

Lab: High-performance computing with Python 2

High-performance computing with Python

Lab Exercise 2

1 Access to the cluster

Download Putty from <ftp://130.39.13.164/putty.exe>

Double click Putty, and then click 'Run' to open its interface. Type in the hostname and choose SSH connection type. You can define a name for this connection in 'Saved session' and click 'Save' to save this connection, so that you don't need to type in them again next time you log in. Then, you can click 'Open' to connect to the cluster. Click 'yes' when you see alert.

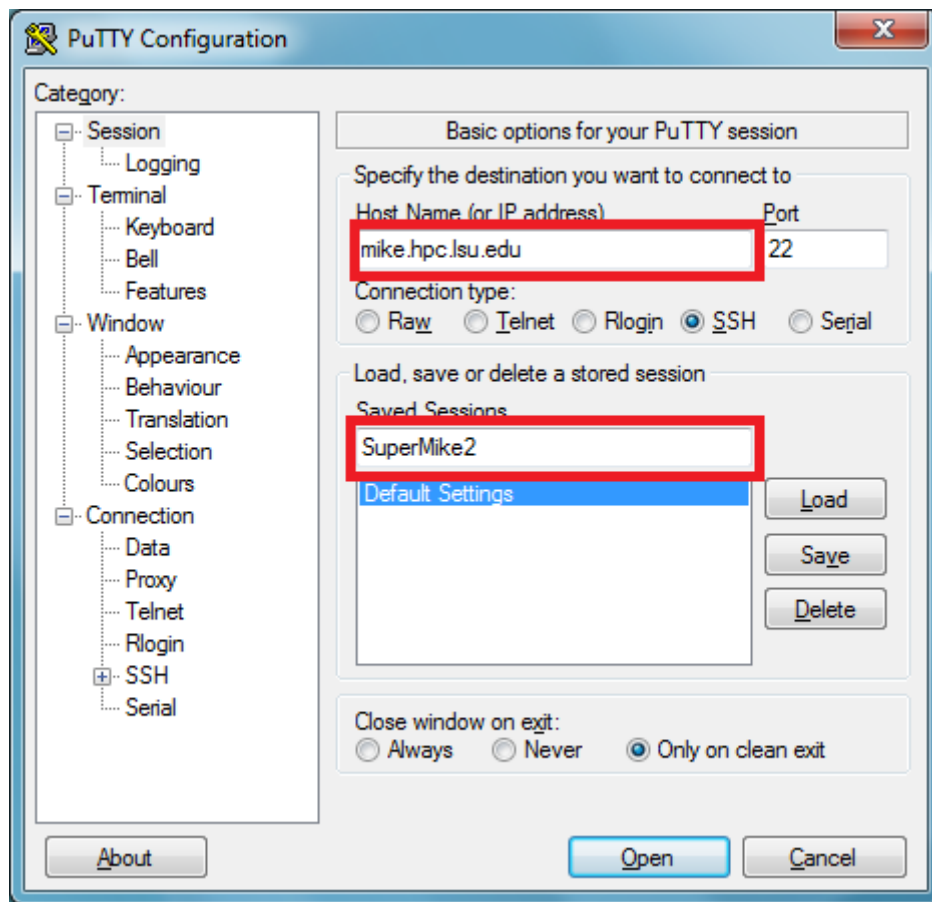
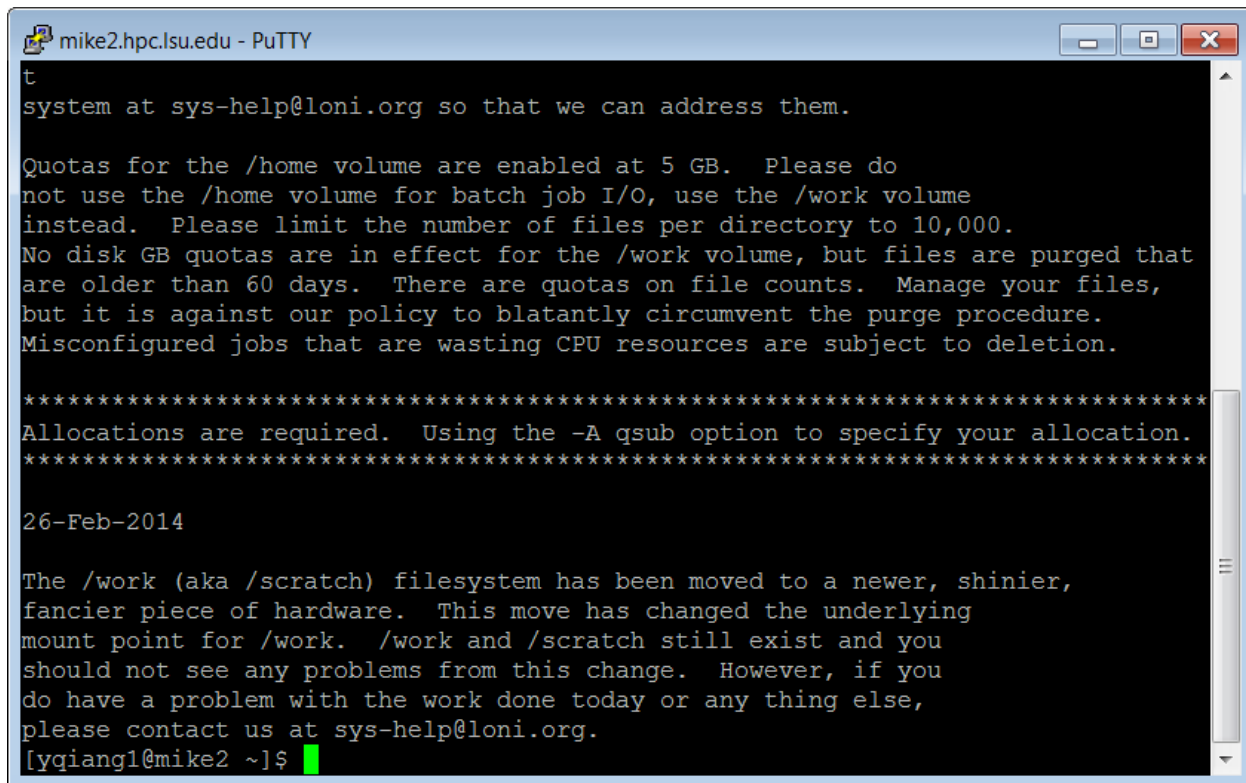


