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

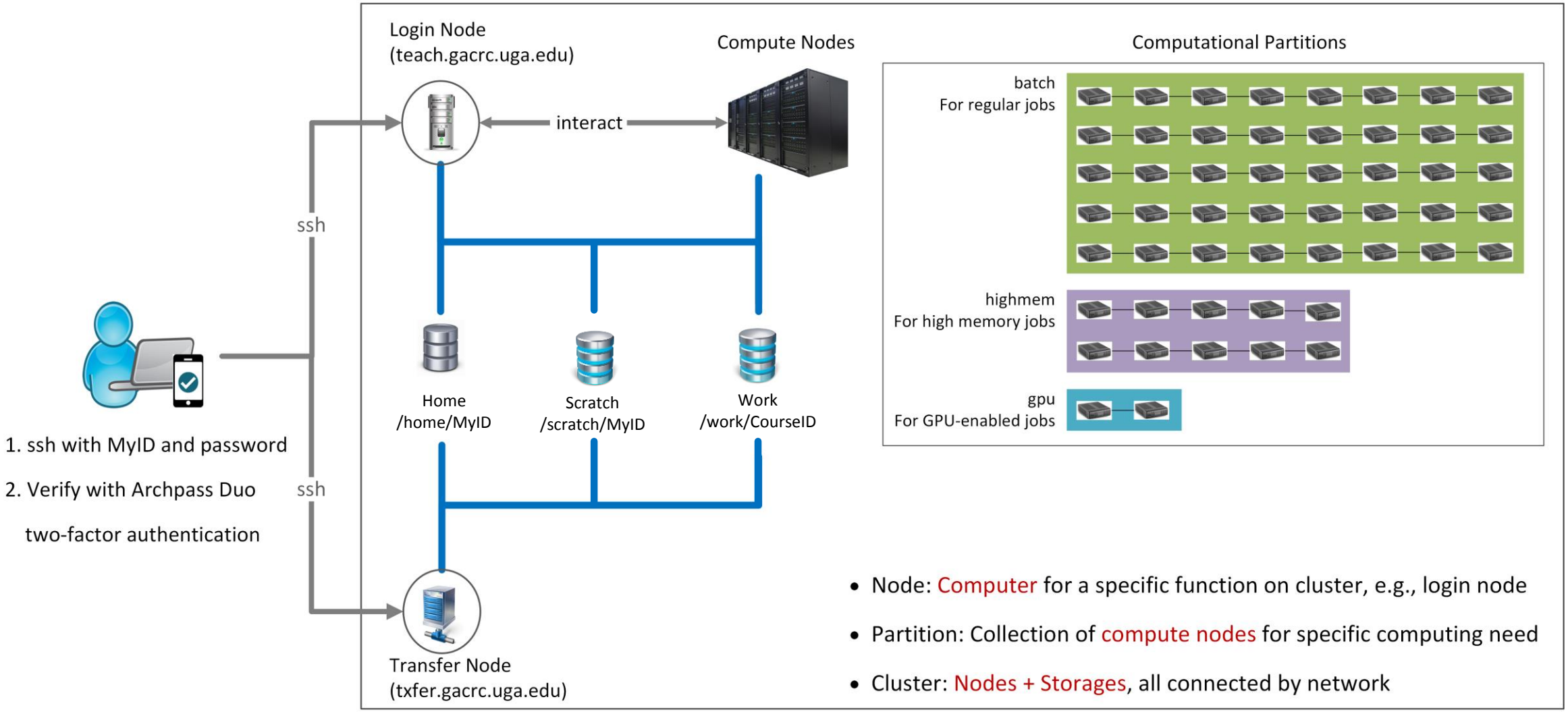
Wiki: http://wiki.gacrc.uga.edu

Support: https://wiki.gacrc.uga.edu/wiki/Getting_Help

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

Kaltura Channel: https://kaltura.uga.edu/channel/GACRC/176125031

Teaching Cluster



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

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 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 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:
   `ssh MyID@teach.gacrc.uga.edu`
2. Change directory to /scratch directory: `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 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 : `nano sub.sh`
7. Submit a job from workDir : `sbatch sub.sh`
8. Check job status : `squeue --me` or Cancel a job : `scancel JobID`

Step1: 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/

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 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…

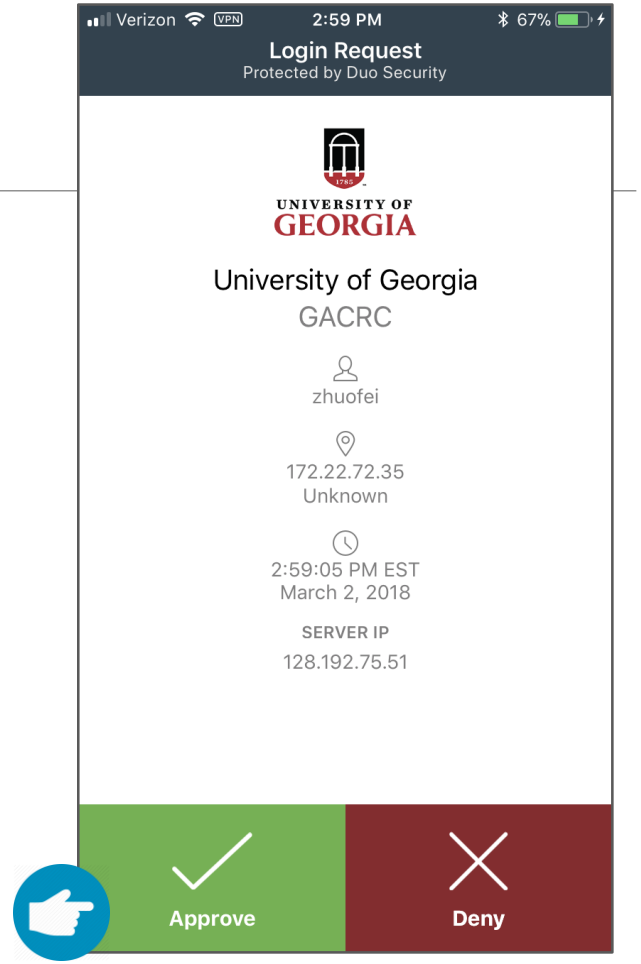
Password: ← 2. Enter your MyID password
When you enter password, no stars or dots will show as you are typing. Please type password carefully!

