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

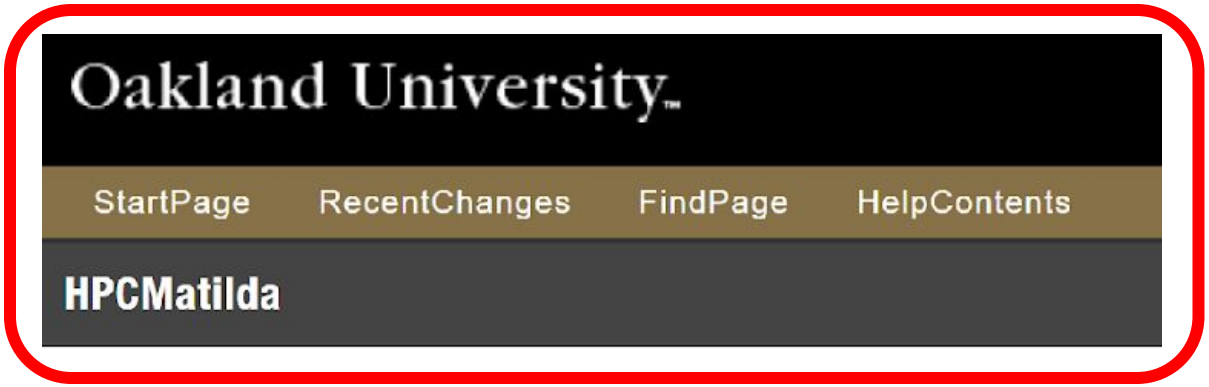
Stephen Villagonzalo
svillagonzalo@oakland.edu

John Johnston
jbjohnston@oakland.edu

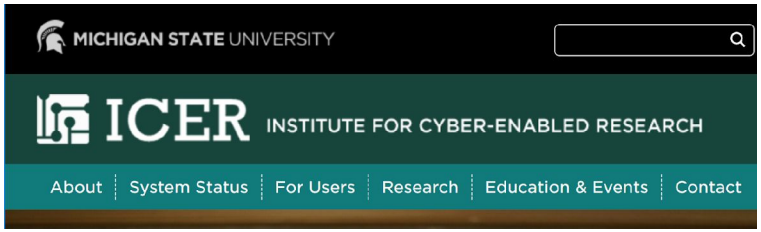
CENTER FOR
DATA SCIENCE
AND
BIG DATA
ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY

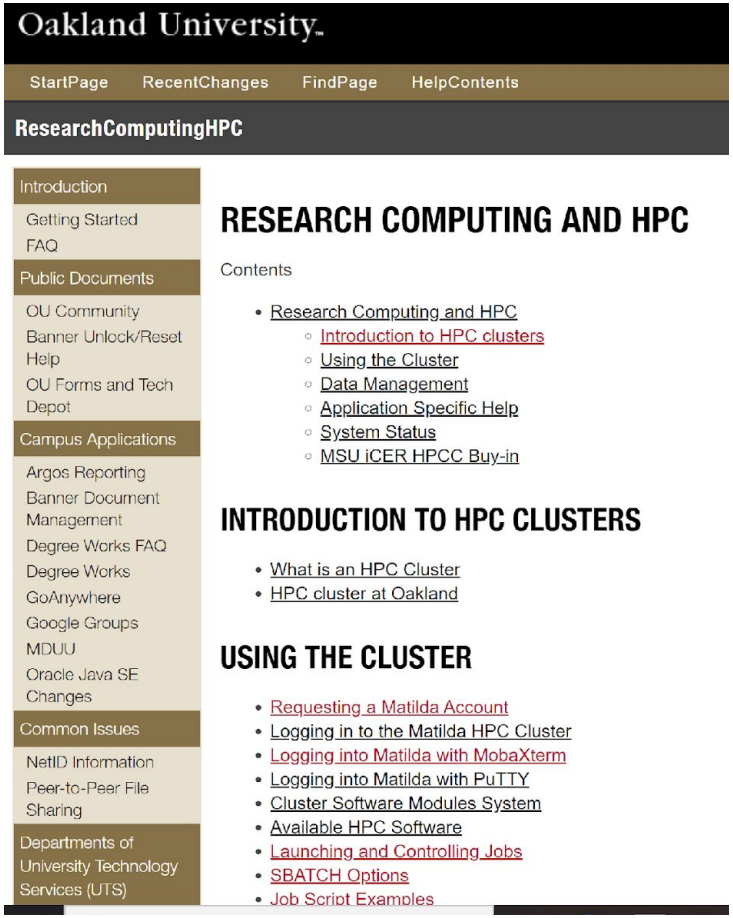


MSU ICER
icer.msu.edu



For general information:

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



Oakland University

StartPage RecentChanges FindPage HelpContents

ResearchComputingHPC

Introduction
Getting Started
FAQ
Public Documents
OU Community
Banner Unlock/Reset Help
OU Forms and Tech Depot
Campus Applications
Argos Reporting
Banner Document Management
Degree Works FAQ
Degree Works
GoAnywhere
Google Groups
MDUU
Oracle Java SE Changes
Common Issues
NetID Information
Peer-to-Peer File Sharing
Departments of University Technology Services (UTS)

RESEARCH COMPUTING AND HPC

Contents

- Research Computing and HPC
  - Introduction to HPC clusters
  - Using the Cluster
  - Data Management
  - Application Specific Help
  - System Status
  - MSU iCER HPCC Buy-in

INTRODUCTION TO HPC CLUSTERS

- What is an HPC Cluster
- HPC cluster at Oakland

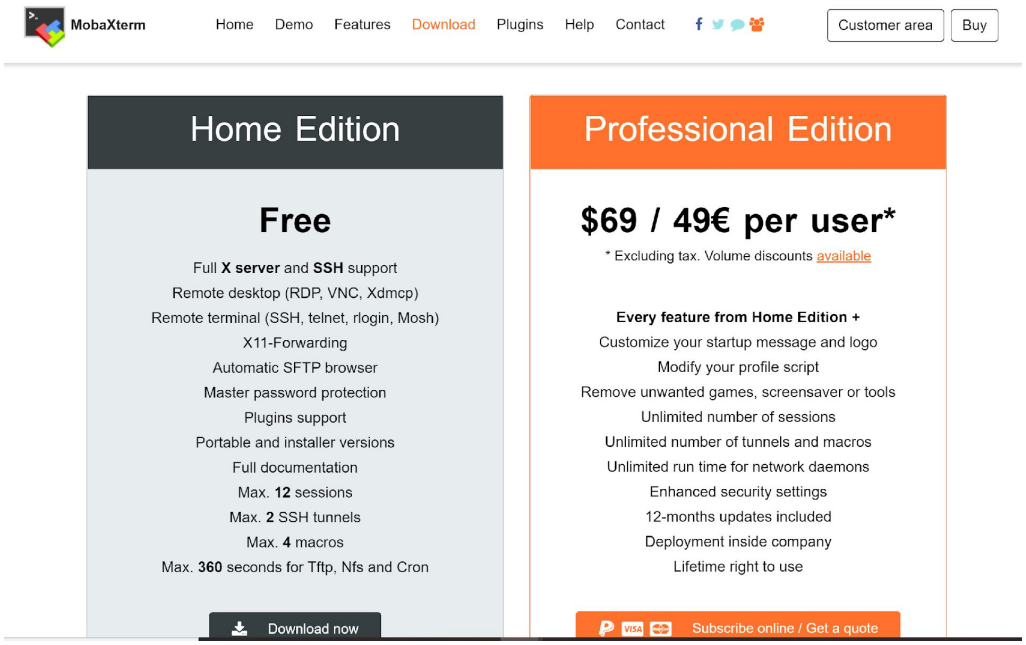
USING THE CLUSTER

- Requesting a Matilda Account
- Logging in to the Matilda HPC Cluster
- Logging into Matilda with MobaXterm
- Logging into Matilda with PuTTY
- Cluster Software Modules System
- Available HPC Software
- Launching and Controlling Jobs
- SBATCH Options
- Job Script Examples

To access the cluster from Windows:

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

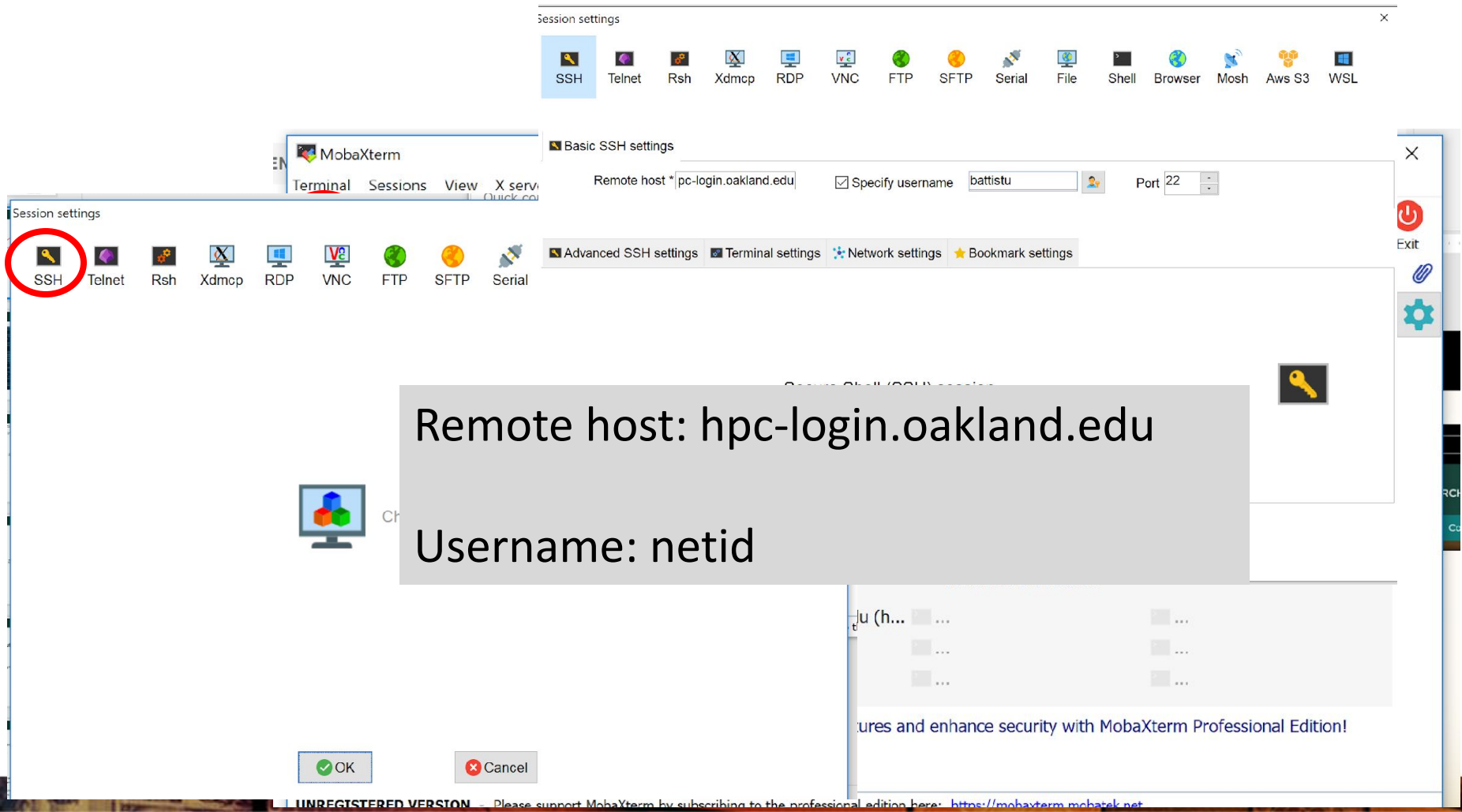
https://mobaxterm.mobatek.net/download.html



CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

Strength in Numbers

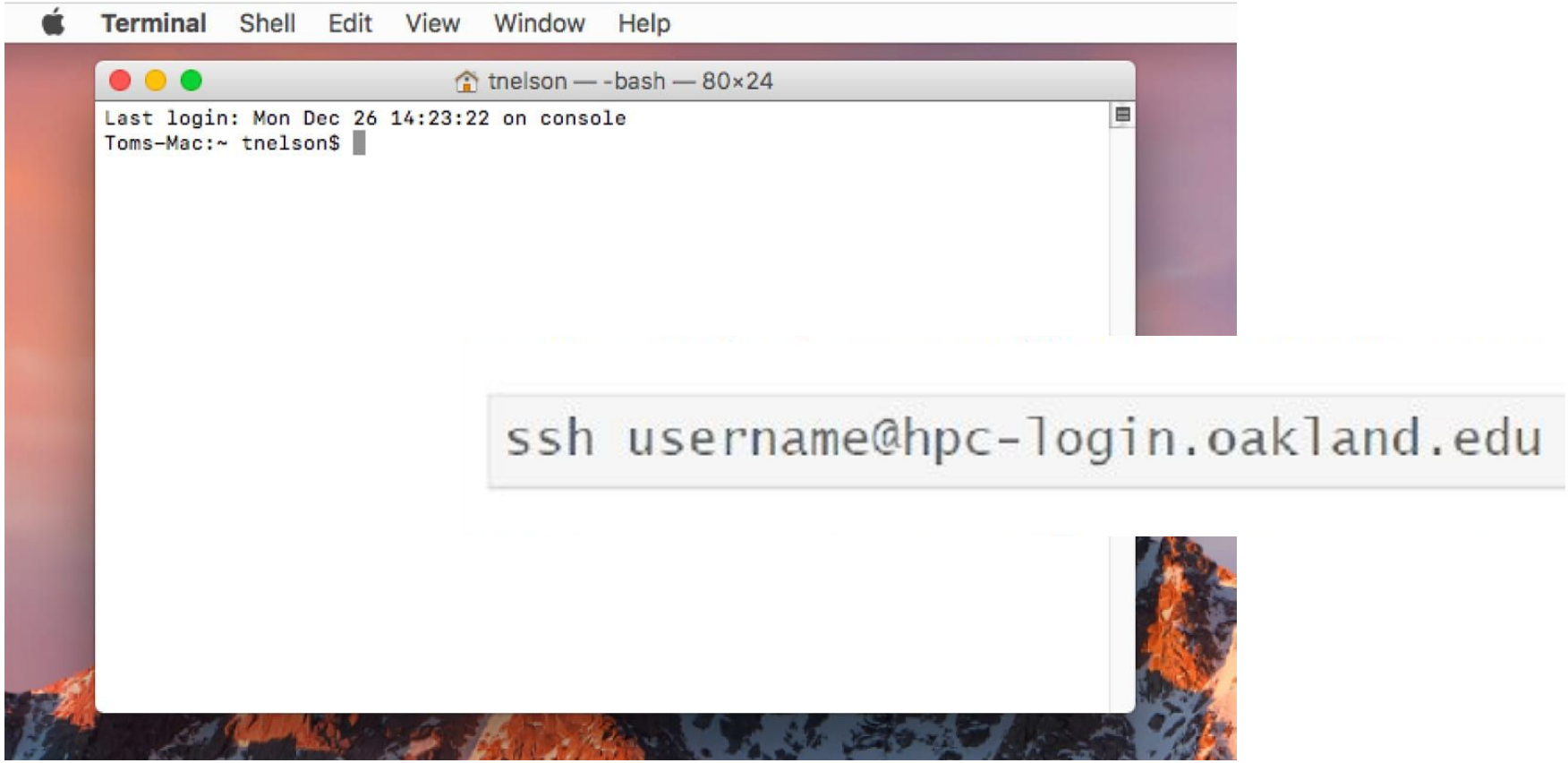
OAKLAND UNIVERSITY



Remote host: hpc-login.oakland.edu

Username: netid

