CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY

HPCs at OU

Fabia U. Battistuzzi

battistu@oakland.edu

Tomas Hajek

hajek@oakland.edu

Stephen Villagonzalo

svillagonzalo@oakland.edu

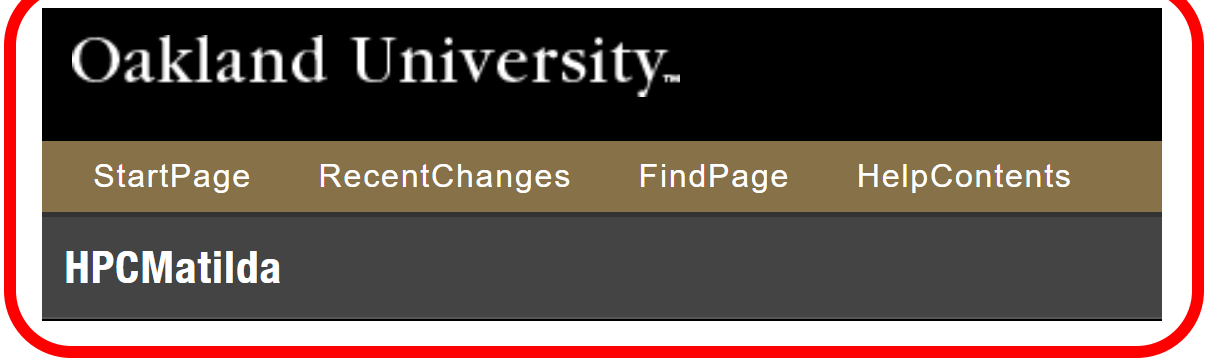
John Johnston

jbjohnston@oakland.edu

CENTER FOR **DATA SCIENCE** AND **BIG DATA** ANALYTICS

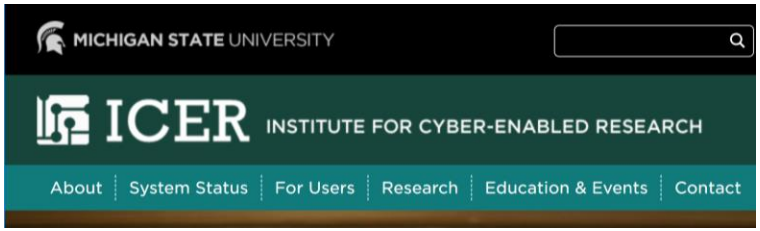
Strength in Numbers

OAKLAND UNIVERSITY



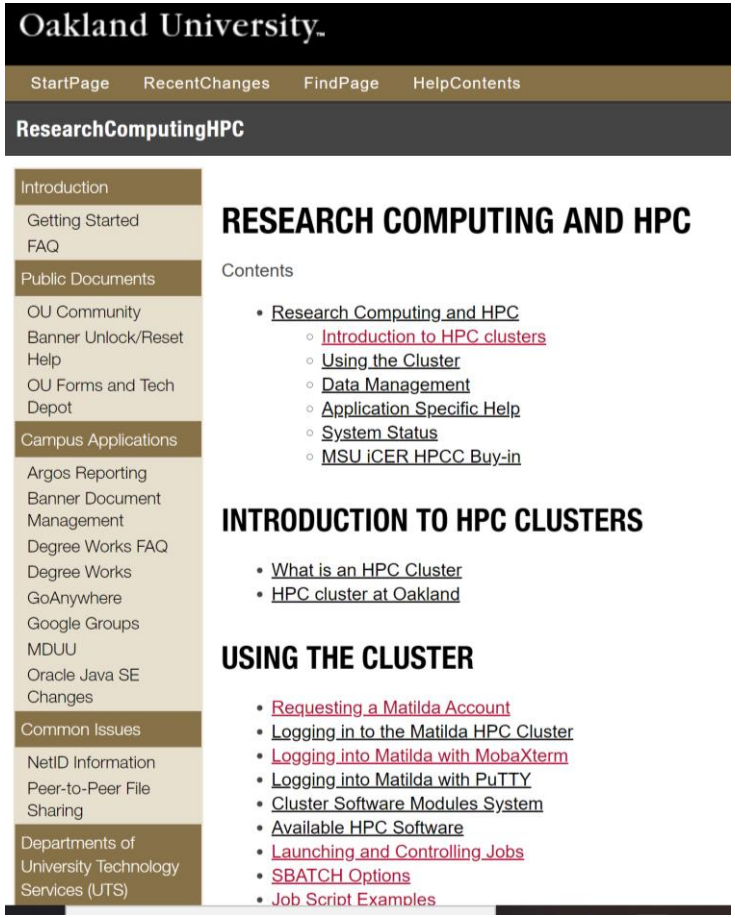
MSU ICER

icer.msu.edu



For general information:

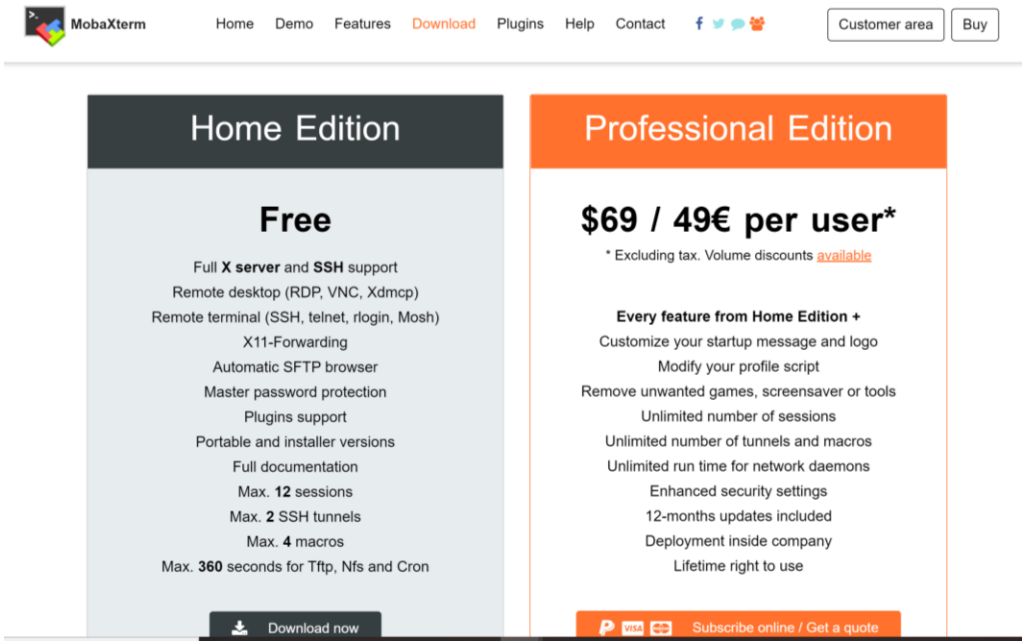
https://kb.oakland.edu/uts/ResearchComputingHPC#Introduction_to_HPC_clusters



