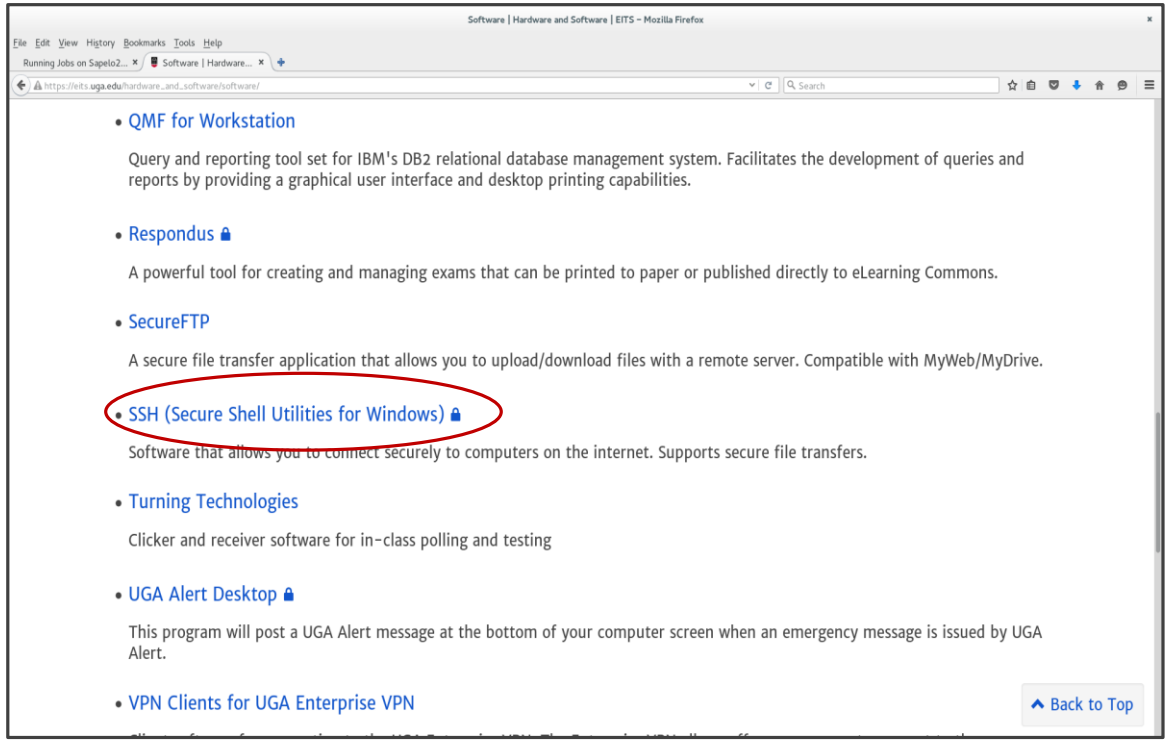
← Input MyID password!

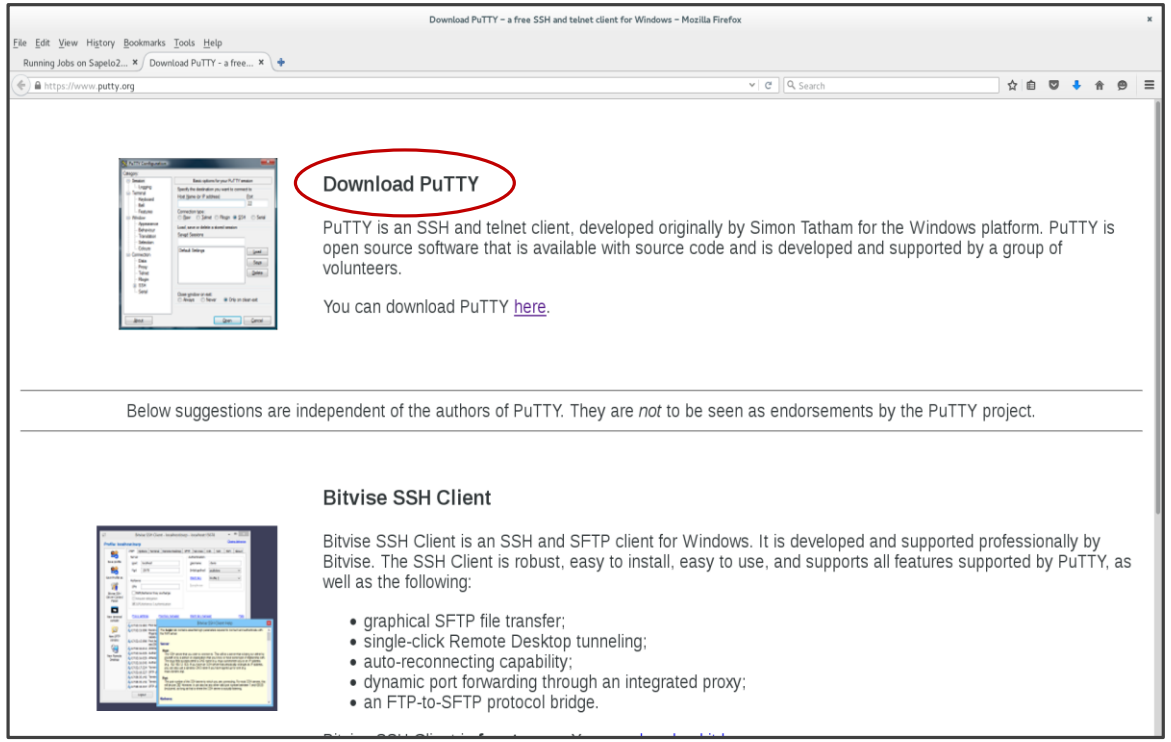
← Select Duo login option!

← I am on sapelo2-sub2!

