Predictive Analytics (ISE529)

CARC High Performance Computing (HPC)

Dr. Tao Ma
ma.tao@usc.edu

Tue/Thu, May 22 - July 1, 2025, Summer

USC Viterbi
School of Engineering
Daniel J. Epstein Department of Industrial and Systems Engineering

USC University of Southern California

Important Website

CARC website : https://www.carc.usc.edu

User Guides: https://www.carc.usc.edu/user-guides

User portal : https://hpcaccount.usc.edu

Running Jobs on CARC Systems
https://www.carc.usc.edu/user-guides/hpc-systems/using-our-hpc-systems/running-jobs

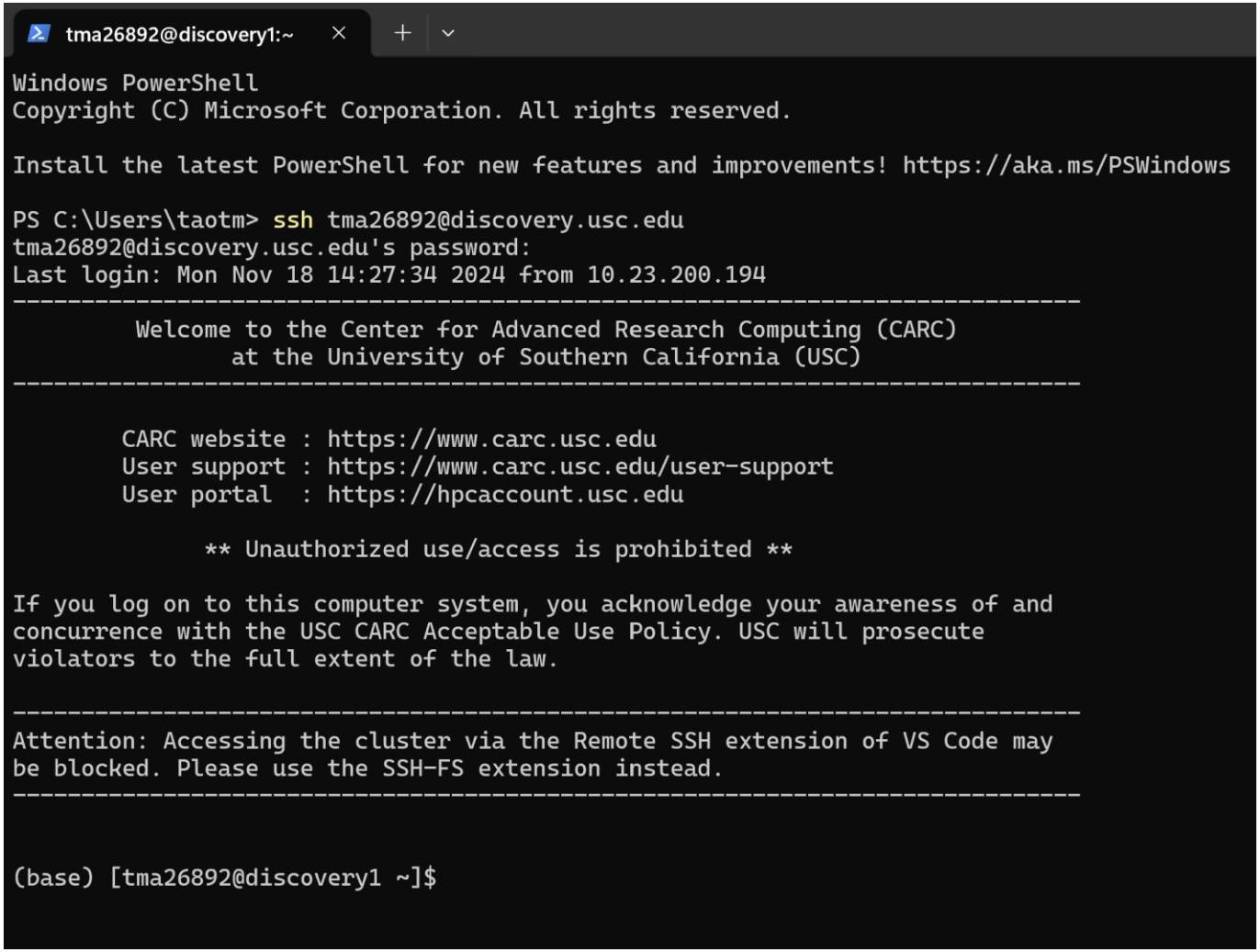
Slurm Cheatsheet: https://www.carc.usc.edu/user-guides/hpc-systems/using-our-hpc-systems/slurm-cheatsheet