To access the cluster from Windows:

[MobaXterm download (home edition)](https://mobaxterm.mobatek.net/download.html)

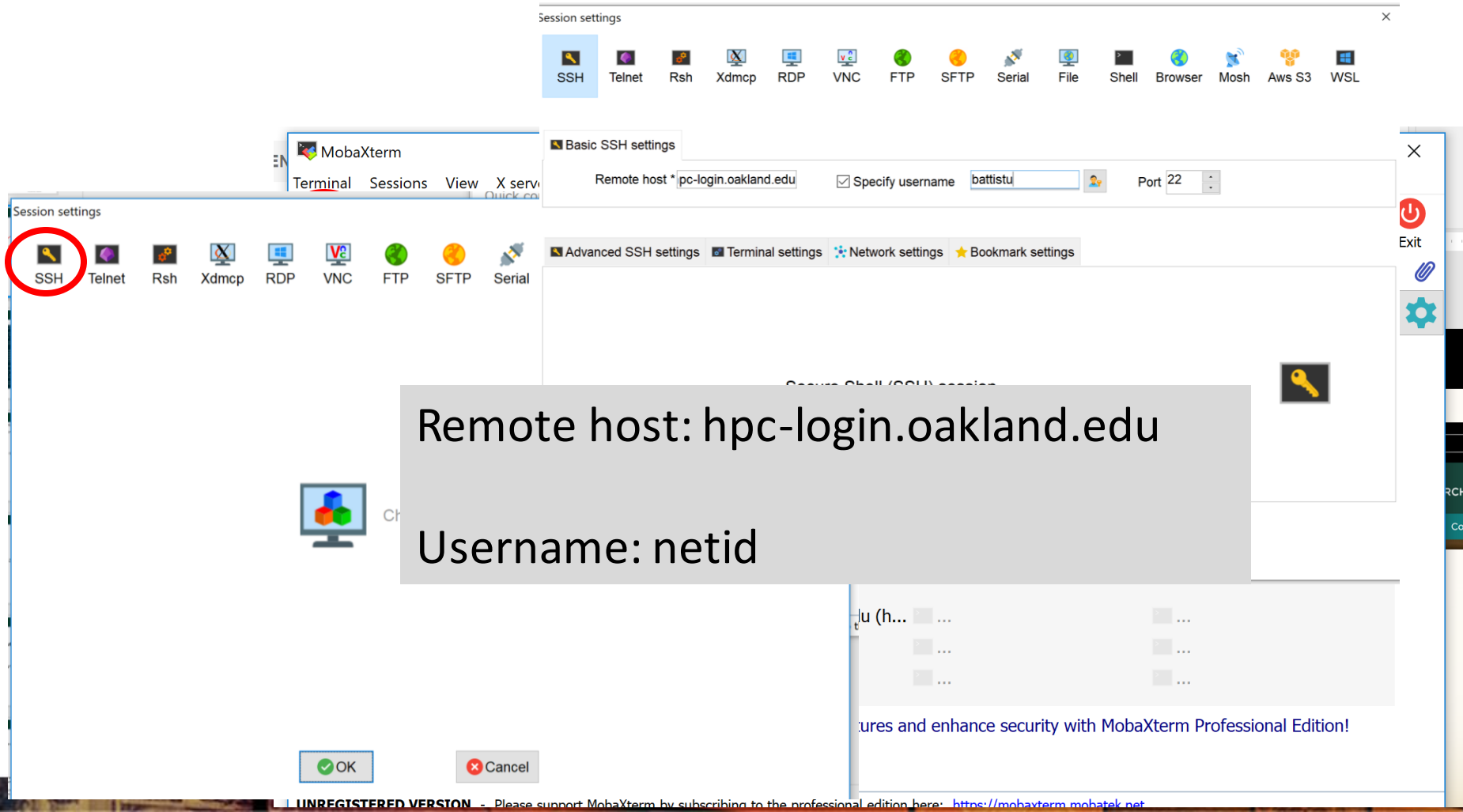
https://mobaxterm.mobatek.net/download.html



CENTER FOR **DATA SCIENCE** AND **BIG DATA** ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY



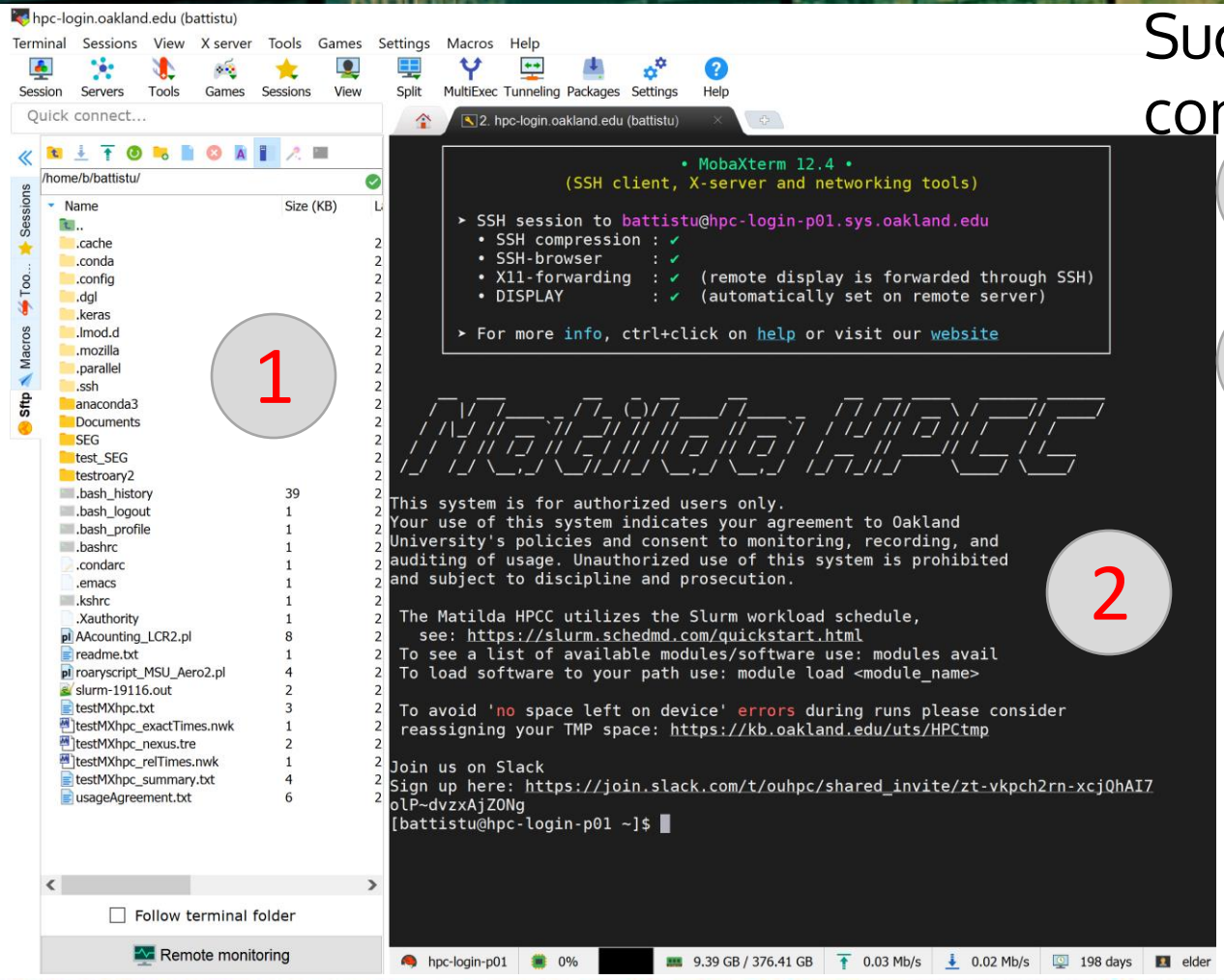
Remote host: hpc-login.oakland.edu

Username: netid

CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

Strength in Numbers

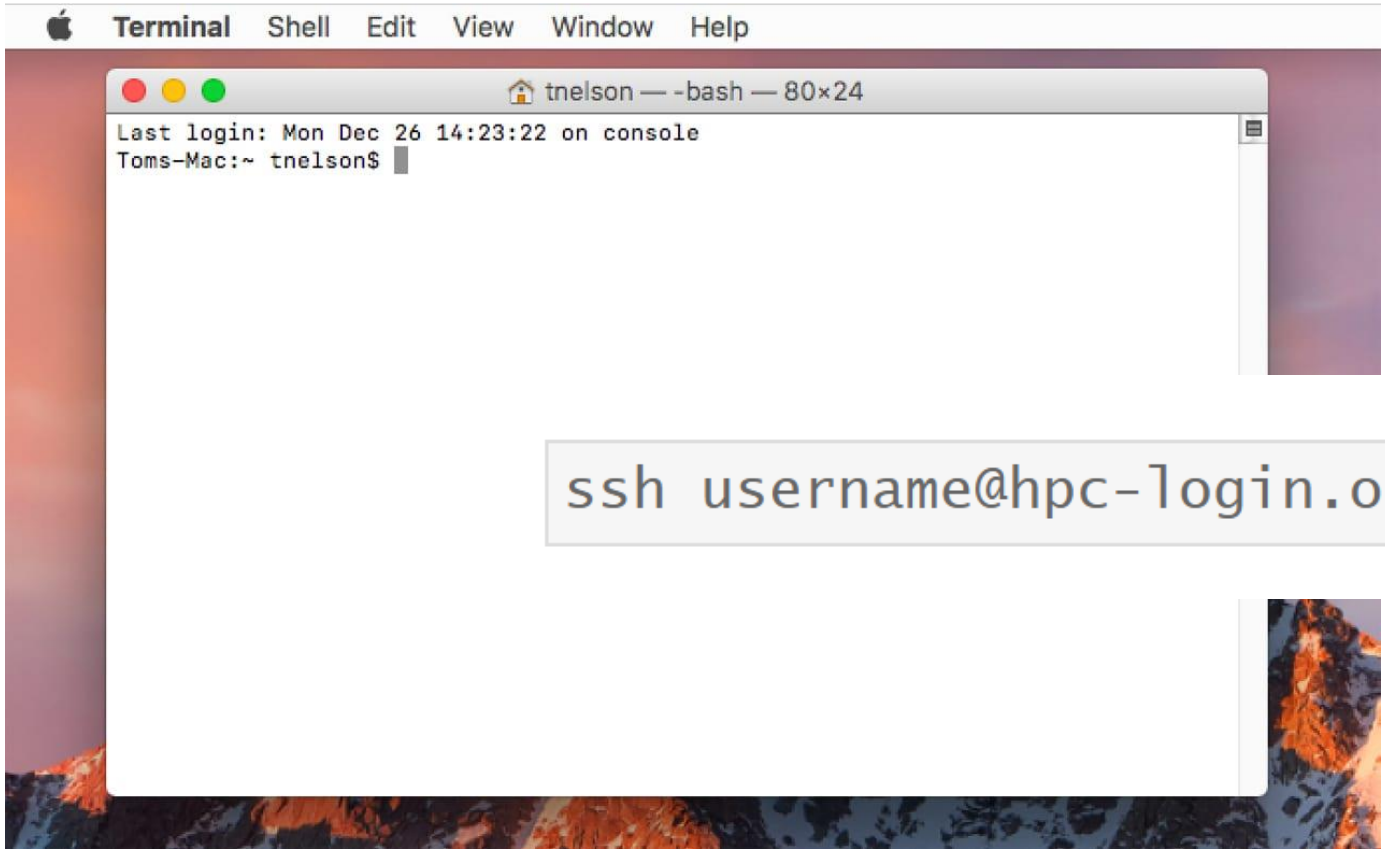
OAKLAND UNIVERSITY



Successful connection!

1. GUI of your home space on the cluster
2. Command line space

To access the cluster on Mac:




```
ssh username@hpc-login.oakland.edu
```

To visualize/transfer files to and from the cluster on Mac:

FileZilla The free FTP solution

Home
FileZilla
Features
Screenshots
Download
Documentation
FileZilla Pro
FileZilla Server
Download
Community
Forum
Wiki
General
FAQ
Support
Contact
License
Privacy Policy
Trademark Policy
Development
Source code
Nightly builds
Translations
Version history
Changelog
Issue tracker
Other projects

Download FileZilla Client

The latest stable version of FileZilla Client is 3

Please select the file appropriate for your platf

macOS

Download FileZilla Client

Requires macOS 10.13.2 or newer

More download options

Other platforms:

Not what you are looking for?