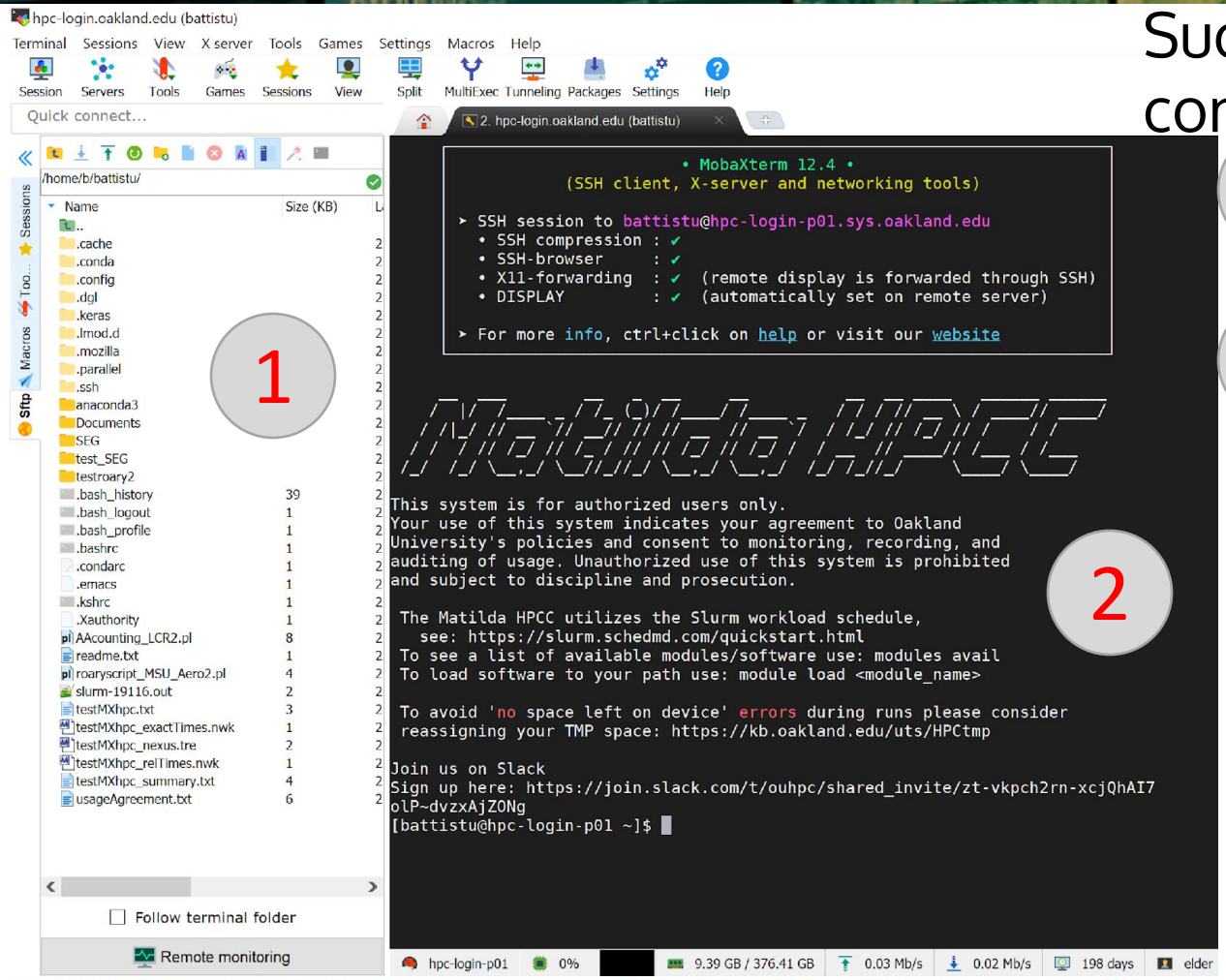
To access the cluster on Mac:



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

Strength in Numbers

OAKLAND UNIVERSITY



Successful connection!