Slurm Job Script Templates: https://www.carc.usc.edu/user-guides/hpc-systems/using-our-hpc-systems/slurm-templates

Connecting to a USC VPN:
https://www.carc.usc.edu/user-guides/quick-start-guides/anyconnect-vpn-setup

Login to HPC-cluster

- Start a terminal in your laptop & login to the HPC system:
  > ssh <username>@discovery.usc.edu



Trouble-shooting for Login problem

- On USC campus, wifi connection must be: USC Security Wireless
- Use VPN if connect from home or off-campus.
- If you encounter the following problem when connect to remote server,

```
PS C:\Users\taotm> ssh tma26892@discovery.usc.edu
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@    WARNING: REMOTE HOST IDENTIFICATION HAS CHANGED!     @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
IT IS POSSIBLE THAT SOMEONE IS DOING SOMETHING NASTY!
Someone could be eavesdropping on you right now (man-in-the-middle attack)!
It is also possible that a host key has just been changed.
The fingerprint for the ED25519 key sent by the remote host is
SHA256:DS3uH28N7dT0tl9BiyfQWC8U2PVCZSaHLWMqm5wLa0U.
Please contact your system administrator.
Add correct host key in C:\\Users\\taotm/.ssh/known_hosts to get rid of this message.
Offending ED25519 key in C:\\Users\\taotm/.ssh/known_hosts:5
Host key for discovery.usc.edu has changed and you have requested strict checking.
Host key verification failed.
```

it is because remote computer's digital fingerprint or SHA256 hash key has changed since you last connected. Please use the command below to update the host fingerprint.

> ssh-keygen -R discovery.usc.edu

- Or download and install **PuTTY**, then use **PuTTY** to login. https://www.putty.org/

INSTALL MINICONDA

Install Miniconda in HPC-cluster

What is Conda?

Conda is an open-source **package** management system and **environment** management system. It allows you to install multiple versions of software packages and their dependencies and switch between them. It's particularly popular among data scientists because it makes it easy to manage Python and R packages.

- Install miniconda:

```
$ wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh
$ chmod 755 Miniconda3-latest-Linux-x86_64.sh
$ ./Miniconda3-latest-Linux-x86_64.sh
```

Add Miniconda to environment path

- Show your current path:
  $ pwd

- Open .bashrc file:

  $ nano ./.bashrc

- Add the following line to .bashrc file:

  export PATH="/home1/**YourLoginID**/miniconda3/bin:$PATH"

- Press key ”Ctrl + X” to exit nano and save the file.

  $ conda init
  $ exit

Reboot environment variables

- If you make any changes to .bashrc file, they won't be applied until .bashrc is re-read. Run the following command to re-read the file.

  > source ./.bashrc

- Or **Logout** and **Login** the system again.

  > ssh: <username>@discovery.usc.edu

VIRTUAL ENVIRONMENT

Create a virtual environment

- Create a virtual environment for Python & packages installation:
  $ mkdir Py310
  $ conda create -n Py310

- List the virtual environments you created:
  $ conda env list

- Activate the environment your just created for Python installation:
  $ conda activate Py310

- Deactivate an active environment whenever you don't need it:
  $ conda deactivate

Remove a virtual environment or Conda

USC Viterbi School of Engineering

- Remove a virtual environment

  $ conda env remove --name Py310

  OR

  $ rm –rf /home1/**yourloginID**/miniconda3/envs/Py310

- Remove the entire miniconda

  $ rm –rf ~/miniconda3

- Delete the following line from .bashrc

  export PATH="/home1/**YourLoginID**/miniconda3/bin:$PATH"