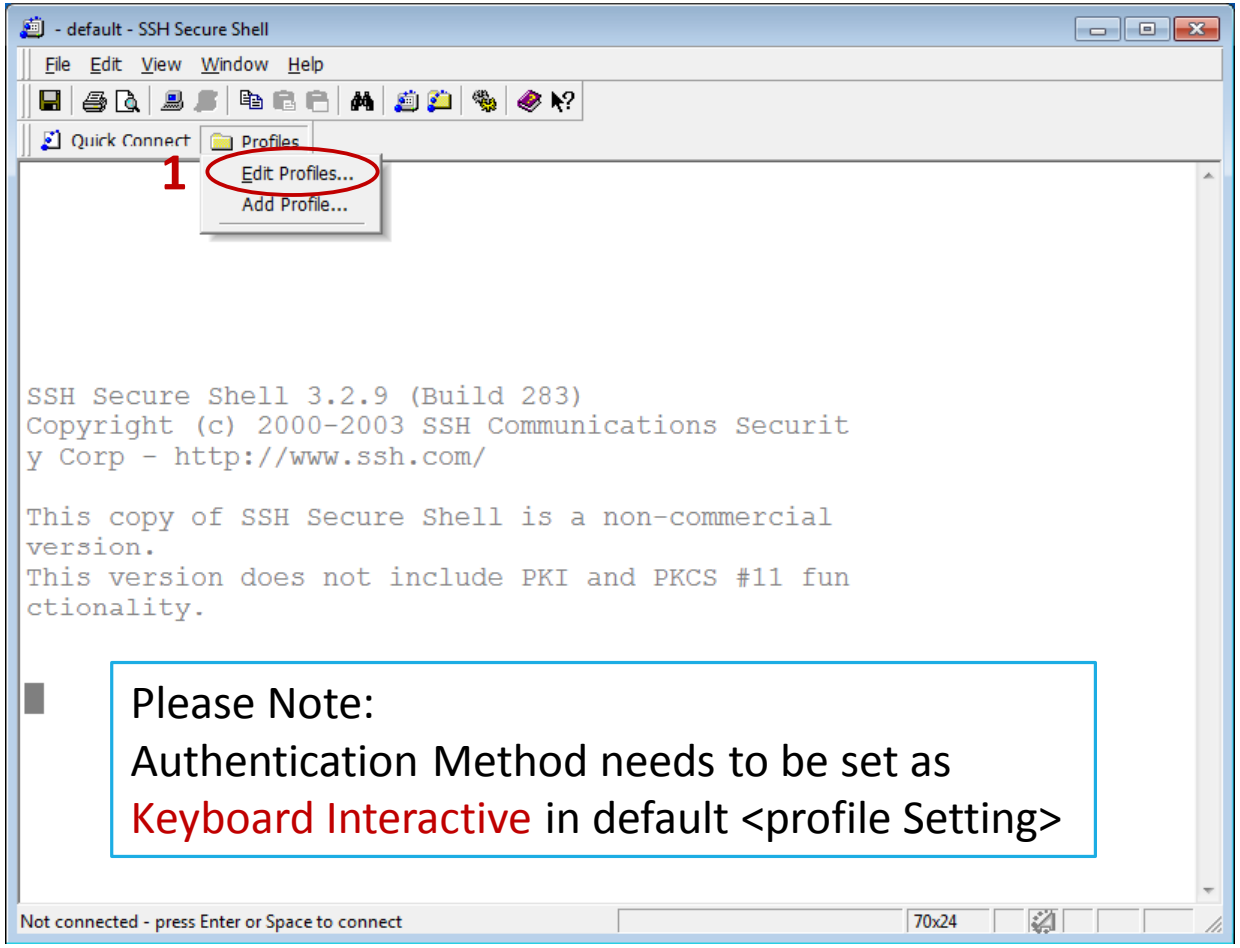
Step1 (Cont.) - Windows using SSH Secure Utilities

1. Download and install SSH Secure Utilities: http://eits.uga.edu/hardware_and_software/software/
2. You can use PuTTY as an alternative: https://www.putty.org/