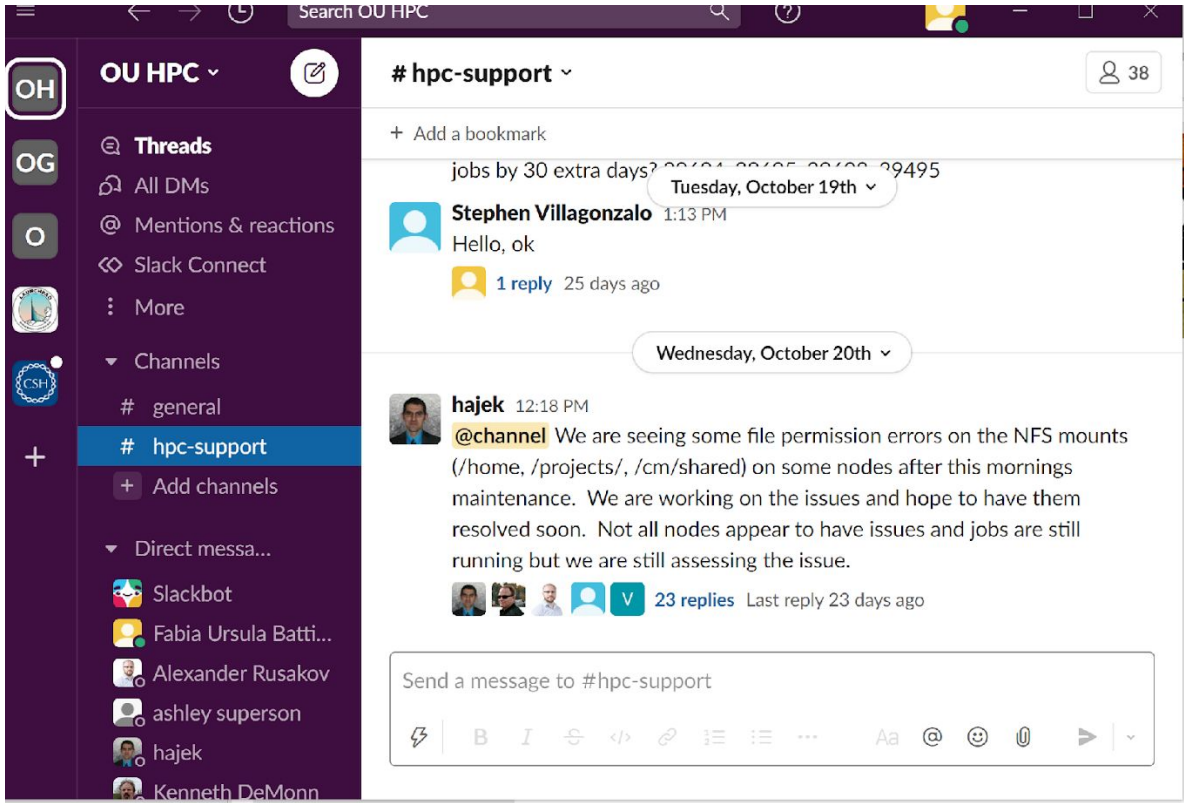
1. GUI of your home space
2. Command line space

CENTER FOR DATA SCIENCE AND BIG DATA ANALYTICS

Strength in Numbers

OAKLAND UNIVERSITY

Technical support: slack



What is available on Matilda:

Home directory: home/firstletternetID/netID (e.g., home/b/battistu) **20GB** (this is where you land when you connect)
Project directory: projects/blab **1TB** (to access this you can do cd /project/blab): this is a shared space for everyone in a project. This has to be activated by the PI of the group