INSTALL PYTHON

Install Python

USC Viterbi School of Engineering

- List and activate the environment where you want to install Python:
  $ conda env list
  $ conda activate Py310

```
(base) [tma26892@discovery1 ~]$ conda env list
# conda environments:
#
base                  *  /home1/tma26892/miniconda3
Py310                    /home1/tma26892/miniconda3/envs/Py310

(base) [tma26892@discovery1 ~]$ conda activate Py310
(Py310) [tma26892@discovery1 ~]$ ls
Miniconda3-latest-Linux-x86_64.sh  new                 simple-AE.py        slurm-27421783.out  websites.txt
knn-demo.py                        read                slurm-27421759.out  test                write
miniconda3                         read-write-print.py slurm-27421767.out  test.slurm
(Py310) [tma26892@discovery1 ~]$
```

- List all versions of Python that are available to install from Conda repository
  $ conda search --full-name python

```
(Py310) [tma26892@discovery1 ~]$ conda search --full-name python
Loading channels: done
# Name                       Version           Build  Channel
python                        2.7.13     hac47a24_15  pkgs/main
python                        2.7.13     heccc3f1_16  pkgs/main
python                        2.7.13     hfff3488_13  pkgs/main
python                        2.7.14     h1571d57_29  pkgs/main
python                        2.7.14     h1571d57_30  pkgs/main
```

Install Python

- Choose the version of Python you want to install:
  $ conda install **python==3.10.15**

```
(Py310) [tma26892@discovery1 ~]$ conda install python==3.10.15
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home1/tma26892/miniconda3/envs/Py310

  added / updated specs:
    - python==3.10.15


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    ca-certificates-2024.9.24  |       h06a4308_0         130 KB
    ld_impl_linux-64-2.40      |       h12ee557_0         710 KB
    openssl-3.0.15             |       h5eee18b_0         5.2 MB
    pip-24.2                   |  py310h06a4308_0         2.3 MB
    python-3.10.15             |       he870216_1        26.8 MB
    setuptools-75.1.0          |  py310h06a4308_0         1.7 MB
    tzdata-2024b               |       h04d1e81_0         115 KB
    wheel-0.44.0               |  py310h06a4308_0         109 KB
    ------------------------------------------------------------
                                           Total:        37.1 MB
```

Install Python

USC Viterbi School of Engineering

- Choose the version of Python you want to install:
  $ conda install **python==3.10.15**

```
The following NEW packages will be INSTALLED:

  _libgcc_mutex      pkgs/main/linux-64::_libgcc_mutex-0.1-main
  _openmp_mutex      pkgs/main/linux-64::_openmp_mutex-5.1-1_gnu
  bzip2              pkgs/main/linux-64::bzip2-1.0.8-h5eee18b_6
  ca-certificates    pkgs/main/linux-64::ca-certificates-2024.9.24-h06a4308_0
  ld_impl_linux-64   pkgs/main/linux-64::ld_impl_linux-64-2.40-h12ee557_0
  libffi             pkgs/main/linux-64::libffi-3.4.4-h6a678d5_1
  libgcc-ng          pkgs/main/linux-64::libgcc-ng-11.2.0-h1234567_1
  libgomp            pkgs/main/linux-64::libgomp-11.2.0-h1234567_1
  libstdcxx-ng       pkgs/main/linux-64::libstdcxx-ng-11.2.0-h1234567_1
  libuuid            pkgs/main/linux-64::libuuid-1.41.5-h5eee18b_0
  ncurses            pkgs/main/linux-64::ncurses-6.4-h6a678d5_0
  openssl            pkgs/main/linux-64::openssl-3.0.15-h5eee18b_0
  pip                pkgs/main/linux-64::pip-24.2-py310h06a4308_0
  python             pkgs/main/linux-64::python-3.10.15-he870216_1
  readline           pkgs/main/linux-64::readline-8.2-h5eee18b_0
  setuptools         pkgs/main/linux-64::setuptools-75.1.0-py310h06a4308_0
  sqlite             pkgs/main/linux-64::sqlite-3.45.3-h5eee18b_0
  tk                 pkgs/main/linux-64::tk-8.6.14-h39e8969_0
  tzdata             pkgs/main/noarch::tzdata-2024b-h04d1e81_0
  wheel              pkgs/main/linux-64::wheel-0.44.0-py310h06a4308_0
  xz                 pkgs/main/linux-64::xz-5.4.6-h5eee18b_1
  zlib               pkgs/main/linux-64::zlib-1.2.13-h5eee18b_1


Proceed ([y]/n)?
```