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

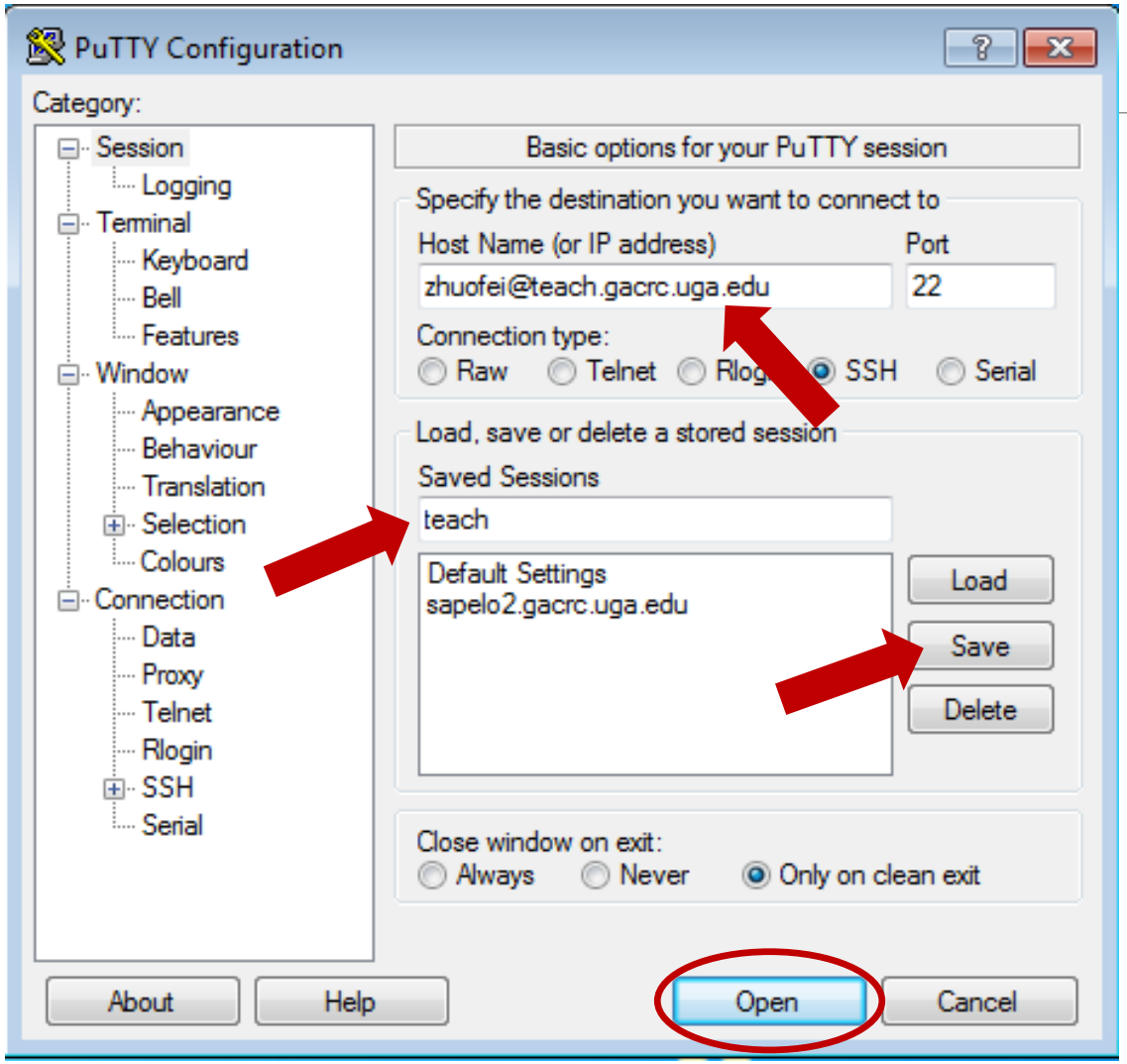
Step1 (Cont.) - Windows using PuTTY

1. Download and install PuTTY: https://www.putty.org/
2. Detailed downloading and installation instructions:

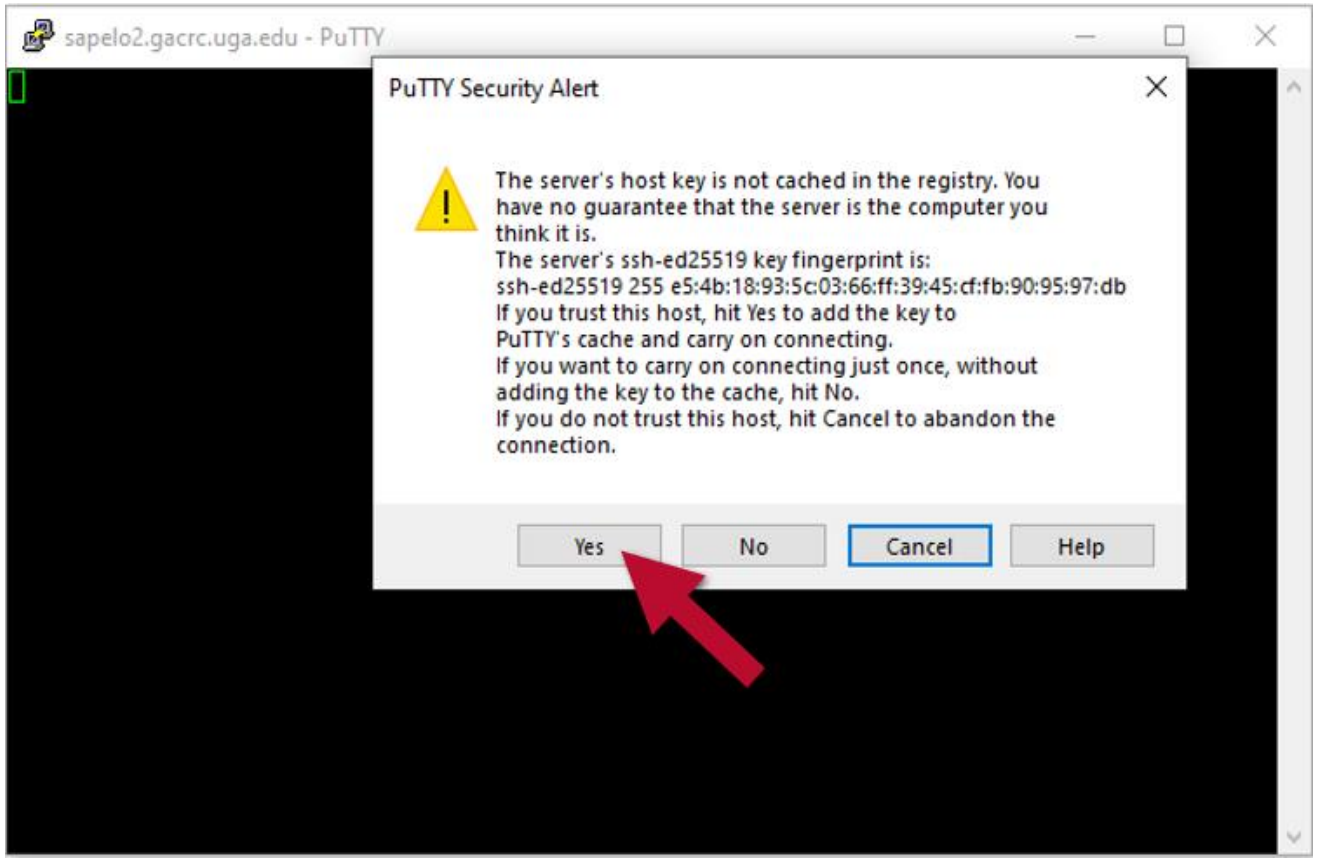
   https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY
3. Detailed configuring and usage instructions:

   https://wiki.gacrc.uga.edu/wiki/How_to_Install_and_Configure_PuTTY#Configuring_PuTTY