Figure 1. Input connection information in Putty.

Next, you will see a command line interface that asks for your username and password. Type in the temporary username and password that you have. If you see messages similar to the following figure, you have logged into the head node of the SuperMike2 cluster of LSU HPC.



```
mike2.hpc.lsu.edu - PuTTY
t
system at sys-help@loni.org so that we can address them.

Quotas for the /home volume are enabled at 5 GB. Please do
not use the /home volume for batch job I/O, use the /work volume
instead. Please limit the number of files per directory to 10,000.
No disk GB quotas are in effect for the /work volume, but files are purged that
are older than 60 days. There are quotas on file counts. Manage your files,
but it is against our policy to blatantly circumvent the purge procedure.
Misconfigured jobs that are wasting CPU resources are subject to deletion.

*****
Allocations are required. Using the -A qsub option to specify your allocation.
*****

26-Feb-2014

The /work (aka /scratch) filesystem has been moved to a newer, shinier,
fancier piece of hardware. This move has changed the underlying
mount point for /work. /work and /scratch still exist and you
should not see any problems from this change. However, if you
do have a problem with the work done today or any thing else,
please contact us at sys-help@loni.org.
[yqiang1@mike2 ~]$
```

Figure 2. Logged in to the SuperMike2 cluster

You can use the commands you learned in the class to browse over the directory hierarchy. Then, type the following command to create a folder ex2 under your home directory.

```
mkdir ~/ex2
```

2 Upload your files

Download the datasets and python file for this exercise from ftp://130.39.13.164/ex2_data.zip , and unzip the data files into a local directory in your computer. This zip package contains: *land_water_la.asc* and *dem_la.asc*), their projection files, 2 Python programs (*ex2_serial.py* and *ex2_parallel.py*) and a PBS job script (*pbs_ex2*).

Download Winscp from <ftp://130.39.13.164/winscp.exe> to your computer.