Show additional download options

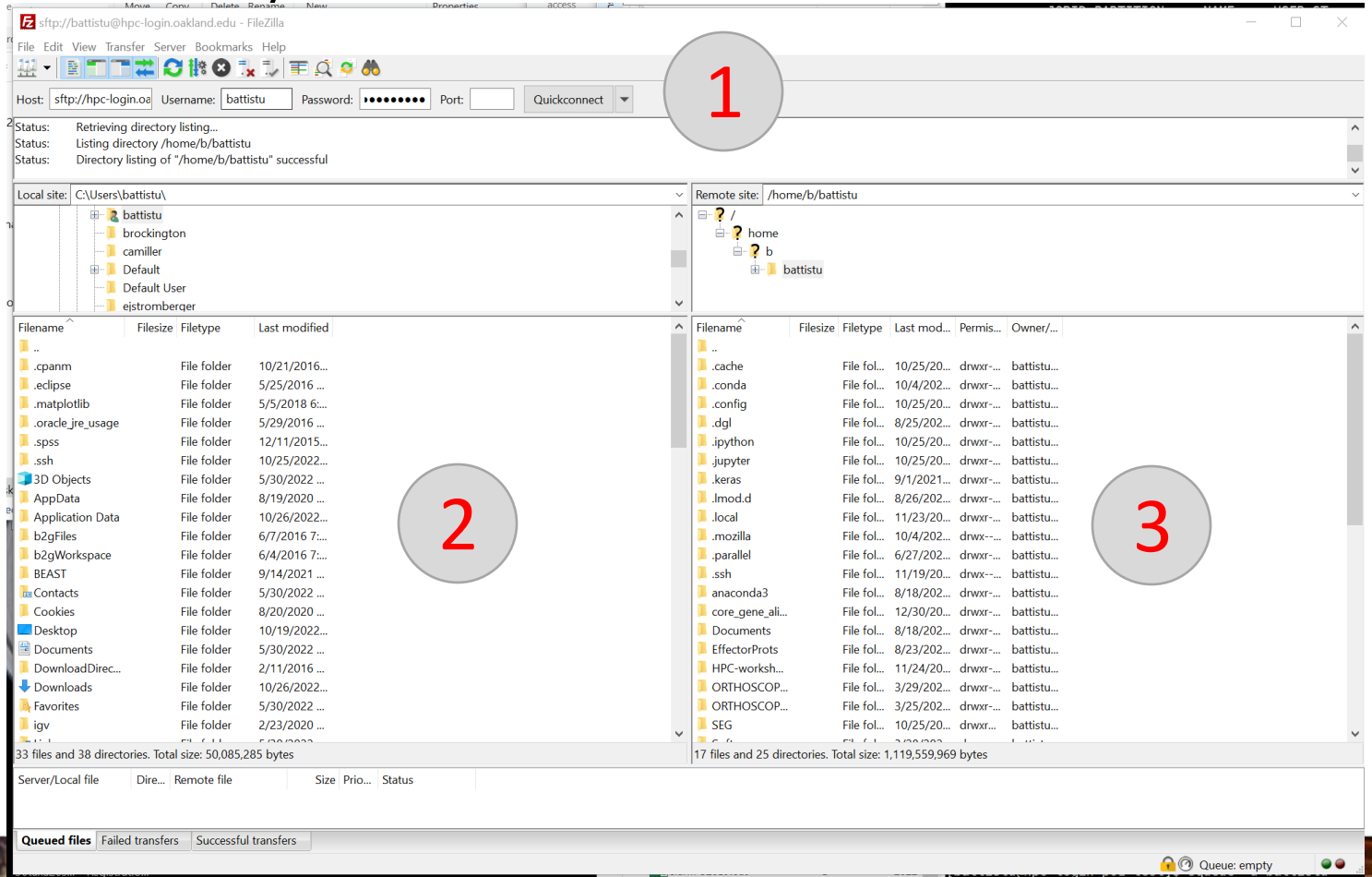
x86)

d page for

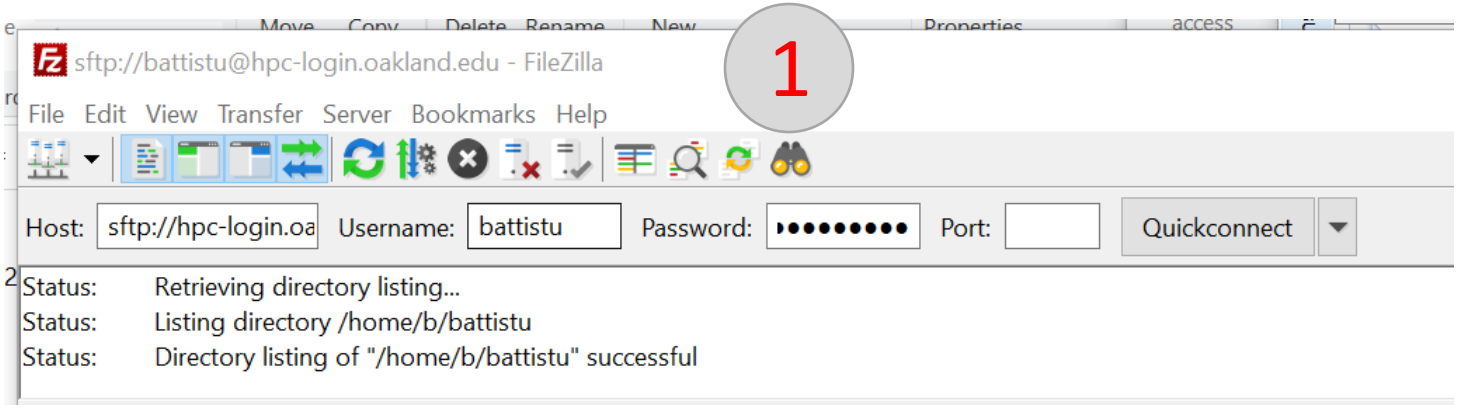
Please select your edition of FileZilla Client

	FileZilla	FileZilla with manual	FileZilla Pro	FileZilla Pro + CLI
Standard FTP	Yes	Yes	Yes	Yes
FTP over TLS	Yes	Yes	Yes	Yes
SFTP	Yes	Yes	Yes	Yes
Comprehensive PDF manual	-	Yes	Yes	Yes
Amazon S3	-	-	Yes	Yes
Backblaze B2	-	-	Yes	Yes
Dropbox	-	-	Yes	Yes
Microsoft OneDrive	-	-	Yes	Yes
Google Drive	-	-	Yes	Yes
Google Cloud Storage	-	-	Yes	Yes
Microsoft Azure Blob + File Storage	-	-	Yes	Yes
WebDAV	-	-	Yes	Yes
OpenStack Swift	-	-	Yes	Yes
Box	-	-	Yes	Yes
Site Manager synchronization	-	-	Yes	Yes
Command-line interface	-	-	-	Yes
Batch transfers	-	-	-	Yes
	Download	Select	Select	Select

To visualize/transfer files to and from the cluster on Mac:

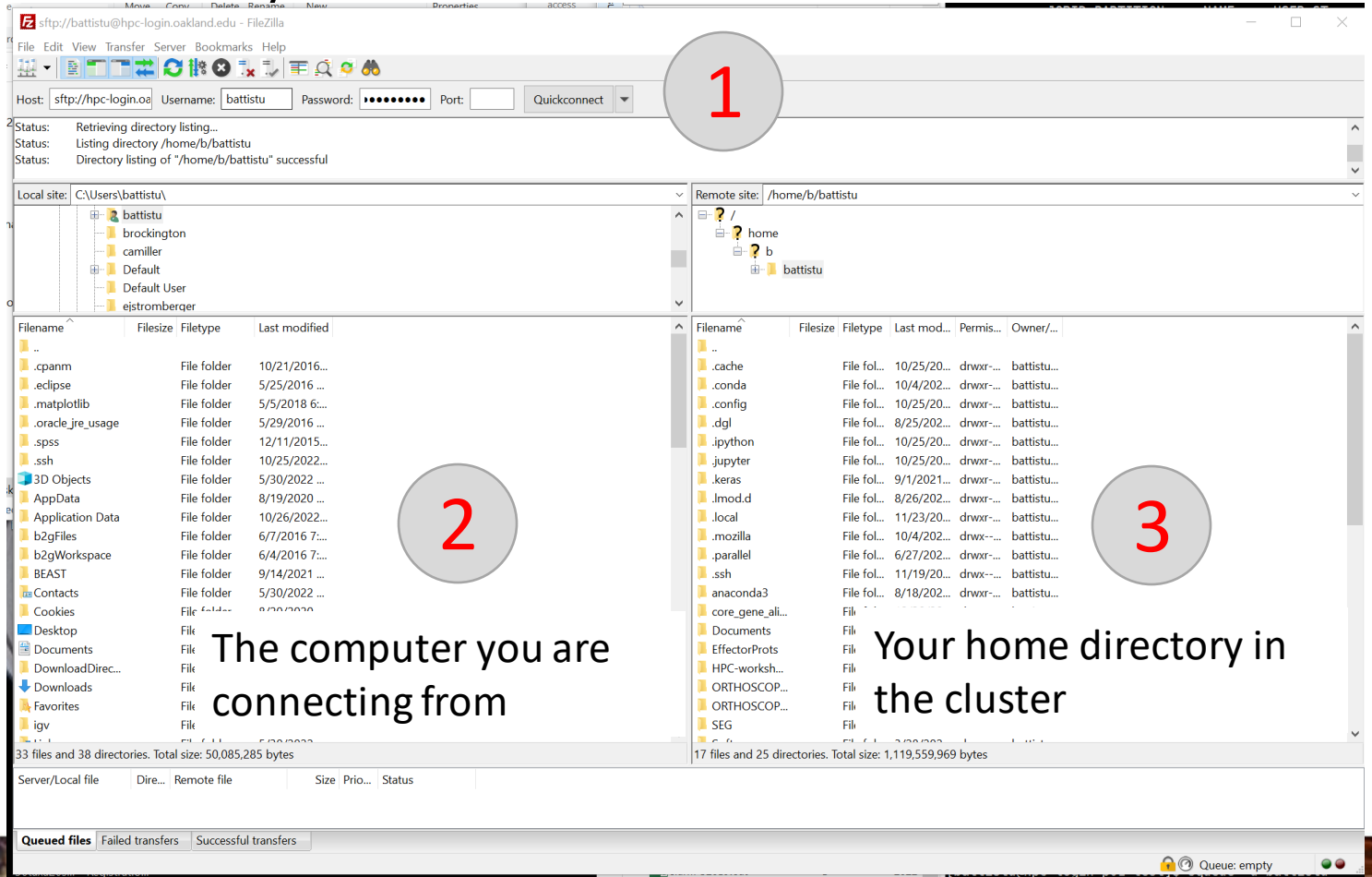


To visualize/transfer files to and from the cluster on Mac:



Host: hpc-login.oakland.edu
Username: netID
Password: netID password
Port: 22

To visualize/transfer files to and from the cluster on Mac:



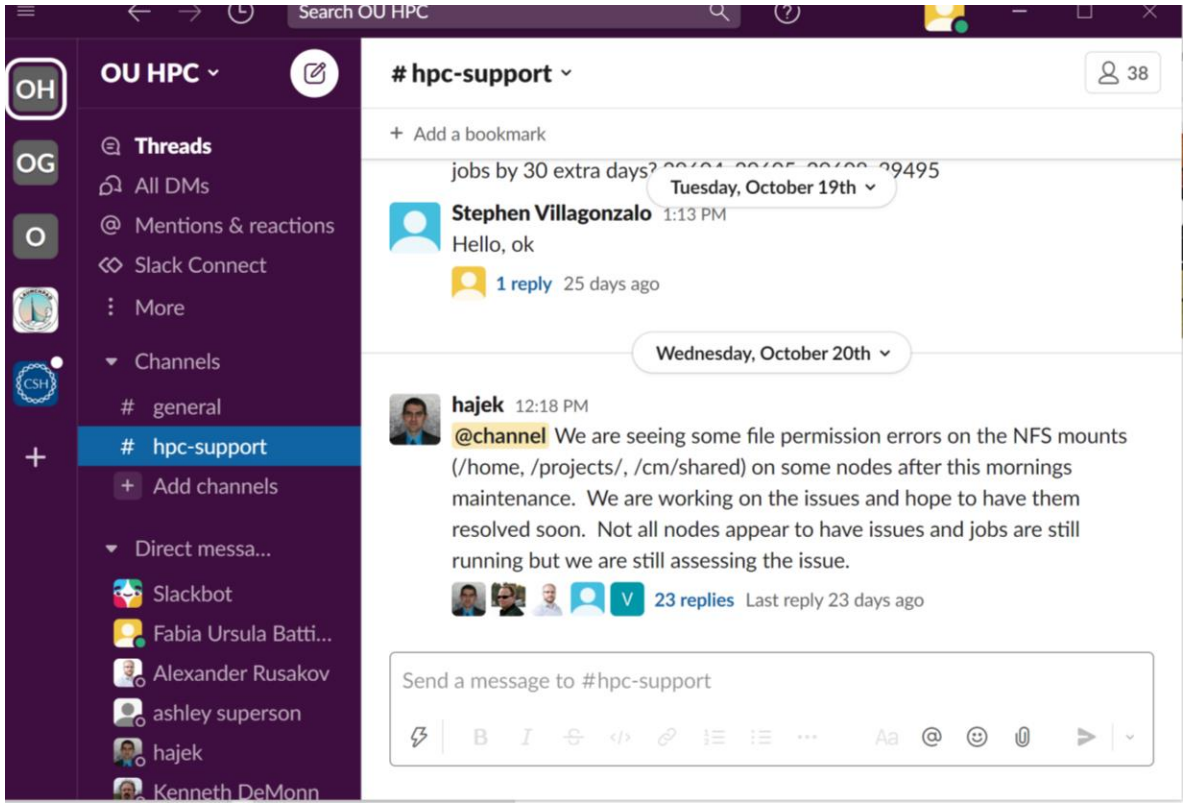
CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