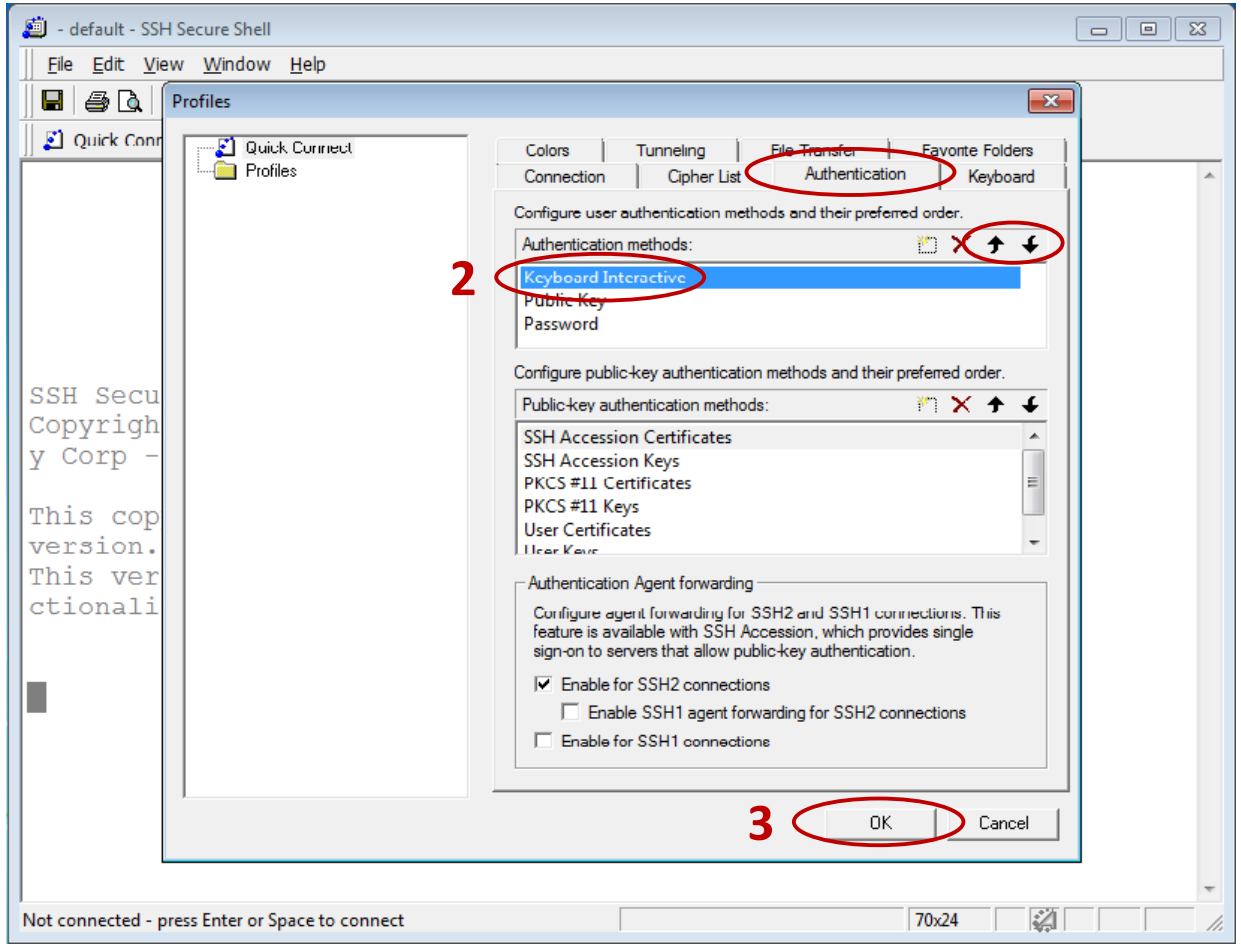




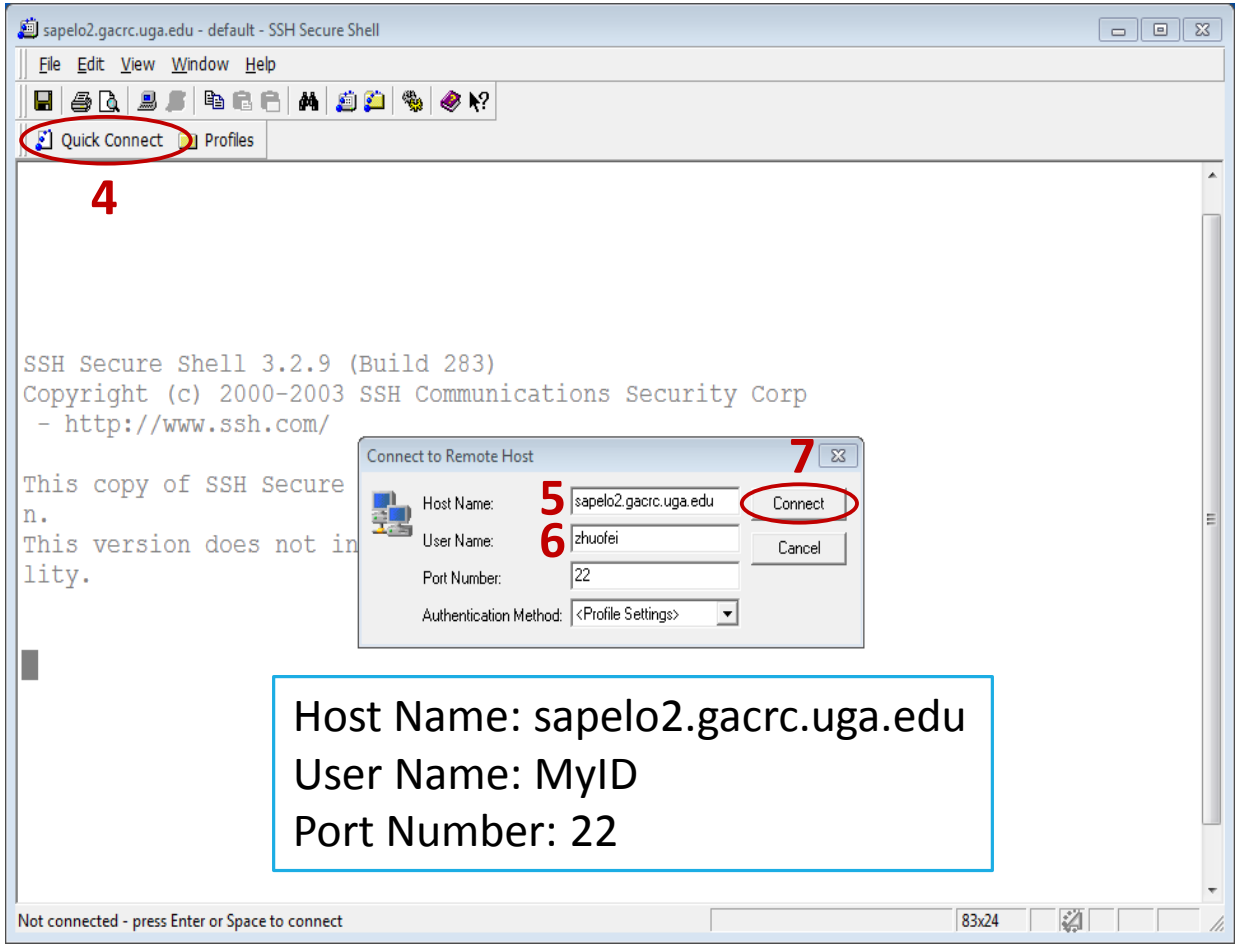
Step1 (Cont.) - Windows using SSH Secure Utilities



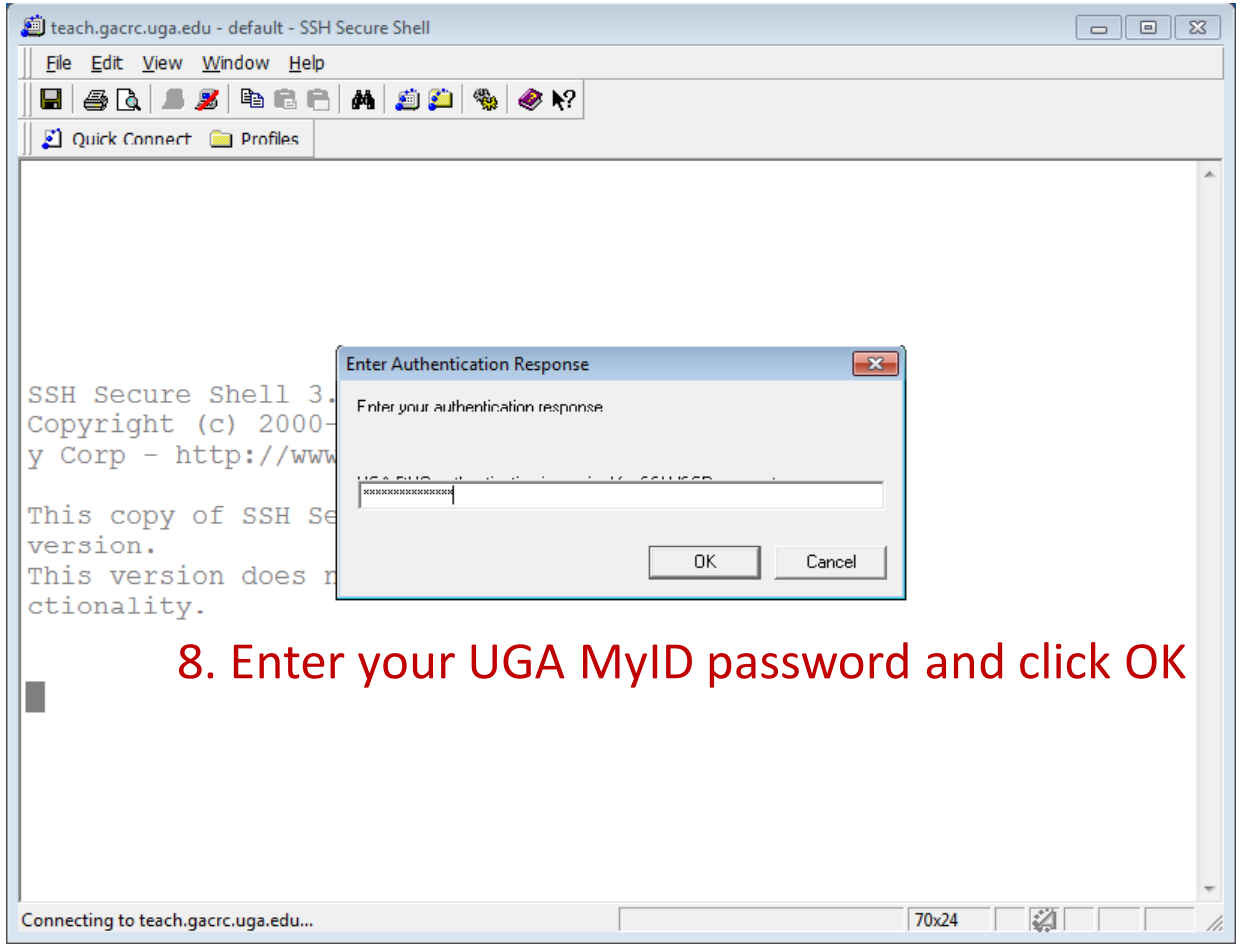
Please Note:
Authentication Method needs to be set as
Keyboard Interactive in default <profile Setting>