Double click Winscp.exe to open its interface, and fill in the connection information like below:

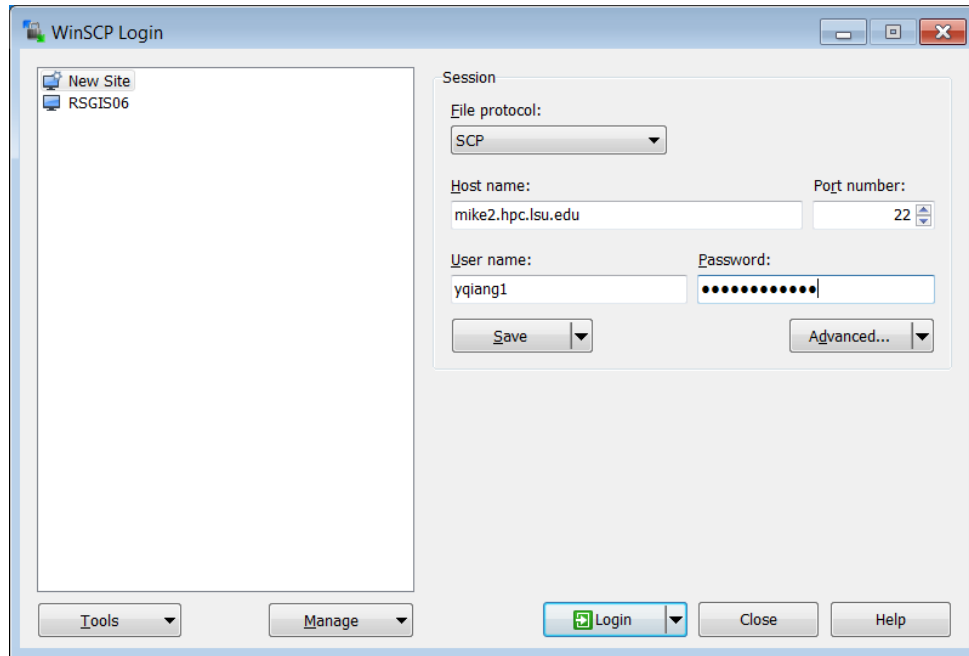


Figure 3. Fill in connection information in Winscp

Click 'Login' on the bottom and click 'Yes' when you see alert. Once connected to the cluster, you will see the transfer interface like below:

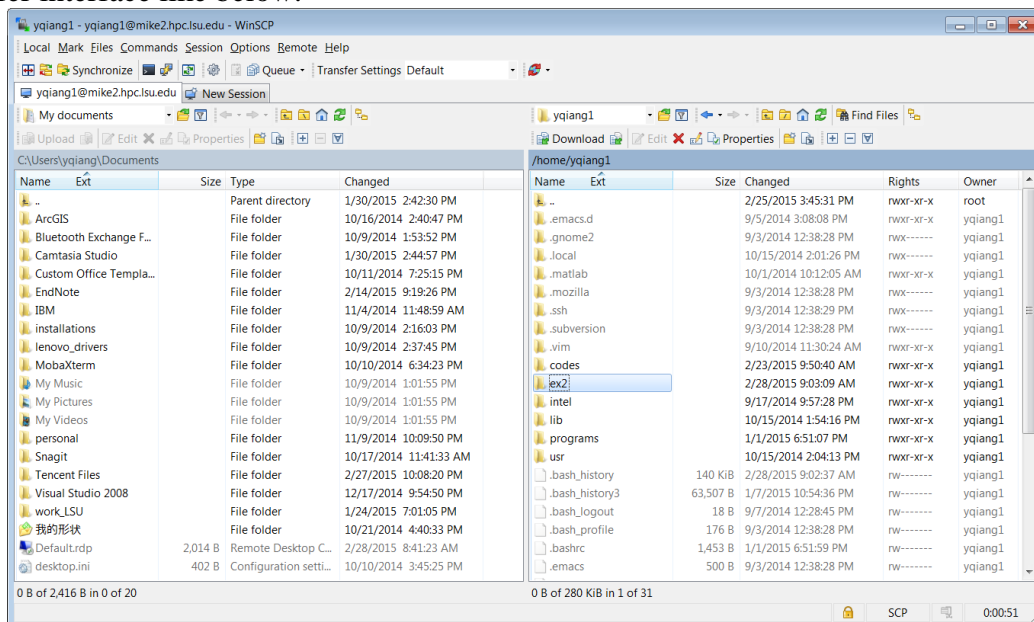


Figure 4. Fill in connection information in Winscp

Browse to the folder that contains the downloaded datasets in the left panel (your local directory). In the right panel, browse to the 'ex2' directory you have created in the cluster. Then, drag all the downloaded files from your local computer into the ex2 directory in the cluster.