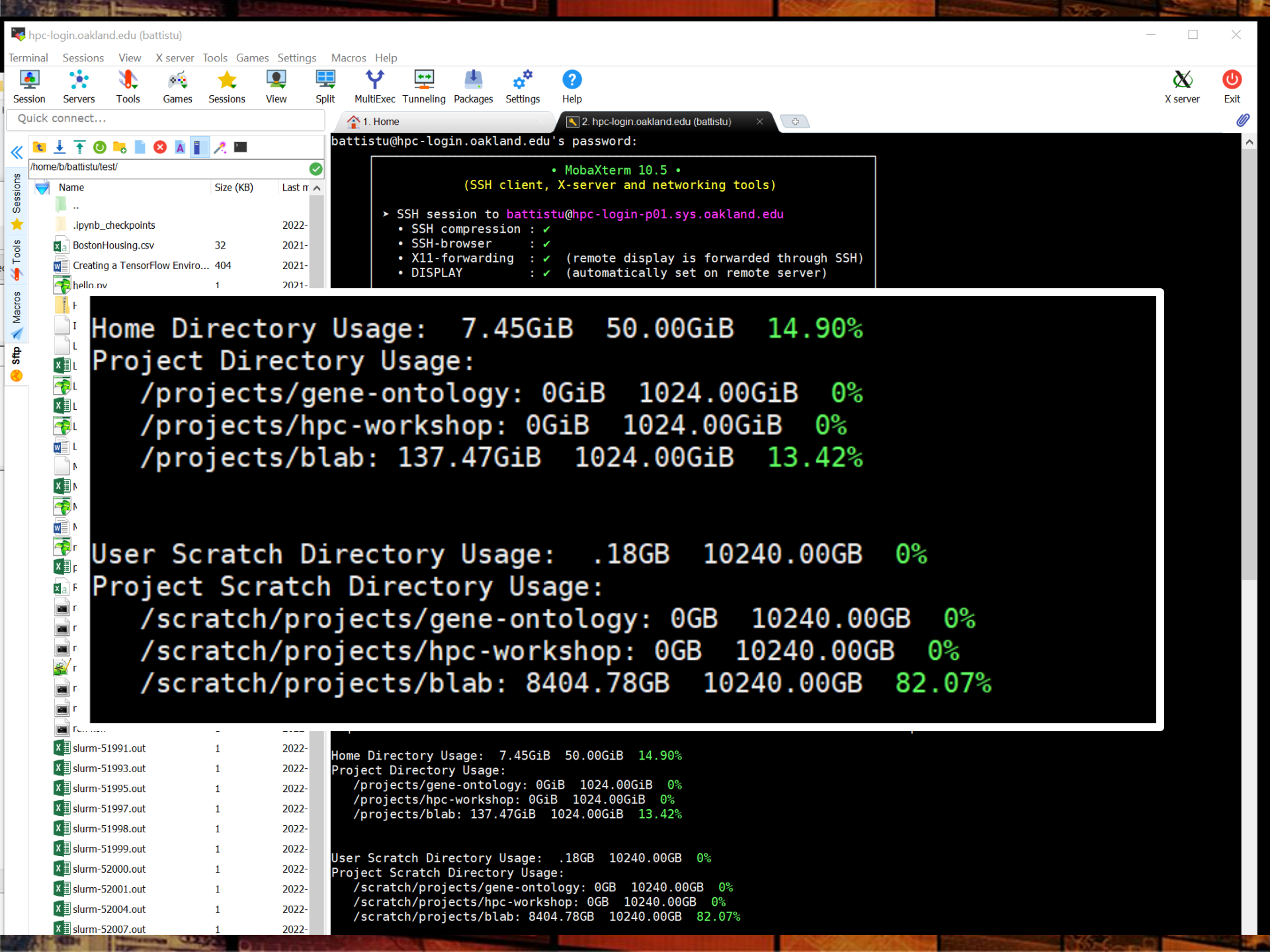
Strength in Numbers

OAKLAND UNIVERSITY

Technical support:

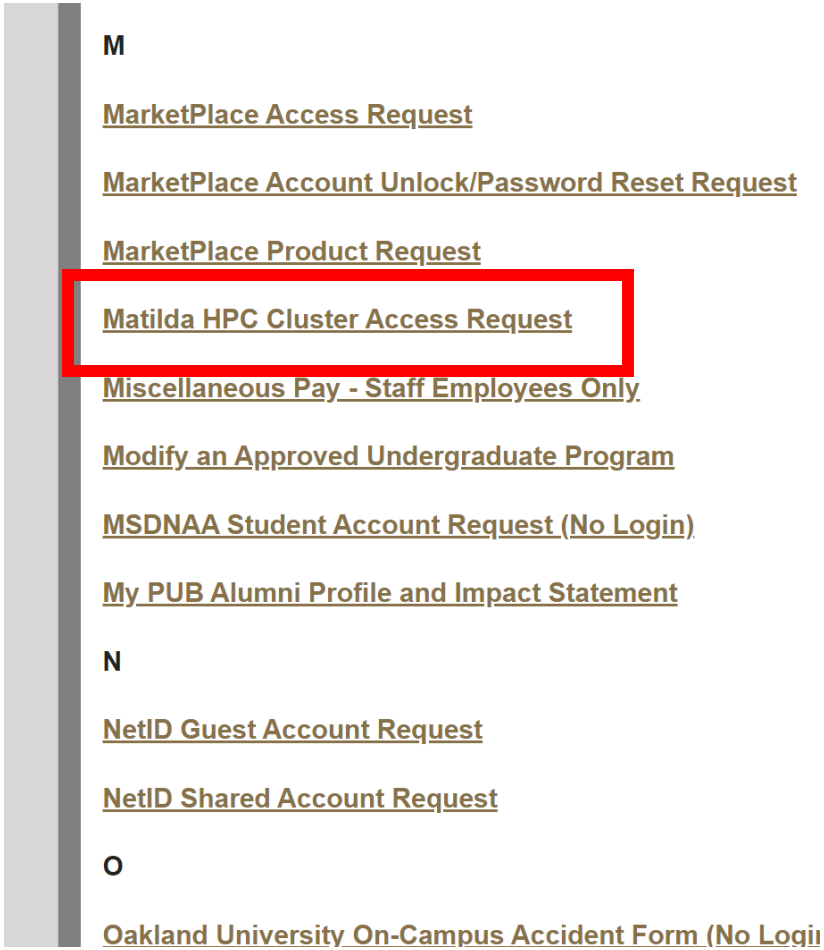
slack




hpc-login.oakland.edu (battistu)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
1. Home
2. hpc-login.oakland.edu (battistu)
/home/b/battistu/test/
Name Size (KB) Last m
.ipynb_checkpoints 2022-
BostonHousing.csv 32 2021-
Creating a TensorFlow Enviro... 404 2021-
hello.py 1 2021-
slurm-51991.out 1 2022-
slurm-51993.out 1 2022-
slurm-51995.out 1 2022-
slurm-51997.out 1 2022-
slurm-51998.out 1 2022-
slurm-51999.out 1 2022-
slurm-52000.out 1 2022-
slurm-52001.out 1 2022-
slurm-52004.out 1 2022-
slurm-52007.out 1 2022-
battistu@hpc-login.oakland.edu's password:
• MobaXterm 10.5 •
(SSH client, X-server and networking tools)
➤ SSH session to battistu@hpc-login-p01.sys.oakland.edu
• SSH compression : ✔
• SSH-browser : ✔
• X11-forwarding : ✔ (remote display is forwarded through SSH)
• DISPLAY : ✔ (automatically set on remote server)
Home Directory Usage: 7.45GiB 50.00GiB 14.90%
Project Directory Usage:
/projects/gene-ontology: 0GiB 1024.00GiB 0%
/projects/hpc-workshop: 0GiB 1024.00GiB 0%
/projects/blab: 137.47GiB 1024.00GiB 13.42%
User Scratch Directory Usage: .18GB 10240.00GB 0%
Project Scratch Directory Usage:
/scratch/projects/gene-ontology: 0GB 10240.00GB 0%
/scratch/projects/hpc-workshop: 0GB 10240.00GB 0%
/scratch/projects/blab: 8404.78GB 10240.00GB 82.07%

To get a Matilda account:
forms.oakland.edu

If you had an account created just for this workshop, it will be deactivated Dec. 6. To request a long-term account follow the instructions



CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY

To access Matilda from outside of campus: *forms.oakland.edu*

Request a VPN account, set up DUO authentication with your phone (or another device)

Software & Hosted Solution Purchasing Checklist (No Login)

Special Account Request (pdf)

Special Credit Offering Request

Student Employee/Intern Confidentiality Agreement

T

Technology Control Plan Template

Touchnet Unlock Account / Password Reset

Travel Authorization Request (UTS)

Tuition Assistance Application

Tuition Assistance Application for Faculty

V

Vendor/Consultant Account Access Request

Virtual Labs Change Request

VPN Access Request

Y

Youth protection Approval Request

Z

Zoom HIPAA Request

To execute an analysis:

Use SLURM to schedule your job in the cluster. This will make the job go in a queue and allow to request resources (# nodes, # of CPUs, types of nodes, etc)

For testing purpose, you can skip the cue and run the job on the login node (**only** for testing purposes!)

Let’s run “Hello world”

"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster

To write the script on windows:

1. Open Notepad++ on Windows (do not use Word) or BioEdit on Mac on your computer (not the cluster)
2. Type: print("Hello world")
3. Save as "helloworld.py"
4. Upload (or drag and drop) the file into your home directory in the cluster

To write the script in Linux:

1. In your home directory type: nano (this is a text editor in linux)
2. Type: print("Hello world")
3. Ctrl+x, give a file name and follow instructions to save it

"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster (python is automatically loaded when you log in but for practice let's go through the process)

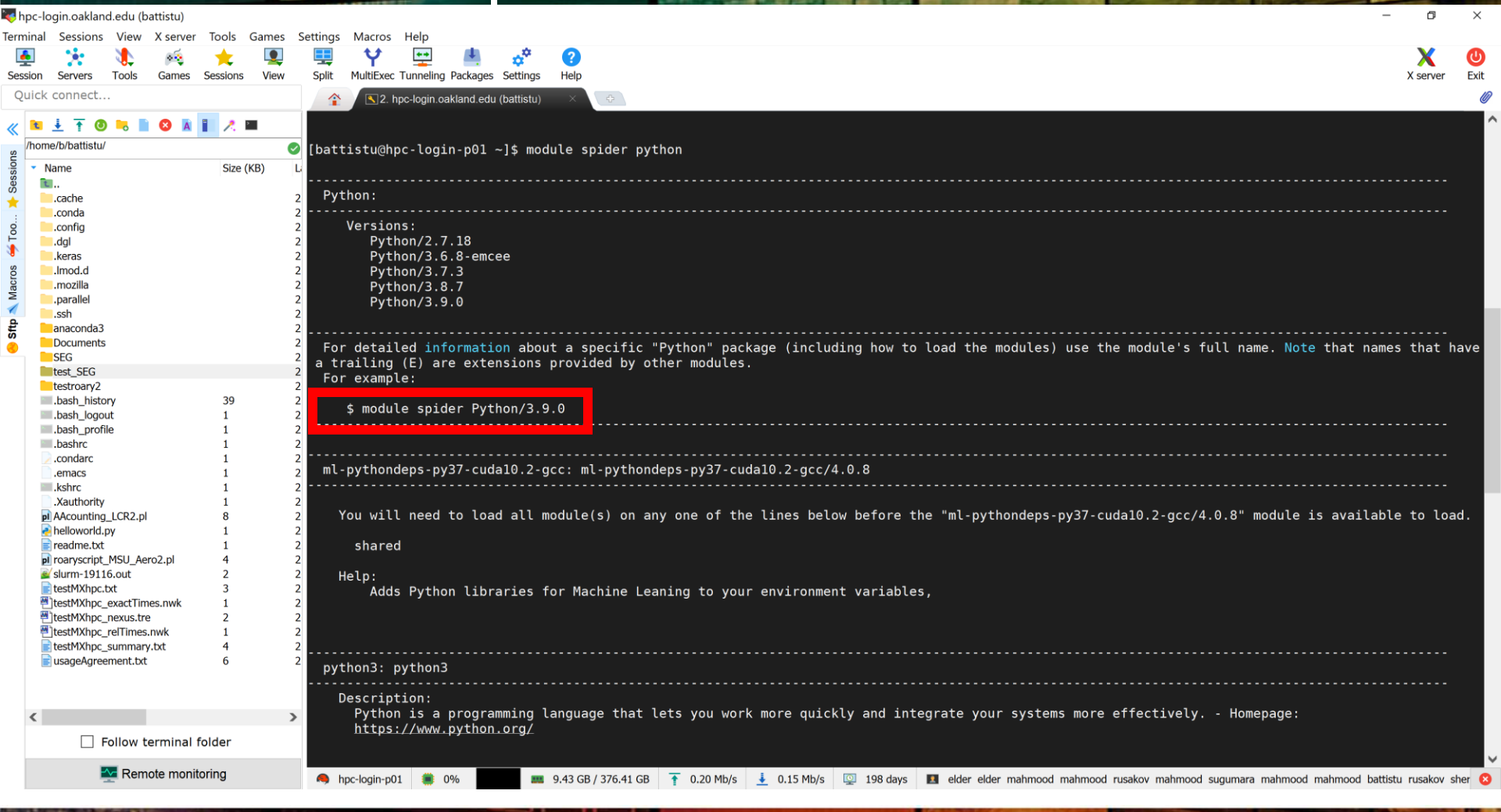
In the cluster type:

```
module spider python
```