Step1 (Cont.) - Windows using SSH Secure Utilities

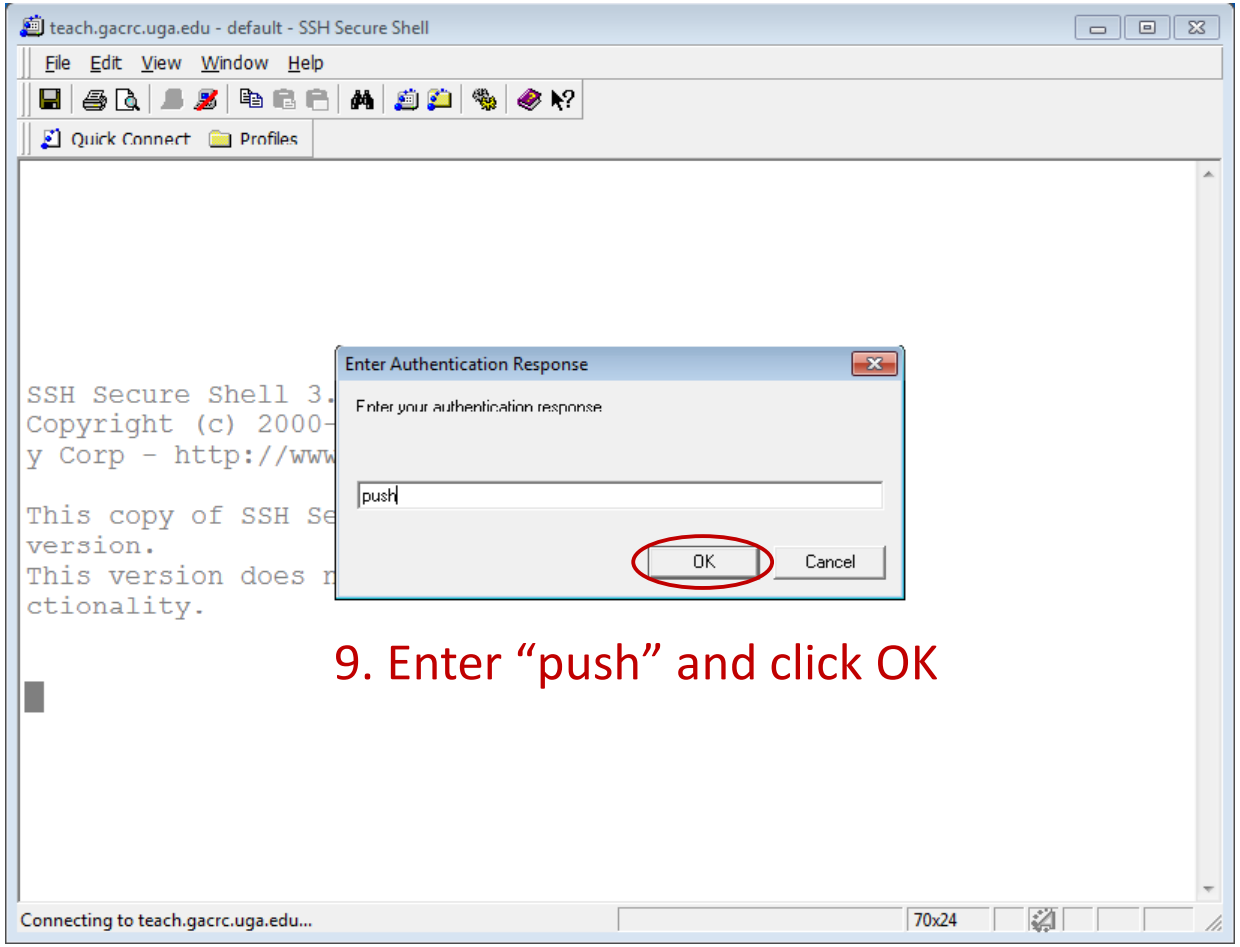


Host Name: sapelo2.gacrc.uga.edu
User Name: MyID
Port Number: 22

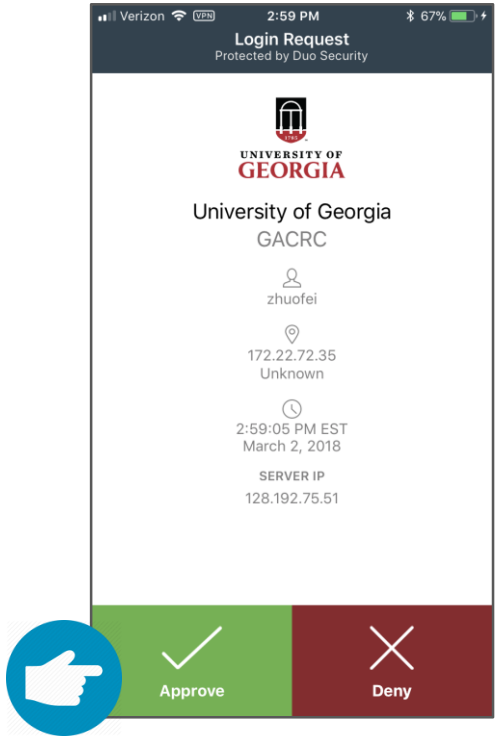


8. Enter your UGA MyID password and click OK

Step1 (Cont.) - Windows using SSH Secure Utilities



9. Enter “push” and click OK



10. Verify login using Duo, then you will log on!

Step2: On Login node change directory to global scratch

- Once you logged on, your current directory will be your home directory

```
zhuofei@sapelo2-sub2 ~$ pwd
/home/zhuofei
```

← this is my home directory!

- Use `cd` command to change your current directory to /scratch/MyID

```
zhuofei@sapelo2-sub2 /$ cd /scratch/zhuofei
zhuofei@sapelo2-sub2 zhuofei$ pwd
/scratch/zhuofei
```

← this is my scratch space!

- Use `ls` command to take a look in /scratch/MyID

```
zhuofei@sapelo2-sub2 zhuofei$ ls
user_test  workDir_Alex  workDir_bk
```

Step3 - 4: Create and cd to a working subdirectory

- Use `mkdir` command to make a subdirectory in /scratch/MyID

```
zhuofei@sapelo2-sub2 zhuofei$ mkdir workDir
zhuofei@sapelo2-sub2 zhuofei$ ls
user_test  workDir  workDir_Alex  workDir_bk
```

- Use `cd` command to change your current directory to /scratch/MyID/workDir

```
zhuofei@sapelo2-sub2 zhuofei$ cd workDir
zhuofei@sapelo2-sub2 workDir$ ls
zhuofei@sapelo2-sub2 workDir$
```

← it is empty!