3 Run your program in an interactive session

The raster files cover a coastal area of Louisiana and are in 30 meter resolution. The *land_water_la.asc* file is the land-water classification in 2010, in which the pixel value of 1 represents land while 0 represents water. The *dem_la.asc* file is a DEM elevation dataset derived from LiDAR images, in which the pixel value represents elevation in meter above the sea level. You can visualize these datasets with ArcGIS. This study area is faced with serious land loss problem caused by multiple natural and human processes. It is also a part of the study area of a NSF project ‘Modeling Coupled Natural and Human Dynamics in a Vulnerable Coastal System’, led by Dr. Nina Lam (<http://www.rsgis.envs.lsu.edu/cnh/>).



Figure 5. The study area

The Python program is a simplified sea level rise simulation model, which simulate the land loss scenario when the sea level rises 0.5 meters. This model assumes that a land pixel will become water if it meets the following two conditions:

- (1) the elevation of the pixel is below 0.5 meter, and
- (2) the pixel has at least 1 water pixel in its adjacent neighborhood.

Now, go back the command line interface of Putty, and request for a 1-hour interactive session with 1 core in 1 compute node in the single queue by typing:

```
qsub -I -V -l walltime=1:00:00,nodes=1:ppn=1 -A hpc_train_2014 -q single
```

You will see the following message when the requested interactive session is allocated to you.

```

Note: Balance and Deposit are measured in CPU-hours
++++ Allocation HPC_startup not available for user yqiang1 ++++
qsub: submit filter returned an error code, aborting job submission.
[yqiang1@mike2 ex2]$ qsub -I -V -l walltime=0:40:00,nodes=1:ppn=1 -A hpc_startup -q single
qsub: waiting for job 309552.mike3 to start
qsub: job 309552.mike3 ready

-----
Running PBS prologue script
-----
User and Job Data:
-----
Job ID:      309552.mike3
Username:    yqiang1
Group:       Users
Date:        04-Mar-2015 20:52
Node:        mike371 (39889)

PBS has allocated the following nodes:
mike371

A total of 1 processors on 1 nodes allocated

Check nodes and clean them of stray processes
-----
Checking node mike371 20:52:11
-> User guannan running job 309472.mike3:state=R:ncpus=4
-> User yqiang1 running job 309552.mike3:state=R:ncpus=1 (This job)
Done clearing all the allocated nodes
-----
Concluding PBS prologue script - 04-Mar-2015 20:52:11
-----
[yqiang1@mike371 ~]$ █ Instead of mike2, mikeXXX indicates that you are in compute node now

```

Figure 6: You are successfully allocated a compute node

Now, please go the directory `ex2` by typing `cd ~/ex2`, and type the following command to start the serial model program:

```
python ex2_serial.py
```

Leave the Putty window open. The program running may take around 20 minutes. When your program is done, a notice and processing time will be printed. Instead of waiting here, you can move to Section 4.

4 Submit your program as a batch job

To submit a batch job, you need a PBS job script to specify your job path and required resources. You can download a template PBS job script from http://130.39.13.164/pbs_ex2 to your local directory. Right click on the file and choose to open it with WordPad. Then, you will see the job script like below:

```

#!/bin/bash
#PBS -l nodes=1:ppn=16
#PBS -l walltime=0:30:00
#PBS -N yqiang1           #give your job a name as you want
#PBS -o yqiang_output     #change to a file name as you want
#PBS -e yqiang_error      #change to a file name as you want
#PBS -q workq
#PBS -A hpc_train_2014
#PBS -M yqiang1@lsu.edu   #Please change to your email!

python /home/yqiang1/ex2/ex2_par.py           #change to your file path

```

This job script basically requested 1 compute node with 16 cores for 30 minute. The job will be placed in the work queue. Now, make changes to the file according to my comments after #. Save the changes and then upload it to the `ex2` directory in the cluster through Winscp. Click 'Yes' to overwrite the existing one.

Double click putty.exe to open another Putty window. Connect to the cluster like what you did in the beginning. You may double click the save connection if you have saved your previous connection session.