Files in these two locations will not be deleted unless you actively delete them

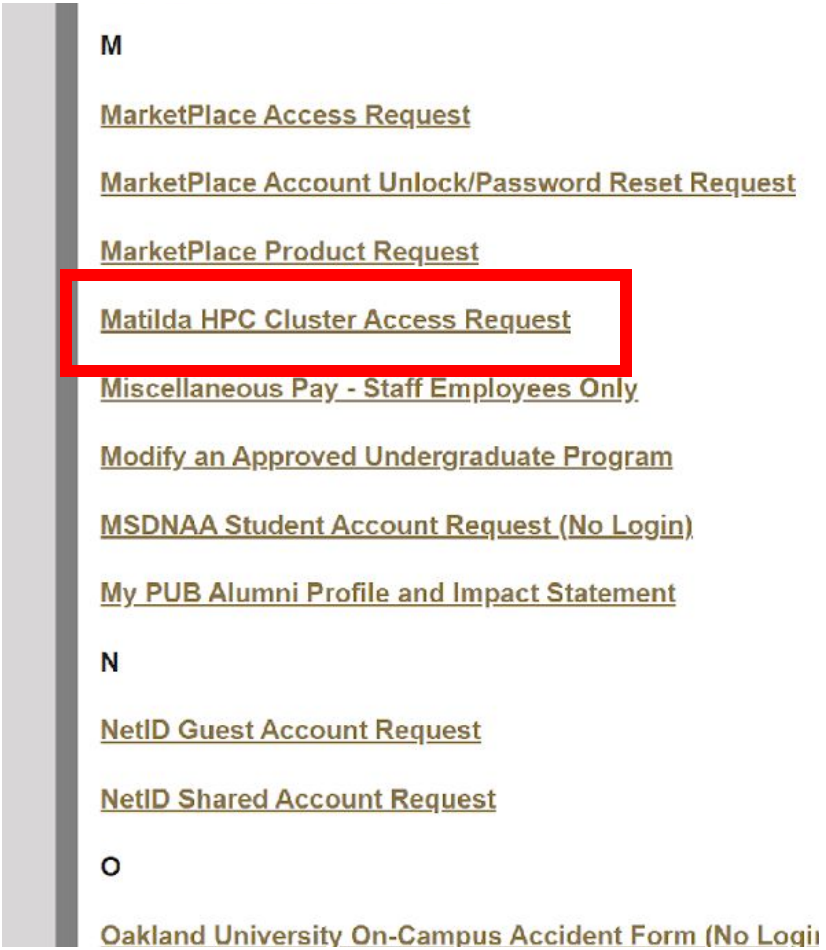
What is available on Matilda:

Then there is the scratch space:
Scratch project: scratch/projects/blab **10TB** (to access this you can do cd /scratch/projects/blab): this is a shared space for all of us
Scratch home: scratch/users/netID (e.g., scratch/users/battistu) **10 TB**

*Inactive files in scratch are **automatically deleted every 45 days** so make sure to regularly download the files*

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++ (do not use Word) on your computer (not the cluster)
2. Type: print("Hello world")
3. Save
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

"Hello world" is a script written in Python.

To run you will need:

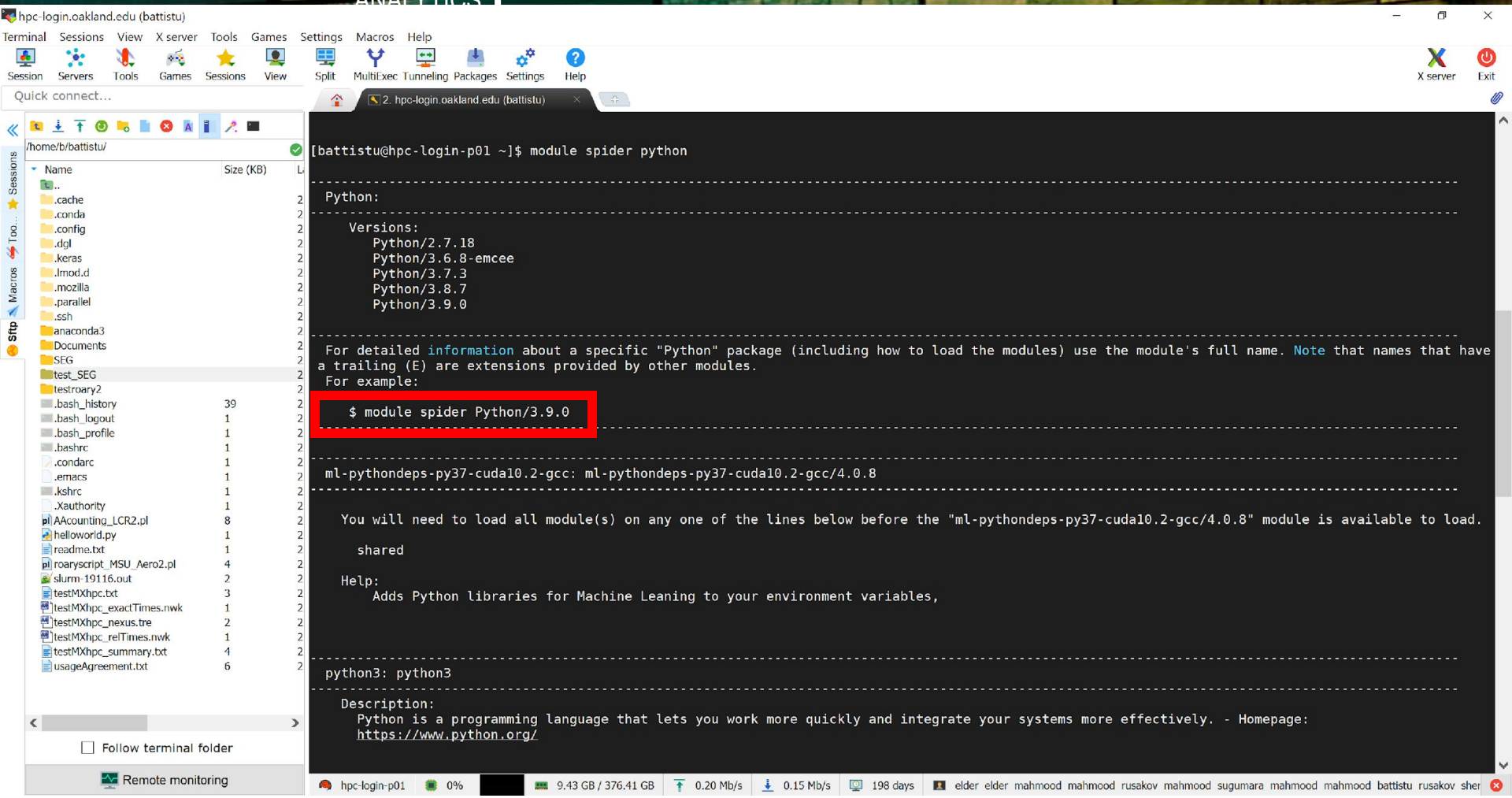
1. The script
2. Python loaded on the cluster (python is automatically loaded when you log in but for practice let's go through the process)

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:

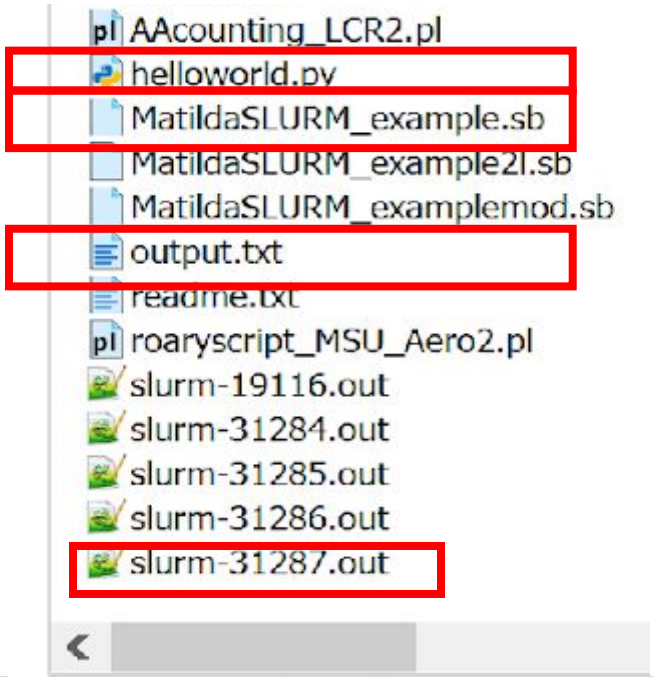
sbatch MatildaSLURM_example.sb

```
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 31827 to cancel a job

Output is created as expected

Slurm-31287.out: check how the job ran; this is where you will find error messages

You just completed your first HPC analysis