Step5: Transfer data from local computer to workDir - Mac/Linux

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

1. You need to connect to cluster's Transfer node (xfer.gacrc.uga.edu)
2. Open Terminal utility on local computer to use scp (-r) [Source] [Target]

E.g. 1: working on local computer, from Local ➔ workDir on cluster

```
scp ./file zhuofei@xfer.gacrc.uga.edu:/scratch/zhuofei/workDir/
```

```
scp -r ./folder/ zhuofei@xfer.gacrc.uga.edu:/scratch/zhuofei/workDir/
```

E.g. 2: working on local computer, from workDir on cluster ➔ Local

```
scp zhuofei@xfer.gacrc.uga.edu:/scratch/zhuofei/workDir/file .
```

```
scp -r zhuofei@xfer.gacrc.uga.edu:/scratch/zhuofei/workDir/folder/ .
```

Step5 (Cont.) - Windows using SSH Secure Utilities

1. You need to connect to cluster's Transfer node (xfer.gacrc.uga.edu)

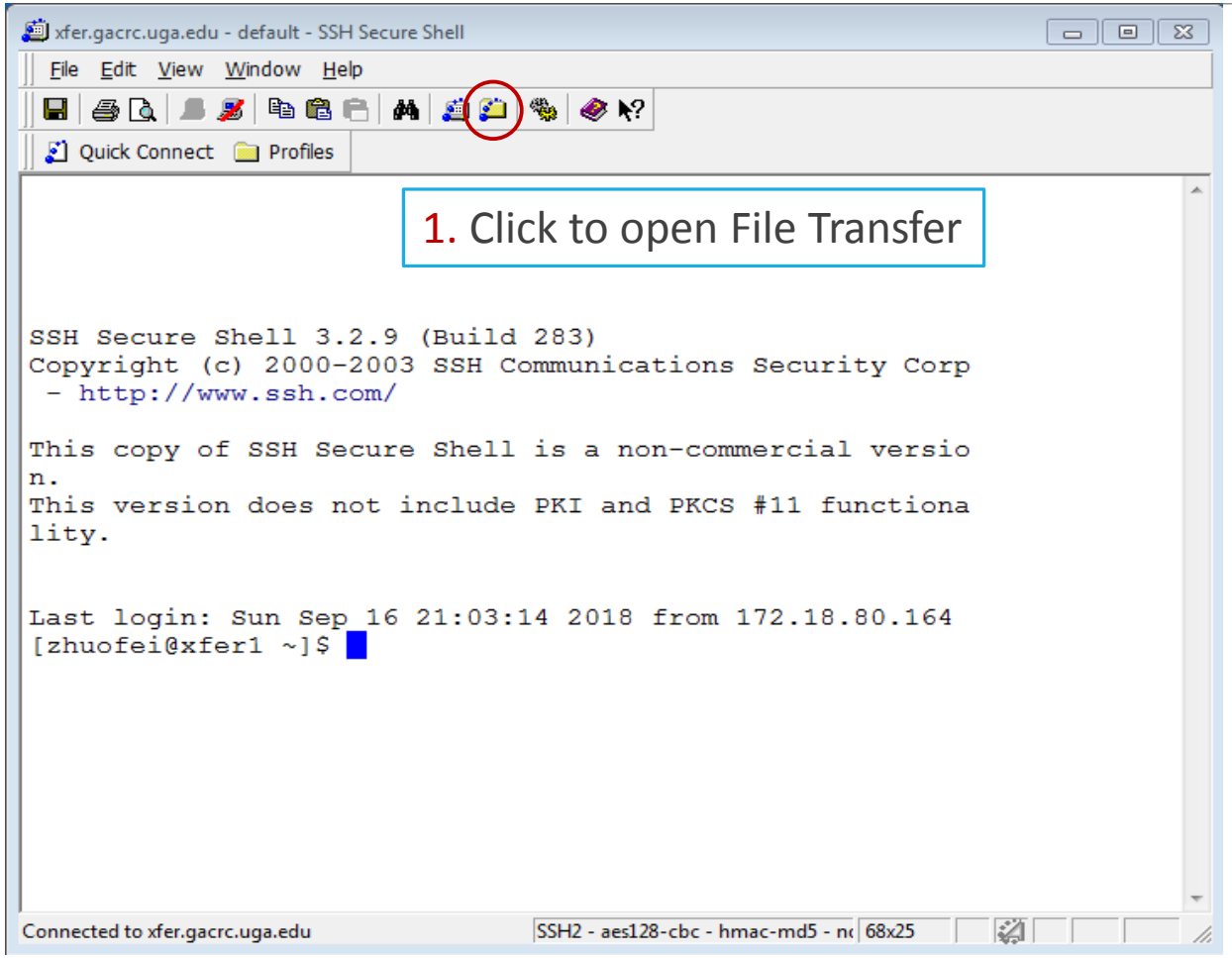
2. Use SSH File Transfer on local computer (alternative FileZilla or WinSCP)

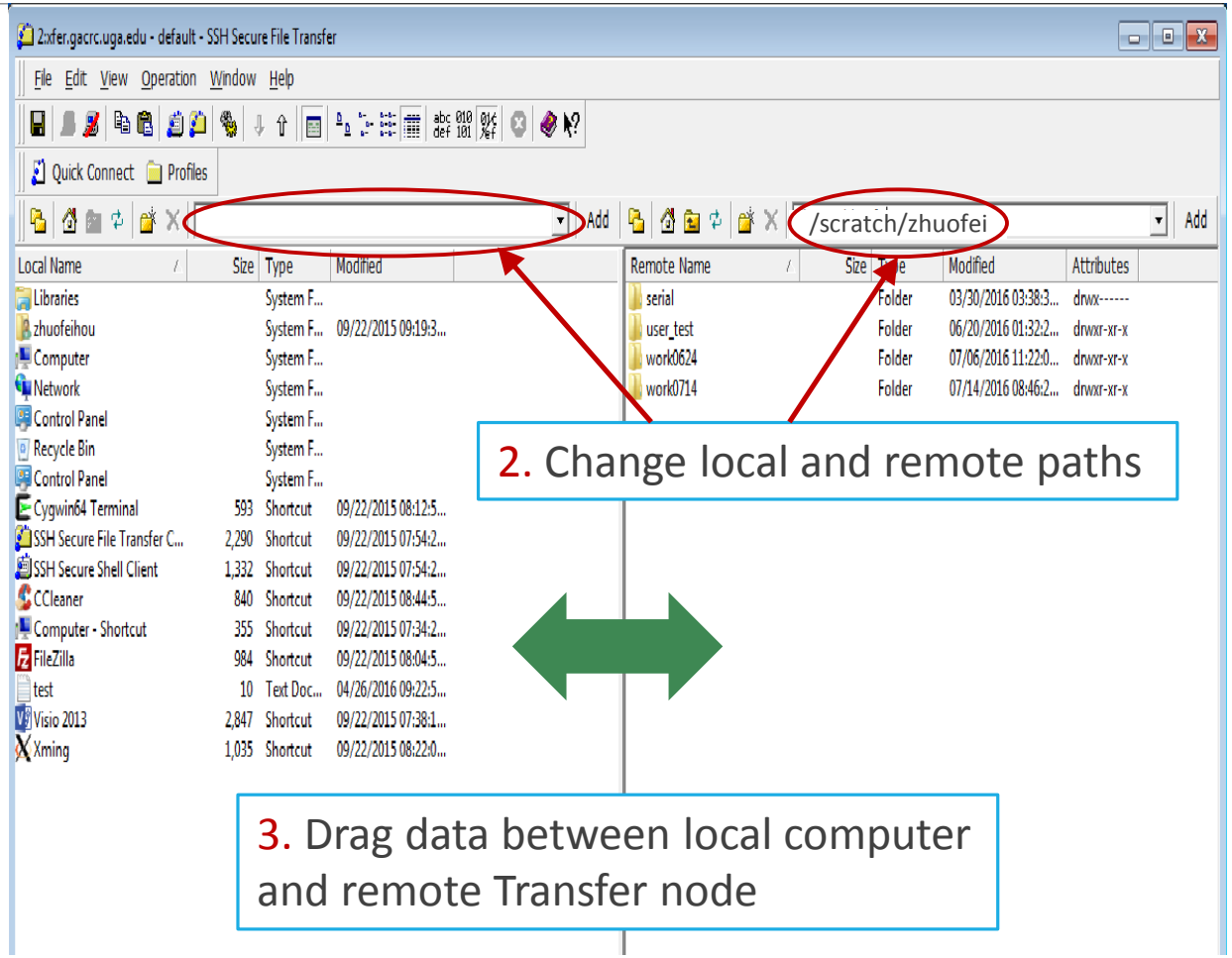
3. Steps 1-10 are the same as steps on page 14-15, except for Host Name in step

