

# Lab 3: Getting started with ROGER

## 1. Outline

In this lab, we will interact with ROGER from various aspects, including its module systems, HPC batch system, and Hadoop/Spark system. We will exercise some of the basic Linux commands to create directly, (remotely) copy data (from and to ROGER). In particular, we will practice how to allocate computation resources via qsub and submit a job with a PBS job script via qsub. Last but not least, we will get to know how to access Hadoop system and some basic file operations.

## 2. Getting familiar with basic commands

- `cd` change directory
  - `pwd` displays your current directory
  - `ls` list the files in your current directory
  - `ls test*` list all of the files starting with “test” in the directory
  - `ls -l` list the files in a directory with details
  - `mkdir name` create a directory
  - `rm` remove (delete) file
  - `rm -r` remove a directory
  - `cp origin destination` copy a file
  - `cp -r origin destination` copy a directory and its contents recursively
  - `mv origin destination` move or rename a file
  - `chmod g+rw filename` set the file to be readable and writeable by group members
  - `module avail` list all available modules
  - `module load module_name` load a specific module
  - `ssh cg-hm08` switch to Hadoop cluster
  - `hdfs dfs -ls directory` list the files in the specified directory of the Hadoop File System
- Note: Hadoop has no concept of a “current” directory
- `scp path/file_name netID@roger-login.ncsa.illinois.edu:path` copy a local file to ROGER
  - `scp netID@roger-login.ncsa.illinois.edu:path\_to\_file destination_path` copy file from ROGER to local file system

## 3. Preparing the dataset

First, you need to login to ROGER with the user name and password

```
>> ssh NetID@roger-login.ncsa.illinois.edu
```

Make a directory for this lab, e.g. lab3. The following command creates the “lab3” folder in your home directory

Notes: In Linux system, the symbol `~` is equivalent to home directory, where the full path is `/home/your_netID`. Whenever you are lost in the file system, type command “`pwd`” to show your current directory.

```
>> mkdir lab3
```

Copy the data from /gpfs\_scratch/geog479/lab3/

```
>> cp -r /gpfs_scratch/geog479/lab3/* ./
```

Or

```
>> cp -r /gpfs_scratch/geog479/lab3/* ~/lab3/
```

Or

```
>> cp -r /gpfs_scratch/geog479/lab3/* /home/NetID/lab3/
```

Make sure you have two folders in your lab3 folder

```
>> ls
```

In particular, if you want to check the size of file:

```
>> cd data
```

```
>> ls -l (list all the files in the current directory with detailed information)
```

```
>> ls -sh (sh will put the file size in human readable format)
```

## 4. Remote server visualization via XServer

Let's back to the lab3 folder and play with the XServer by launching firefox from ROGER

```
>> cd ../ (assuming you are currently at the data folder and want to jump to the parent folder)
```

Or

```
>> cd ~/lab3/
```

```
>> firefox/firefox --no-remote
```

Note that because MobaXterm has already enabled XServer capabilities, you can directly see the firefox browser pops up. To use regular command console with XServer, you need to specify the XServer capability when you login to ROGER

```
>> ssh -X NetID@roger-login.ncsa.illinois.edu
```

## 5: Remotely copy data from/to ROGER

The command you will use is called scp:

- copy data from ROGER to local disk  
open a new tab in MobaXterm

```
>> scp NetID@roger-login.ncsa.illinois.edu:~/lab3/data/dem3/u39w120.tif Desktop/
```

Now, open QGIS or ArcGIS to view the DEM data (u39w120.tif), which we will process later.

- copy data from local disk to ROGER

you can choose any file to upload to ROGER, in this case, let's create a text file (example.txt)

```
>> nano example.txt
```

Type any words you would like and save it

```
>> scp example.txt NetID@roger-login.ncsa.illinois.edu:~/lab3/
```

Now, switch back to your ROGER login and see the copied file.

## 6: Running jobs/tasks via QSUB

There are, in general, two ways to use QSUB to allocate resources in ROGER: qsub interactive and qsub script using PBS (portable batch system: <https://hpcc.usc.edu/support/documentation/running-a-job-on-the-hpcc-cluster-using-pbs/>).

- qstat usage

```
>> qstat
```

```
>> qstat -f
```

```
>> qstat -f job_id
```

- Use qsub interactively

```
>> qsub -l -l walltime=0:30:00
```

Notice the change of computing node you are current in: e.g. [netID@computing\_node]

Let's run TauDEM in ROGER to process the DEM files

You need to load certain modules in order to, e.g., run MPI programs or use specific libraries

```
>> module load mpich taudem
```

Make a directory for output file (note: it is required by taudem)

```
>> mkdir ~/lab3/output
```

Here, let's compare the difference of using MPI and without MPI

### **Normal (parallel) execution**

```
>> pitremove -z $HOME/lab3/data/dem3/ -fel $HOME/lab3/output/
```

Once it is done, check the content in ~/lab3/output

## MPI execution

```
>>mpirun -n 20 pitremove -z $HOME/lab3/data/dem3/ -fel $HOME/lab3/output/
```

(You may encounter some message saying there was an error, it is alright in this case, the installed taudem is outdated).

Again, check the content in ~/lab3/output

Now, let's rename the file and copy it to our desktop to visualize the difference between the original DEM file.

```
>> cd ~/lab3/output
```

```
>> mv file_name pitremoved.tif
```

Open a new tab in MobaXterm (why we do it in this way?)

```
>> cd Desktop
```

```
>> scp NetID@roger-login.ncsa.illinois.edu:~/lab3/output/pitremoved.tif ./
```

## 7. Writing and Submitting (HPC batch) job script

A common approach to submit and run your jobs is through PBS script

An example script is placed in ~/lab3/script

```
>> cd ~/lab3/script
```

View the content

```
>> nano ~/lab3/script/job.sh
```

Edit the script, such as job function, your email address, etc.

Submit your job

```
>> qsub job.sh
```

Job submitted! Now you can check the status

```
>> qstatq
```

Note that if you want to write your own script from scratch, make sure your script is executable. For example: `chmod +x script_name.sh`

## 8. Check ROGER statistics

Launch your preferred web browser, and visit the following two links:

1: <http://roger-stats.ncsa.illinois.edu>

2: <http://clustat.ncsa.illinois.edu/roger/>

## 9. Hadoop file system

We have enabled your ROGER accounts to access Hadoop cluster. To access the Hadoop cluster

```
>> ssh cg-hm08
```

You can checkout the directory in the Hadoop system

```
>> hdfs dfs -ls
```

By default, the command above only checks this directory `/user/NetID/`, you can issue this command to see all the users:

```
>> hdfs dfs -ls /user/
```

If you do not have folder yet, you need to create a folder the same name as your NetID

```
>> hdfs dfs -mkdir /user/netID
```

And now check it again

```
>> hdfs dfs -ls
```

It is currently empty, right? Let's pop some data into the file system

```
>> hdfs dfs -copyFromLocal your_file /user/NetID/file_name_of_your_choice
```

Or simply

```
>> hdfs dfs -copyFromLocal your_file
```

Now you want to download the file in Hadoop File System to your local directory:

```
>> hdfs dfs -getmerge file_path_in_hdfs path_to_local_directory
```

For example:

```
>> hdfs dfs -getmerge /user/NetID/example.txt example2.txt
```

Lastly, since we are sharing the resources with many other users in ROGER, let's keep the file system tidy and try not to waste the spaces, meaning delete files if you will not use them in the future.

```
>> hdfs dfs -rm -r /user/file_name
```