Step1 (Cont.) - Windows using PuTTY

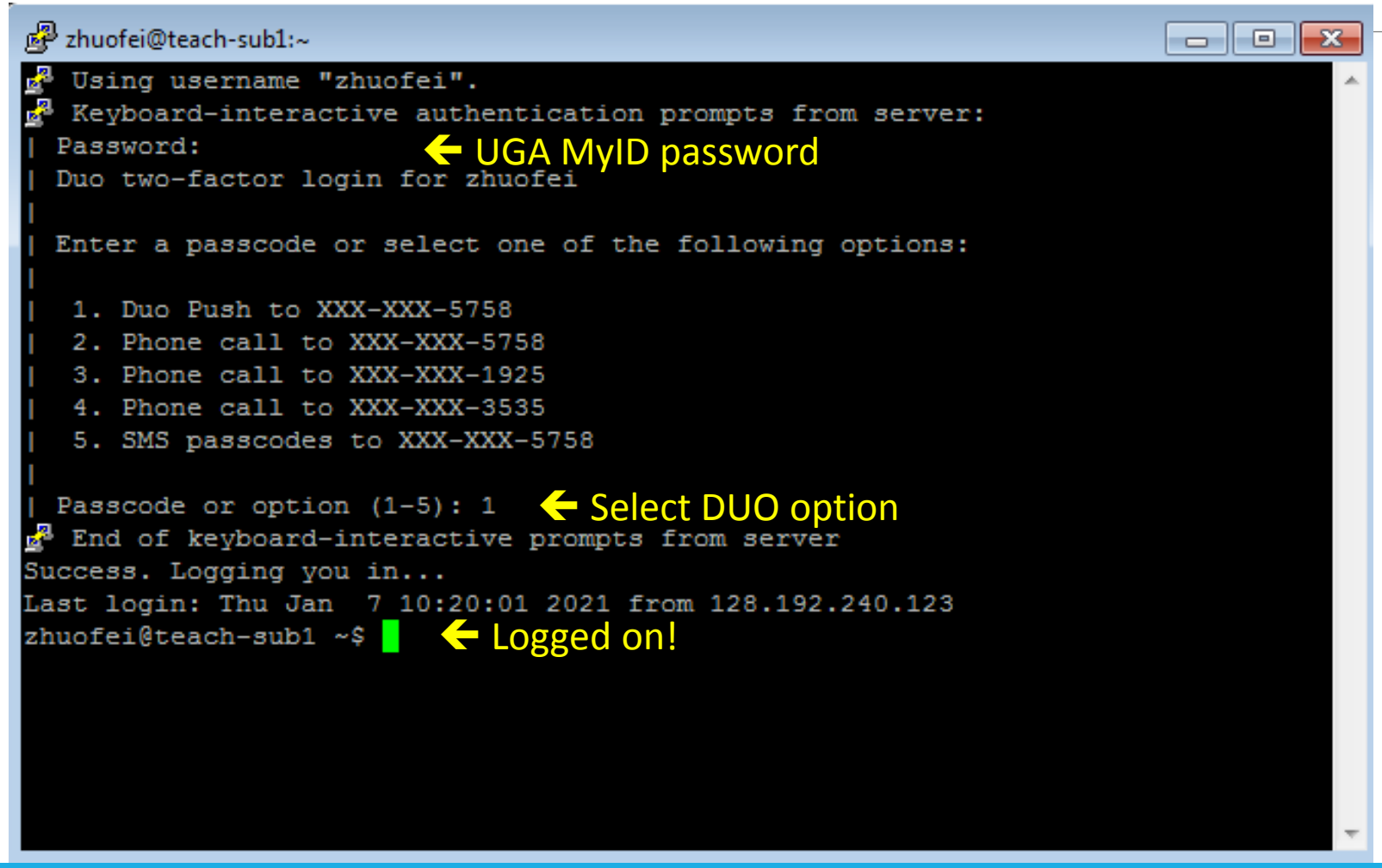


The first time you connect to login node, PuTTY will give you this security alert window. Please click "Yes"



Step1 (Cont.) - Windows using PuTTY

Next you will enter your UGA MyID password and initiate DUO authentication procedure:



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 .
```

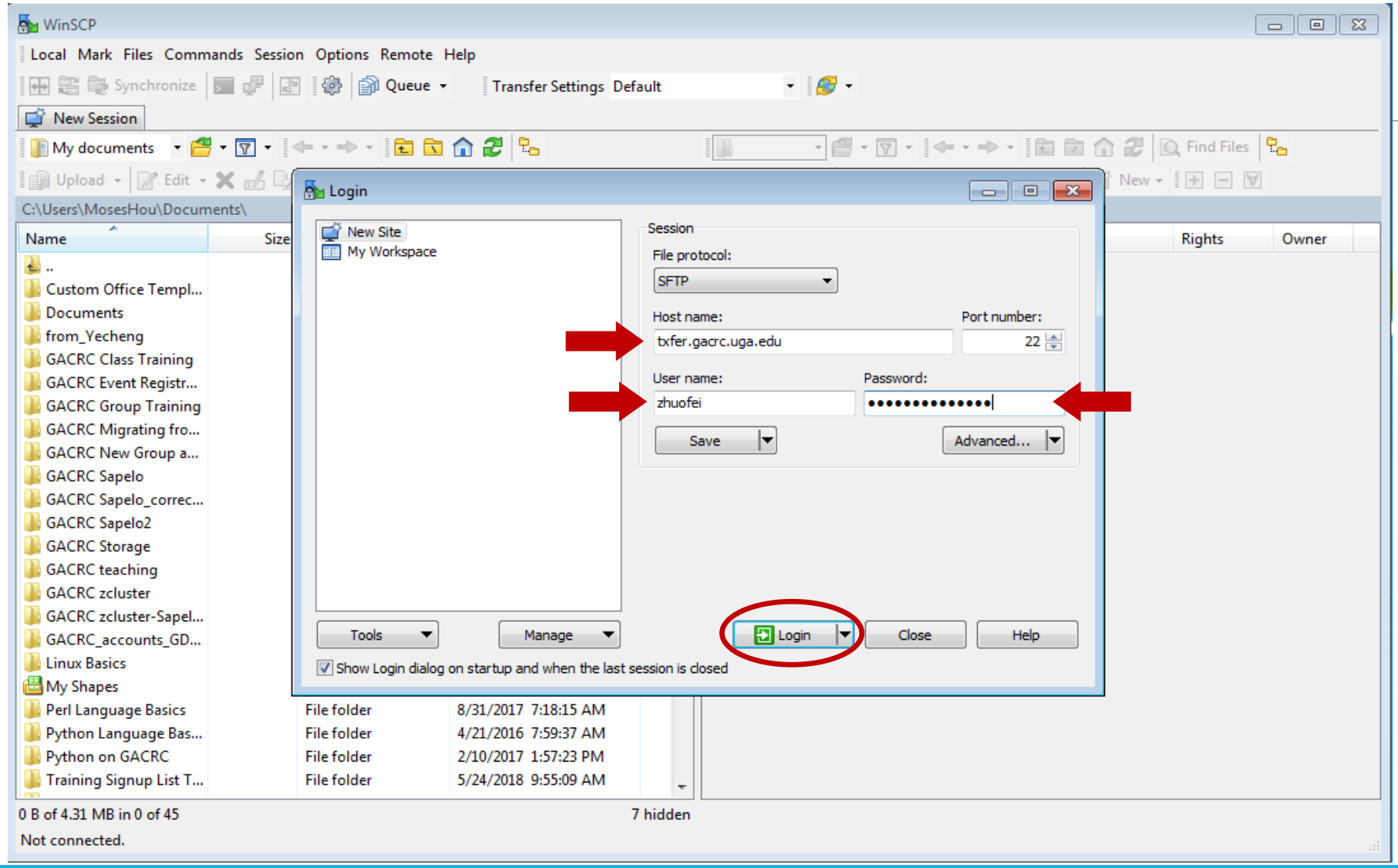
```
scp -r zhuofei@txfer.gacrc.uga.edu:/home/zhuofei/workDir/folder/ .
```

Step 5 (Cont.) - Windows using WinSCP

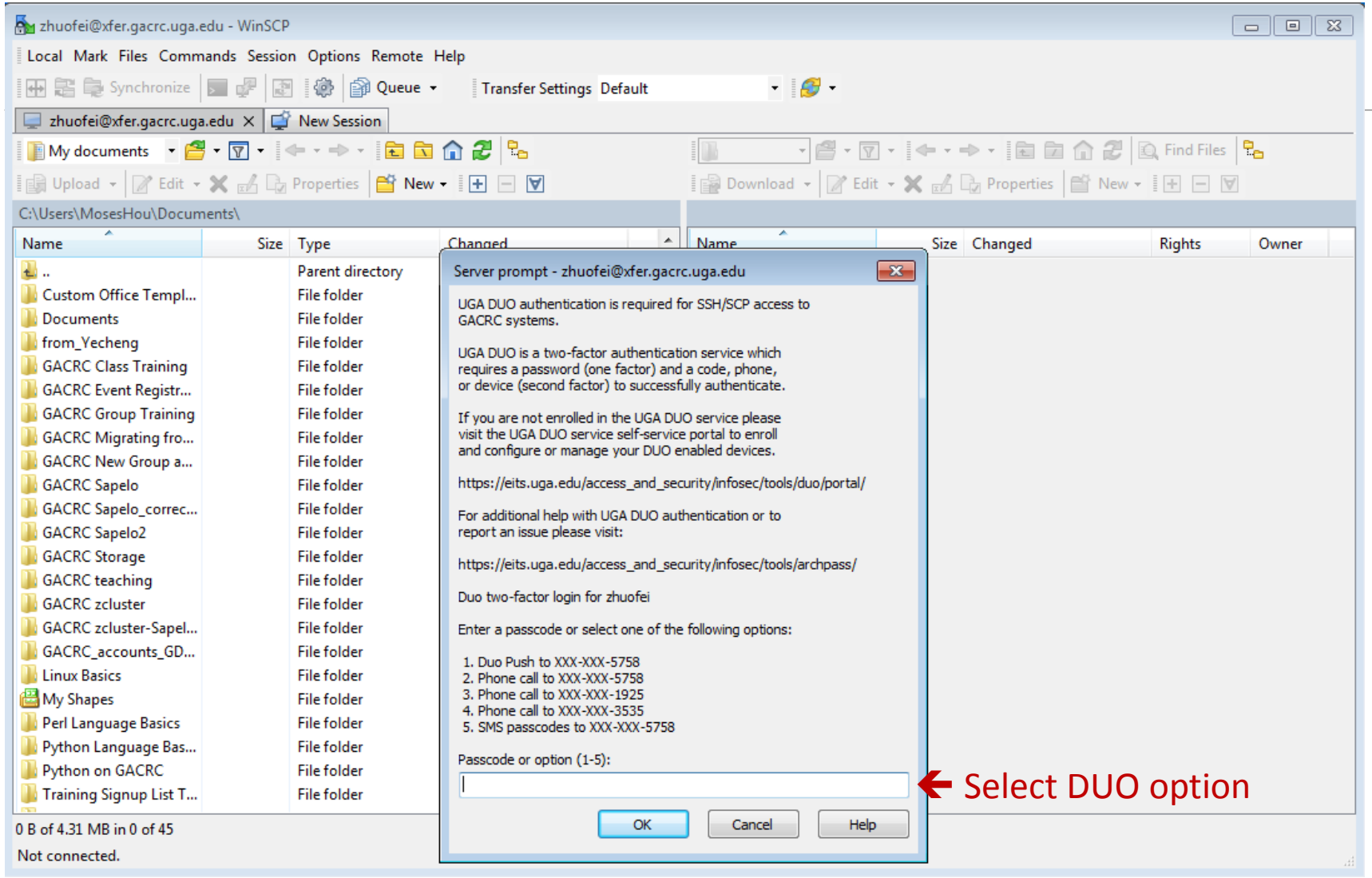
https://wiki.gacrc.uga.edu/wiki/Transferring_Files#Using_WinSCP_2