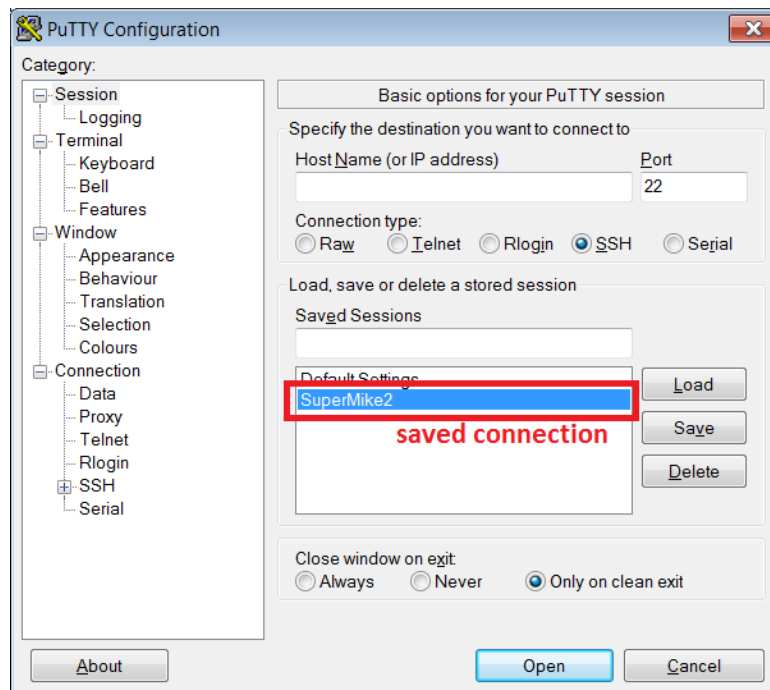


Figure 7. Connect to the cluster by double clicking the saved session.

When you have logged in to the cluster, go to `ex2` directory, and type:

```
qsub pbs_ex2
```

Now, the job is successfully submitted to SuperMike2 cluster. As the model program in `ex2_par.py` has been parallelized to 16 processes, it should finish much sooner.

When you are waiting for your jobs, you can type in following commands to check information about your submitted jobs:

```
qstat -f <jobid> – print details of your submitted batch job
qstat -n -u <user> – print information of nodes assigned to the < user >
checkjob <jobid> – print the running status of your job
qshow -j <jobid> – check the health of your running job
```

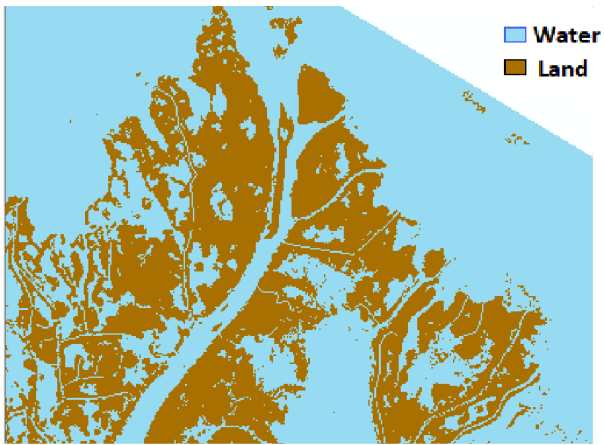
You will receive a notice in your email when the job is done or an error occurs. You can see the processing time in the output file specified in the PBS job script, for example, by typing:

```
cat yqiang_output          #type the output file name in the pbs job file
```

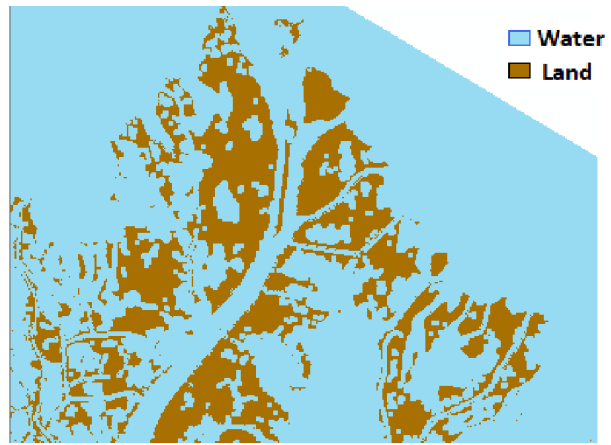
Remember to change `yqiang_output` to your output file name. You may notice that the processing time of the serial program is much shorter than that of the serial program.

5 Download your outputs

When your programs in the interactive session and submitted job are done, you will see the outputs (`land_water_la2.asc` and `land_water_la3.asc`) generated in the `ex2` folder in the cluster. You can download either one (with the corresponding `prj` file) to your computer using Winscp, and load it in ArcGIS to see the difference between the simulation output and the original land cover map (`land_water_la.asc`)



land_water_la.asc



land_water_la2.as

Figure 8: Comparison between original land cover (left) and simulated land cover (right) in the coast