   Host Name: xfer.gacrc.uga.edu

4. Once you log on, use File Transfer of SSH Secure Utilities, as shown on next page

Step5 (Cont.) - Windows using SSH Secure Utilities





Step5 (Cont.): Transfer data on cluster to workDir

- Log on to Transfer node (xfer.gacrc.uga.edu)
  - ✓ Mac/Linux: ssh MyID@xfer.gacrc.uga.edu (page 12)
  - ✓ Windows: use SSH Secure Utilities (page 19)
- Landing folder: /home/MyID (Home)
- You can transfer data between following directories on cluster using cp or mv:
  1. /home/MyID (Home)
  2. /scratch/MyID (Scratch)
  3. /work/abclab (Work)
  4. /project/abclab (Project)
- Most file systems on Transfer are *auto-mounted* upon *the first time full-path access*, e.g., cd /project/abclab/

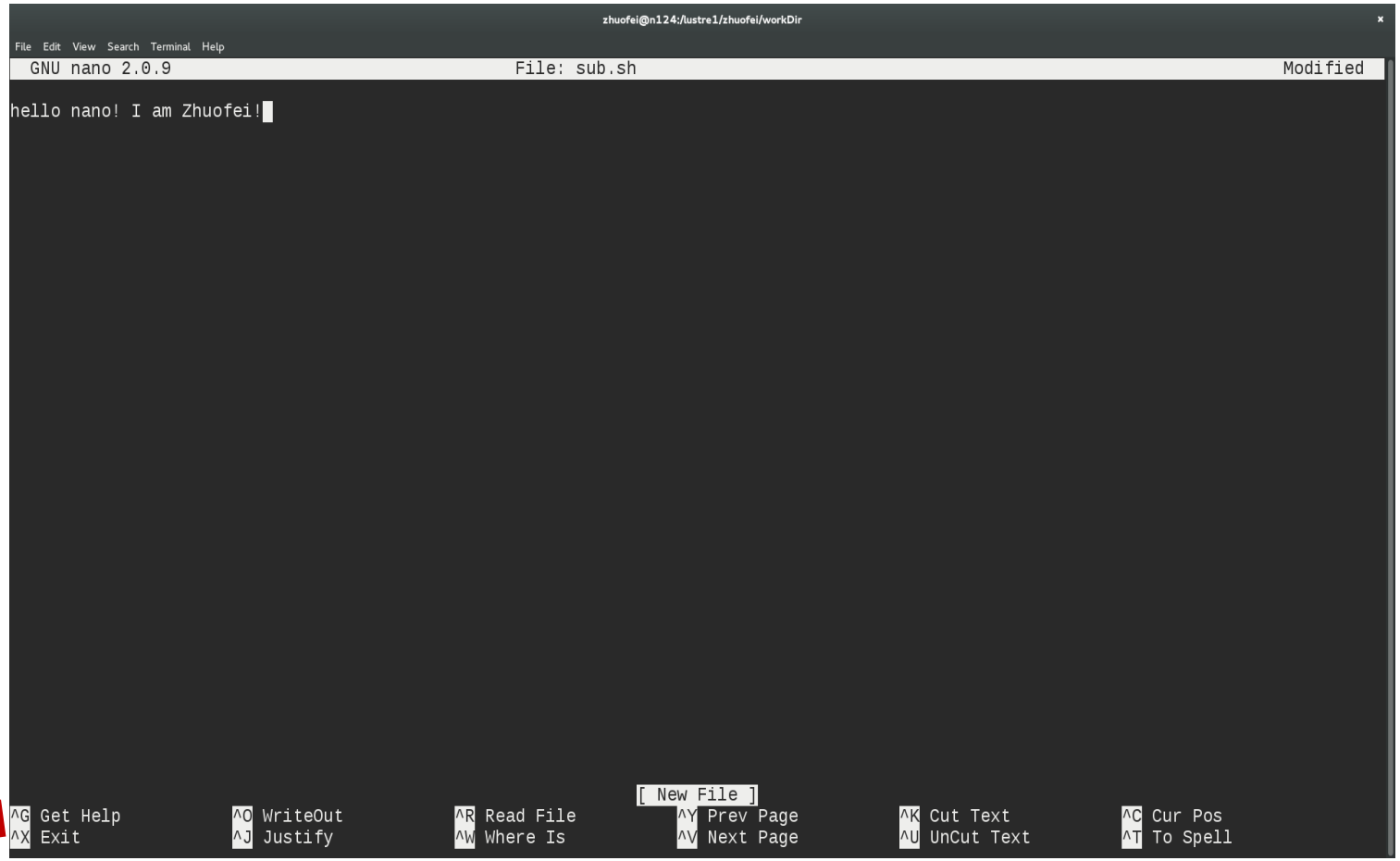
Step6: Make a job submission script in workDir

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

```
$ nano sub.sh
```

nano is a small and friendly text editor on Linux.

Ctrl-x to save file and quit from nano



Step6 (Cont.)

1. Sample script on GACRC Wiki Software page:

   https://wiki.gacrc.uga.edu/wiki/Bowtie2-Sapelo2

2. Modify it as needed for your computing

To run this example, you need to copy 3 files into your current working dir:

```
cp  /usr/local/training/sub.sh   .
cp  /usr/local/training/myreads.fq   .
cp  -r  /usr/local/training/index   .
```

zhuofei@sapelo2-sub2:/lustre1/zhuofei/workDir

File Edit View Search Terminal Help

```
  GNU nano 2.3.1              File: sub.sh                        Modified

#PBS -S /bin/bash              ← default Linux bash shell
#PBS -q batch                  ← use batch queue
#PBS -N bowtie2_test           ← job name
#PBS -l nodes=1:ppn=1          ← 1 compute node and 1 core from the node
#PBS -l mem=2gb                ← total RAM memory (enforced!)
#PBS -l walltime=1:00:00       ← job running time on cluster (enforced!)

#PBS -M zhuofei@uga.edu
#PBS -m ae

cd $PBS_O_WORKDIR

module load Bowtie2/2.3.3-foss-2016b

time bowtie2 -p 1 -x ./index/lambda_virus -U ./myreads.fq -S output.sam

^G Get Help  ^O WriteOut  ^R Read File^Y Prev Page^K Cut Text  ^C Cur Pos
^X Exit      ^J Justify   ^W Where Is ^V Next Page^U UnCut Tex^T To Spell
```