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
3. Alternative FileZilla https://wiki.gacrc.uga.edu/wiki/Transferring_Files#Using_FileZilla_2

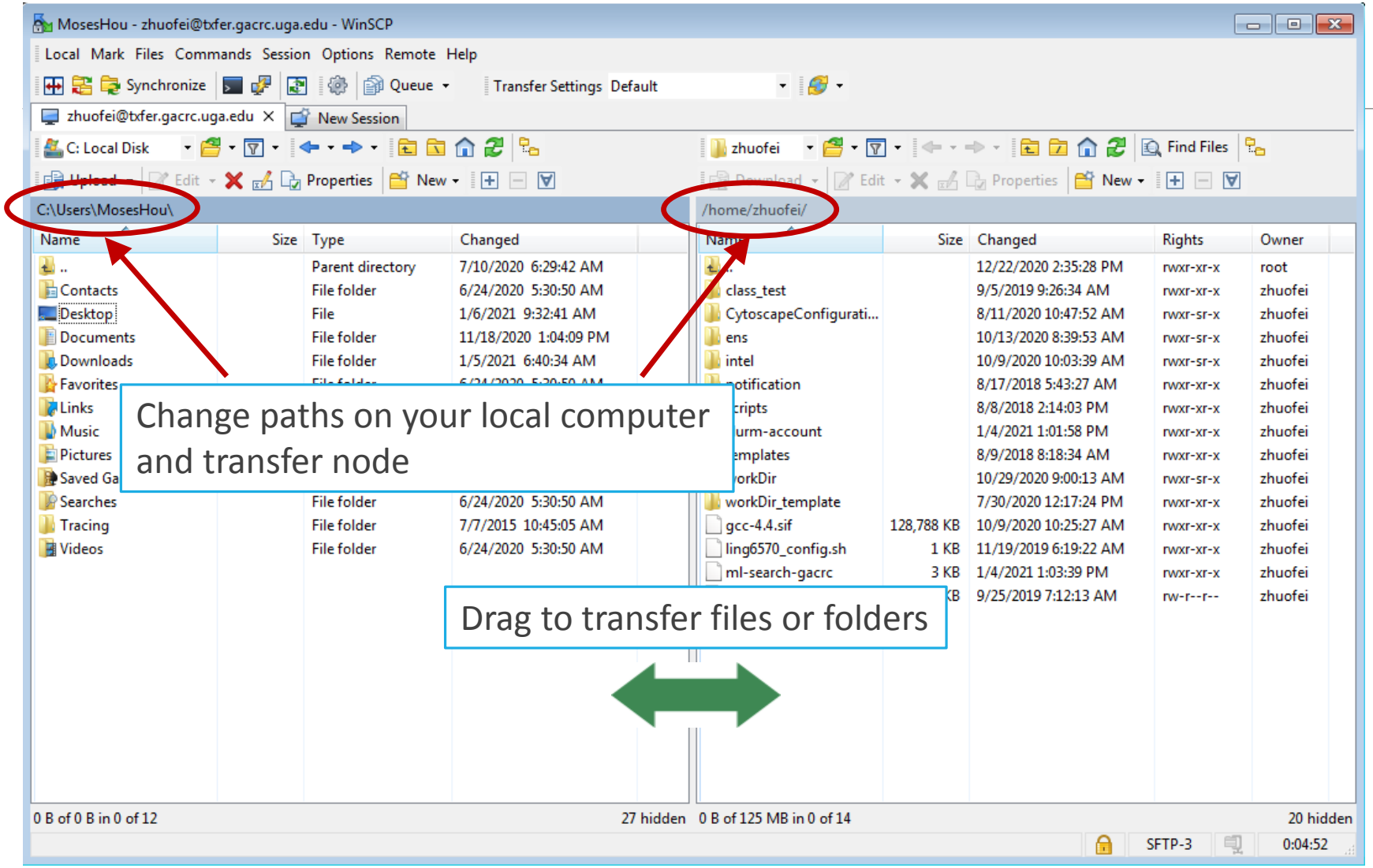
Step 5 (Cont.) - Windows using WinSCP



Step 5 (Cont.) - Windows using WinSCP



Step 5 (Cont.) - Windows using WinSCP



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
```