Install Python

- Check all versions of Python have been installed in HPC-cluster including the ones installed by the system administrator and the version you just installed:

$ compgen -c python

```
(Py310) [tma26892@discovery1 ~]$ compgen -c python
python3.1
python3.10
python3
python
python3-config
python3.10-config
python3
python2
python2.7
python3.6m
python3.6
python
(Py310) [tma26892@discovery1 ~]$
```

- Verify the version 3.10.15 of Python you installed:

$ python

```
(Py310) [tma26892@discovery1 ~]$ python
Python 3.10.15 (main, Oct  3 2024, 07:27:34) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

INSTALL PYTHON LIBRARIES

Install Python Libraries

$ conda install -n Py310 pandas

```
(Py310) [tma26892@discovery1 ~]$ conda install -n Py310 pandas
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home1/tma26892/miniconda3/envs/Py310

  added / updated specs:
    - pandas
```

$ conda install -n Py310 scikit-learn

```
(Py310) [tma26892@discovery1 ~]$ conda install -n Py310 scikit-learn
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home1/tma26892/miniconda3/envs/Py310

  added / updated specs:
    - scikit-learn
```

Install Python Libraries

USC Viterbi School of Engineering

- Install python libraries from **conda repository**:

$ conda install -n Py310 matplotlib
$ conda install -n Py310 xlrd

OR

- Install libraries with **pip repository**

$ pip install --upgrade pip
$ pip install matplotlib
$ pip install pandas
$ pip install scikit-learn
$ pip install **tensorflow-cpu==2.10.0**
$ pip install keras
$ pip install wget
$ pip install umap-learn

List Python Libraries

USC Viterbi School of Engineering

- List python libraries that have been installed in your virtual environment:

$ conda list