CENTER FOR **DATA SCIENCE** AND **BIG DATA** ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY



"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster

```
[battistu@hpc-login-p01 ~]$ module spider Python/3.9.0

----------------------------------------------------------------------------------------------------------------------------------------
  Python: Python/3.9.0
----------------------------------------------------------------------------------------------------------------------------------------

    You will need to load all module(s) on any one of the lines below before the "Python/3.9.0" module is available to load.

      shared

    Help:
        Adds  Python3.9.0 to path,




[battistu@hpc-login-p01 ~]$
```

"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster

```
battistu@hpc-login-p01 ~]$ module list

urrently Loaded Modules:
 1) shared   2) DefaultModules   3) dot   4) slurm/slurm/19.05.8   5) default-environment   6) gcc/9.2.0


battistu@hpc-login-p01 ~]$
```

"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster

```
battistu@hpc-login-p01 ~]$
battistu@hpc-login-p01 ~]$ module load Python/3.9.0
battistu@hpc-login-p01 ~]$ module list

currently Loaded Modules:
  1) shared   2) DefaultModules   3) dot   4) slurm/slurm/19.05.8   5) default-environment   6) gcc/9.2.0   7) Python/3.9.0



battistu@hpc-login-p01 ~]$
```

"Hello world" is a script written in Python.

To run you will need:

1. The script
2. Python loaded on the cluster

Case sensitive
Do not use spaces
Meaningful names

To run your script type:

`python3 helloworld.py`

```
[battistu@hpc-login-p01 ~]$ python3 helloworld.py
Hello world
[battistu@hpc-login-p01 ~]$
```

"Hello world" is a script written in Python.

To submit your job you will need:

1. The script
2. The SLURM file
3. Python loaded on the cluster

```
#!/bin/bash --login

########### SBATCH Lines for Resource Request ############
SBATCH --time=00:01:00        # walltime
#SBATCH --nodes=1             # number of nodes
#SBATCH --ntasks=1            # number of tasks
#SBATCH --cpus-per-task=40    # number of CPUs (or cores) per task
#SBATCH --mem-per-cpu=2G      # memory required per allocated CPU
#SBATCH --job-name HelloWorld  # job name for easier identification
```

########## Command Lines to Run ##########

```
module load Python/3.9.0
cd /home/b/battistu          ### directory where your code is located

python3 helloworld.py > output.txt      ### call your executable

scontrol show job $SLURM_JOB_ID     ### write job info to output file
```

```
#!/bin/bash --login

########## SBATCH Lines for Resource Request ##########

#SBATCH --time=00:01:00             # limit of wall clock time - how long the job will run in hrs:min:sec (same as -t). Max 168:00:00 (7 days)
#SBATCH --nodes=1                  # number of different nodes - could be an exact number or a range of nodes (same as -N)
#SBATCH --ntasks=1                  # number of tasks - how many tasks (nodes) that you require (same as -n)
#SBATCH --cpus-per-task=40           # number of CPUs (or cores) per task (same as -c). Max for most nodes: 40
#SBATCH --mem-per-cpu=2G            # memory required per allocated CPU (or core) - amount of memory (in bytes)
#SBATCH --job-name HelloWorld  # you can give your job a name for easier identification (same as -J)

########## Command Lines to Run ##########
module load Python/3.9.0

cd /home/b/battistu             ### change to the directory where your code is located

python3 helloworld.py > outut.txt          ### call your executable

scontrol show job $SLURM_JOB_ID      ### write job information to output file

```

Upload the SLURM file into your home directory

Run the job:

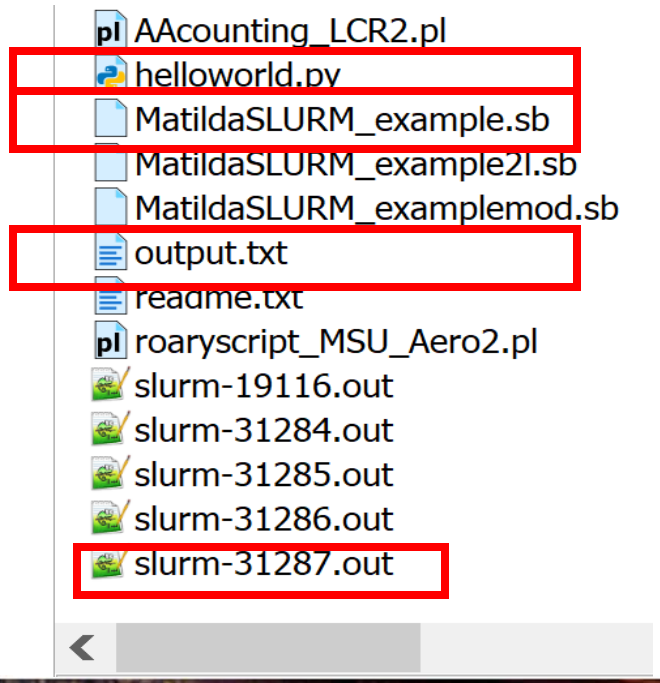
```
sbatch MatildaSLURM_example.sb
```

```
battistu@hpc-login-p01 ~]$
battistu@hpc-login-p01 ~]$ sbatch MatildaSLURM_example.sb
batch: error: Batch script contains DOS line breaks (\r\n)
batch: error: instead of expected UNIX line breaks (\n).
battistu@hpc-login-p01 ~]$
```

```
module spider dos2unix
module load shared
module load dos2unix/7.4.2
dos2unix MatildaSLURM_example.sb
```

Repeat sbatch command

```
battistu@hpc-login-p01 ~]$
battistu@hpc-login-p01 ~]$
battistu@hpc-login-p01 ~]$
battistu@hpc-login-p01 ~]$ sbatch MatildaSLURM_example.sb
ubmitted batch job 31287
battistu@hpc-login-p01 ~]$ squeue -u battistu
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
battistu@hpc-login-p01 ~]$
```



`squeue -u netid` to check job status

`scancel job#` to cancel a job

Output is created as expected

Slurm-job#.out: check how the job ran; this is where you will find error messages

You just completed your first HPC analysis