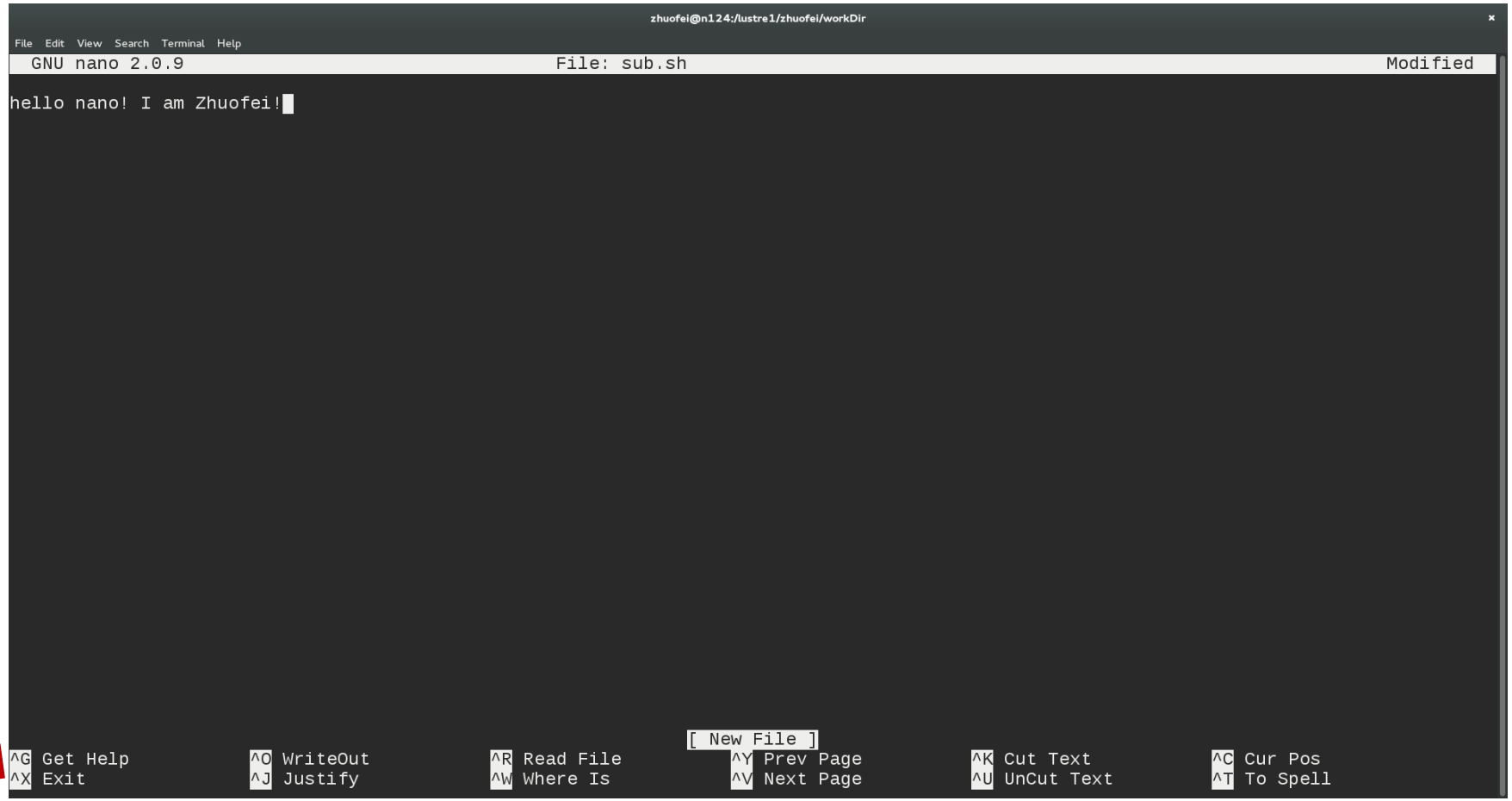
Step 6: Make a job submission script in workDir using nano

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.)