Step7: Submit a job from workDir using qsub

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

```
zhuofei@sapelo2-sub2 workDir$ pwd
/scratch/zhuofei/workDir
zhuofei@sapelo2-sub2 workDir$ ls
index  myreads.fq  sub.sh
zhuofei@sapelo2-sub2 workDir$ qsub sub.sh
11943.sapelo2
```

sub.sh is job submission script to

1. specify computing resources:
2. load software using module load
3. run any Linux commands you want to run
4. run the software

Step8: Check job status using qstat_me

https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_Sapelo2

```
zhuofei@sapelo2-sub2 workDir$ qstat_me
Job ID                  Name             User      Time Use S Queue
------------------- ---------------- --------- -------- - -----
11943.sapelo2           bowtie2_test     zhuofei   00:06:40 C batch
11944.sapelo2           bowtie2_test     zhuofei   00:05:17 R batch
11946.sapelo2           bowtie2_test     zhuofei   00:12:51 R batch
11947.sapelo2           bowtie2_test     zhuofei          0 R batch
11948.sapelo2           bowtie2_test     zhuofei          0 Q batch
```

R : job is running
C : job completed (or canceled or crashed) and is not longer running. Jobs stay in this state for 24 hour
Q : job is pending, waiting for resources to become available

Note: “Time Use” is the CPU time, instead of the wall-clock time of your job staying on cluster!

Step8 (Cont.): Cancel job using qdel

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

```
zhuofei@sapelo2-sub2 workDir$ qdel 11947
zhuofei@sapelo2-sub2 workDir$ qstat_me
Job ID                  Name             User            Time Use S Queue
------------------- ---------------- ------------- -------- - -----
11943.sapelo2           bowtie2_test     zhuofei         00:06:40 C batch
11944.sapelo2           bowtie2_test     zhuofei         00:05:17 R batch
11946.sapelo2           bowtie2_test     zhuofei         00:12:51 R batch
11947.sapelo2           bowtie2_test     zhuofei         00:00:09 C batch
11948.sapelo2           bowtie2_test     zhuofei                0 Q batch
```

job 11947 status is changed from R to C

C status will stay in list for ~24 hour

Step8 (Cont.): Check Job using qstat -n -u MyID

https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_Sapelo2

```
zhuofei@sapelo2-sub2 workDir$ qstat -n -u zhuofei

dispatch.ecompute:
                                                                  Req'd   Req'd        Elap
Job ID          Username  Queue  Jobname        SessID  NDS  TSK  Memory  Time      S  Time
-------------- -------- ----- ------------ ------ --- --- ------ -------- -  --------
12175.sapelo2 zhuofei   batch  bowtie2_test  132442  1    1    2gb     01:00:00 R  00:23:44
   n238/0  ← job is running on node238/CPU0
12176.sapelo2 zhuofei   batch  bowtie2_test  67226   1    1    2gb     01:00:00 R  00:20:44
   n237/0
12177.sapelo2 zhuofei   batch  bowtie2_test  119643  1    1    2gb     01:00:00 R  00:05:44
```

Note: “Elap Time” is the wall-clock time, instead of the CPU time, which qstat_me can give you!

Step8 (Cont.): Check all Jobs on cluster using qstat

https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_Sapelo2

```
zhuofei@sapelo2-sub2 workDir$ qstat
Job ID                    Name             User            Time Use S Queue
------------------------- ---------------- --------------- -------- - -----
11267.sapelo2              L80-500          jx57780         164:32:5 R batch
11269.sapelo2              L80-502          jx57780         164:55:5 C batch
11270.sapelo2              L80-503          jx57780         165:38:5 C batch
11607.sapelo2              canu             gd98309         3414:46: R bergman_q
11726.sapelo2              ..._3_constoptTS sm39091         3157:30: R wheeler_q
11729.sapelo2              ..._2_constoptTS sm39091         2731:29: R wheeler_q
11790.sapelo2              sp2_run19b_dye   castelao        4412:52: C batch
11804.sapelo2              ...e-4_Nnoise=64 cotter          98:26:20 R batch
11806.sapelo2              ...e-4_Nnoise=64 cotter          98:14:22 R batch
11987.sapelo2              ..._th_W18-T5-L4 qbcg            08:02:40 C batch
11989.sapelo2              matlabjob        zhyw86                 0 C batch
11990.sapelo2              ..._1_constoptTS sm39091         445:42:1 R wheeler_q
11991.sapelo2              ..._1_constoptTS sm39091         444:51:4 R wheeler_q
11992.sapelo2              ..._cl_W18-T3-D1 qbcg            03:04:21 C batch
.
.
```

Note: qstat command will give you a long list of all jobs from all users on cluster!

Workflow Diagram

1. Linux/Mac user:
`ssh MyID@sapelo2.gacrc.uga.edu`

Windows user:

Login

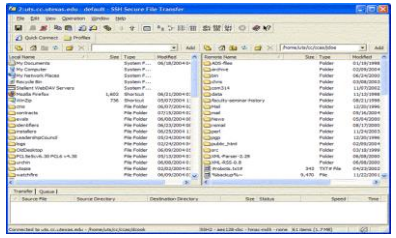
2. `cd /scratch/MyID`

3. `mkdir ./workDir`
4. `cd ./workDir`

lustre1

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

Windows user:



6. `nano ./sub.sh`

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -N bowtie2_test
#PBS -l nodes=1:ppn=1
#PBS -l mem=2gb
#PBS -l walltime=1:00:00