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 .

```
#!/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
```

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

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_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 8: Check job status using squeue

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

```
$ squeue --me -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 tcn18
161       batch   testBLAS   zhuofei   RUNNING   0:02   2:00:00      1 tcn18

$ squeue --me
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 tcn18
161       batch   testBLAS   zhuofei  R   0:17      1 tcn18
```

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 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

```
$ squeue --me -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 tcn18
168   batch   testBLAS  zhuofei  RUNNING   3:14    2:00:00      1 tcn19

$ 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 tcn18
168   batch   testBLAS  zhuofei  RUNNING   3:32    2:00:00      1 tcn19

$ squeue --me -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 tcn19
```

Step8 (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

JobID  JobName  User     Partition  NodeList  AllocNodes  NTask  NCPUS  ReqMem  MaxVMSize  State      CPUTime   Elapsed   Timelimit  ExitCode
------ -------- -------- ---------- --------- ----------- ------ ------ ------- ---------- ---------- --------- --------- ---------- --------
33562  testBLAST zhuofei  batch      tcn18     1           4              20Gn               COMPLETED  00:27:48  00:06:57  02:00:00   0:0

$ seff 33562        # Check computing resources used by a COMPLETED job
Job ID: 33562
Cluster: gacrc-teach
User/Group: zhuofei/gacrc-instruction
State: COMPLETED (exit code 0)
Nodes: 1
Cores per node: 4
CPU Utilized: 00:03:33
CPU Efficiency: 12.77% of 00:27:48 core-walltime
Job Wall-clock time: 00:06:57
Memory Utilized: 15.65 GB
Memory Efficiency: 78.29% of 20.00 GB
```

Step 8 (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

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
```

GACRC Wiki http://wiki.gacrc.uga.edu
Kaltura Channel https://kaltura.uga.edu/channel/GACRC/176125031

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:

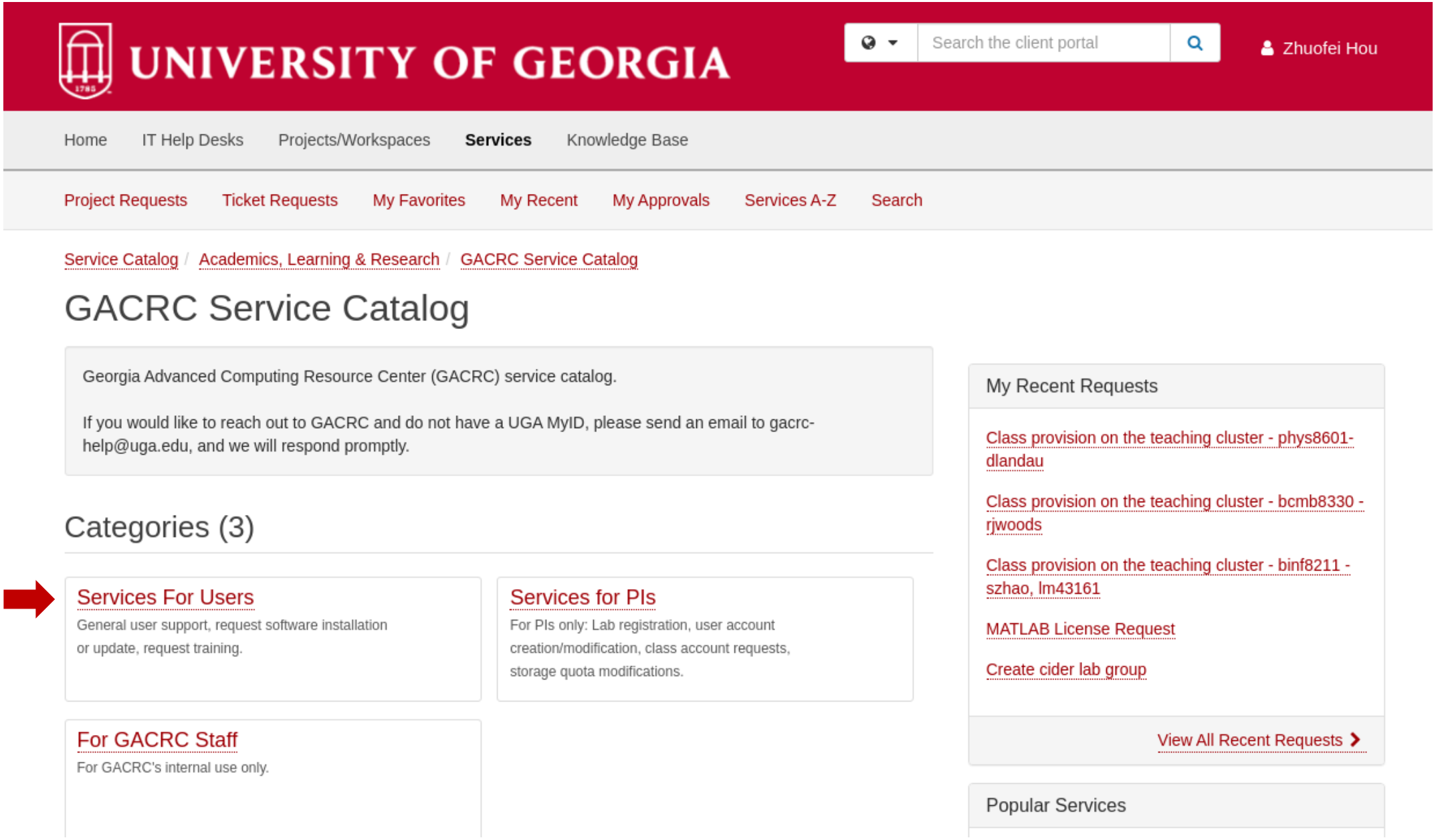
Please tell us details of your question or problem, including but not limited to:

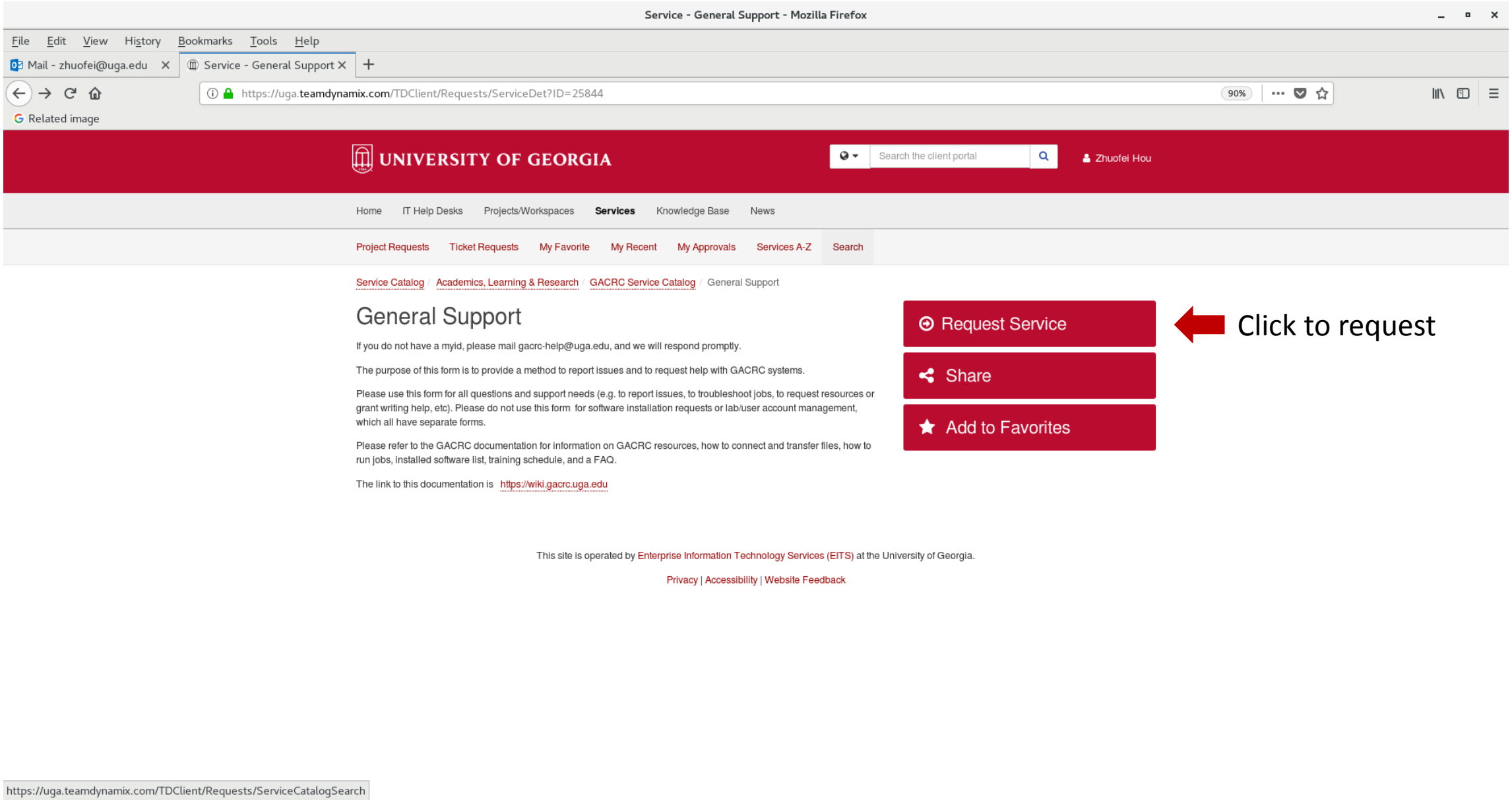
- ✓ Your user name
- ✓ Your job ID
- ✓ Your working directory
- ✓ The partition name and command you used to submit the job

➢ 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!


UNIVERSITY OF GEORGIA
Search the client portal
Zhuofei Hou
Home
IT Help Desks
Projects/Workspaces
Services
Knowledge Base
Project Requests
Ticket Requests
My Favorites
My Recent
My Approvals
Services A-Z
Search
Service Catalog / Academics, Learning & Research / GACRC Service Catalog
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.
My Recent Requests
Class provision on the teaching cluster - phys8601-dlandau
Class provision on the teaching cluster - bcmb8330 - rjwoods
Class provision on the teaching cluster - binf8211 - szhao, lm43161
MATLAB License Request
Create cider lab group
View All Recent Requests
Popular Services


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
90%
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

Need Support? http://help.gacrc.uga.edu