#PBS -M yourMyID@uga.edu
#PBS -m ae

cd $PBS_O_WORKDIR

module load Bowtie2/2.3.3-foss-2016b……
```

7. `$ qsub sub.sh`

8. `$ qstat_me or qdel JobID`

GACRC Wiki http://wiki.gacrc.uga.edu

Running Jobs: https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_Sapelo2

Monitoring Jobs: https://wiki.gacrc.uga.edu/wiki/Monitoring_Jobs_on_Sapelo2

Job Submission Queue: https://wiki.gacrc.uga.edu/wiki/Job_Submission_Queues

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

Transfer File: https://wiki.gacrc.uga.edu/wiki/Transferring_Files

Linux Command: https://wiki.gacrc.uga.edu/wiki/Command_List

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

User Account Request: https://wiki.gacrc.uga.edu/wiki/User_Accounts

GACRC Support

https://uga.teamdynamix.com/TDClient/Requests/ServiceCatalog?CategoryID=11593

➢ 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 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

Services (11)

Account Creation
For a research group's PI to request user accounts for group members on the GACRC computing systems.

Class Account Creation
For an instructor to request user accounts for students attending a course that will need to use GACRC computing systems.

Class Account Modification
For instructors to request changes to be made in previously requested class account.

Computing Lab Modification/Deletion

General Internal

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

Lab Creation
For a research group's PI to register a computing lab on the GACRC computing systems

Modify/Delete Account
For PIs to request changes in or deletion of user accounts on GACRC computing systems.

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

My Recent Requests

- home directory is not fully provisioned: ss57215
- GACRC Sapelo2 New Lab/Use Account Request 2018-11-14_preTraining
- GACRC Sapelo2 Cluster New Lab/Use Account Request 2018-11-05_preTraining
- provision 5 user accounts for ugahelpdesk group
- GACRC Sapelo2 New Lab/Use Account Request 2018-10-22_preTraining

View All Recent Requests >

Popular Services

- EITS Help Desk Support Request
- MyID Account Request
- Change Request
- 02 Restricted VPN Access
- Terry Classroom & Meeting Room Support

View All Popular Services >

My Recently Visited Services

- Modify/Delete Account
- Class Account Creation

Service - General Support - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Mail - zhuofei@uga.edu | Service - General Support

https://uga.teamdynamix.com/TDClient/Requests/ServiceDet?ID=25844

Related image

UNIVERSITY OF GEORGIA

Search the client portal

Zhuofei Hou

Home | IT Help Desks | Projects/Workspaces | **Services** | Knowledge Base | News

Project Requests | Ticket Requests | My Favorite | My Recent | My Approvals | Services A-Z | Search

Service Catalog / Academics, Learning & Research / GACRC Service Catalog / General Support

General Support

If you do not have a myid, please mail gacrc-help@uga.edu, and we will respond promptly.

The purpose of this form is to provide a method to report issues and to request help with GACRC systems.

Please use this form for all questions and support needs (e.g. to report issues, to troubleshoot jobs, to request resources or grant writing help, etc). Please do not use this form for software installation requests or lab/user account management, which all have separate forms.

Please refer to the GACRC documentation for information on GACRC resources, how to connect and transfer files, how to run jobs, installed software list, training schedule, and a FAQ.

The link to this documentation is https://wiki.gacrc.uga.edu

Request Service

Share

Add to Favorites

Click to request

This site is operated by Enterprise Information Technology Services (EITS) at the University of Georgia.

Privacy | Accessibility | Website Feedback

https://uga.teamdynamix.com/TDClient/Requests/ServiceCatalogSearch

General Support - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Mail - zhuofei@uga.edu | General Support

https://uga.teamdynamix.com/TDClient/Requests/TicketRequests/NewForm?ID=MNeY5EiDcMY_

Related image

UNIVERSITY OF GEORGIA

Search the client portal

Zhuofei Hou

Home | IT Help Desks | Projects/Workspaces | Services | Knowledge Base | News

Service Catalog / Academics, Learning & Research / GACRC Service Catalog / General Support

General Support

+ Show Help − Hide Help

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

Short Description *

Email *

MyID *

Phone Number *

Support Needed For

- [ ] Galaxy
- [ ] Sapelo2
- [ ] Teaching Cluster
- [ ] Work Filesystem
- [ ] Home Filesystem
- [ ] Scratch Filesystem
- [ ] Project Filesystem
- [ ] Xfer Nodes
- [ ] Other

Lab *

Appendix: Examples of Batch Serial/Threaded/MPI Job Scripts

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

- Components you need to run a job:
  - Software already installed (cluster software or the one installed by yourself)
  - Job submission script to
    1. specify computing resources:
       - ✓ number of nodes and cores
       - ✓ amount of memory
       - ✓ maximum wallclock time
    2. load software using ml load (for cluster software)
    3. run any Linux commands you want to run, e.g., pwd, mkdir, cd, echo, etc.
    4. run the software
  - Input data for analysis, if have
- Common queueing commands you need:
  - qsub, qstat_me, qstat, qdel
  - qstat -f, showq

Example 1: Serial job script running NCBI Blast+ using 1 CPU

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -N testBlast
#PBS -l nodes=1:ppn=1
#PBS -l mem=20gb
#PBS -l walltime=48:00:00

cd $PBS_O_WORKDIR

ml load BLAST+/2.6.0-foss-2016b-Python-2.7.14

time blastn [options] …
```

- → Linux default shell (bash)
- → Queue name (batch)
- → Job name (testBlast)
- → Number of nodes (1), number of cores (1), node feature is NOT needed!
- → Maximum amount of RAM memory (20 GB) is enforced by the cluster!
- → Maximum wall-clock time (48 hours) for the job, default 6 minutes
- → Compute node will use the directory from which the job is submitted as the working directory, i.e., /lustre1/MyID/workDir
- → Load the module of ncbiblast+, version 2.6.0
- → Run blastn with 'time' command to measure the amount of time it takes to run the application

https://wiki.gacrc.uga.edu/wiki/BLAST%2B-Sapelo2

*Example 2: Threaded job script running NCBI Blast+ using 4 CPUS

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -N testBlast
#PBS -l nodes=1:ppn=4
#PBS -l mem=20gb
#PBS -l walltime=480:00:00

#PBS -M jsmith@uga.edu
#PBS -m ae
#PBS -j oe

cd $PBS_O_WORKDIR

ml load BLAST+/2.6.0-foss-2016b-Python-2.7.14

time blastn -num_threads 4 [options] ...
```

➔ Number of nodes (1), number of cores (4)
Number of cores requested (4) = Number of threads (4)

➔ Email address to receive a notification for computing resources
➔ Send email notification when job aborts (a) or terminates (e)
➔ Standard error file (testBlast.e12345) will be merged into standard out file (testBlast.o12345)

➔ Run blastn with 4 threads (-num_threads 4)

*Example 3: MPI job script running RAxML using 2 full nodes

```
#PBS -S /bin/bash
#PBS -q batch
#PBS -N testRAxML
#PBS -l nodes=2:ppn=28
#PBS -l walltime=120:00:00
#PBS -l mem=100gb

cd $PBS_O_WORKDIR

ml load RAxML/8.2.11-foss-2016b-mpi-avx

mpirun –np 50 raxmlHPC-MPI-AVX [options]
```

➔ Number of nodes (2), number of cores (28)
Total cores requested = 2 × 28 = 56
We suggest, Number of MPI Processes (50) ≤ Number of cores requested (56)

➔ To run raxmlHPC-MPI-AVX, MPI version using OpenMPI

➔ Run raxmlHPC-MPI-AVX with 50 MPI processes (–np 50), default 56

Guideline Tips

- Do NOT use Login node to run CPU/memory intensive tasks directly ➔ submit job to queue!
- Do NOT use Login Node to transfer data between your local computer and cluster ➔ use Transfer node!
- Do NOT use Home for storing job data ➔ use /scratch/MyID
- Do NOT park data in Scratch or Local Scratch ➔ clean up when job finished or exits from node
- Do NOT park data permanently in Project ➔ download data to your local drive
- NO large memory job running on batch queue ➔ use highmem_q
- NO small memory job running on highmem_q queue ➔ use batch
- In general, <u>number of threads you want to run with a parallel job = number of cores (ppn) requested</u>

Practical Tips

- Each directory should not have too many files inside! A rule of thumb would be to try to keep no more than a few tens of thousands of files (<10000 would be even better) in any single directory which is accessed frequently

All files are in ONE single dir!

Files are organized in subdirs!

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/