```
(Py310) [tma26892@discovery1 ~]$ conda list
# packages in environment at /home1/tma26892/miniconda3/envs/Py310:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                        main
_openmp_mutex             5.1                       1_gnu
absl-py                   2.1.0           py310h06a4308_0
astunparse                1.6.3                      py_0
blas                      1.0                         mkl
bottleneck                1.4.2           py310ha9d4c09_0
brotli                    1.0.9                h5eee18b_8
brotli-bin                1.0.9                h5eee18b_8
brotli-python             1.0.9           py310h6a678d5_8
bzip2                     1.0.8                h5eee18b_6
c-ares                    1.19.1               h5eee18b_0
ca-certificates           2024.9.24            h06a4308_0
certifi                   2024.8.30       py310h06a4308_0
charset-normalizer        3.3.2              pyhd3eb1b0_0
contourpy                 1.2.0           py310hdb19cb5_0
cycler                    0.11.0             pyhd3eb1b0_0
cyrus-sasl                2.1.28               h52b45da_1
dbus                      1.13.18              hb2f20db_0
expat                     2.6.3                h6a678d5_0
flatbuffers               24.3.25              h6a678d5_0
fontconfig                2.14.1               h55d465d_3
fonttools                 4.51.0          py310h5eee18b_0
freetype                  2.12.1               h4a9f257_0
```

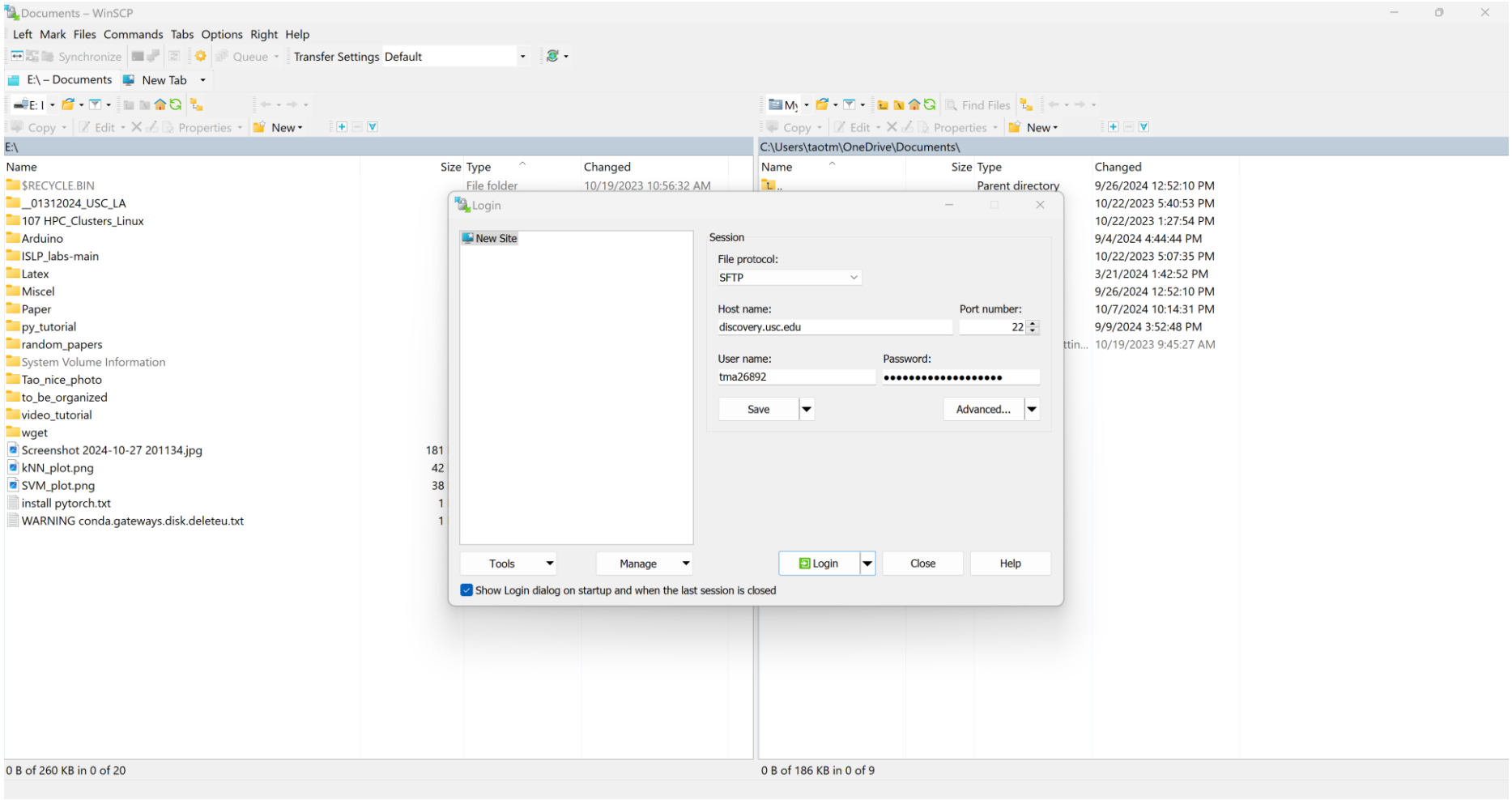
```
scikit-learn              1.5.1
scipy                     1.14.1
setuptools                75.1.0
sip                       6.7.12
six                       1.16.0
snappy                    1.2.1
sqlite                    3.45.3
tbb                       2021.8.0
tensorboard               2.17.0
tensorboard-data-server   0.7.0
tensorflow                2.17.0
tensorflow-base           2.17.0
tensorflow-estimator      2.17.0
termcolor                 2.1.0
threadpoolctl             3.5.0
tk                        8.6.14
tomli                     2.0.1
tornado                   6.4.1
typing-extensions         4.11.0
typing_extensions         4.11.0
tzdata                    2024b
unicodedata2              15.1.0
urllib3                   2.2.3
```

FILE TRANSFER

Install WinSCP

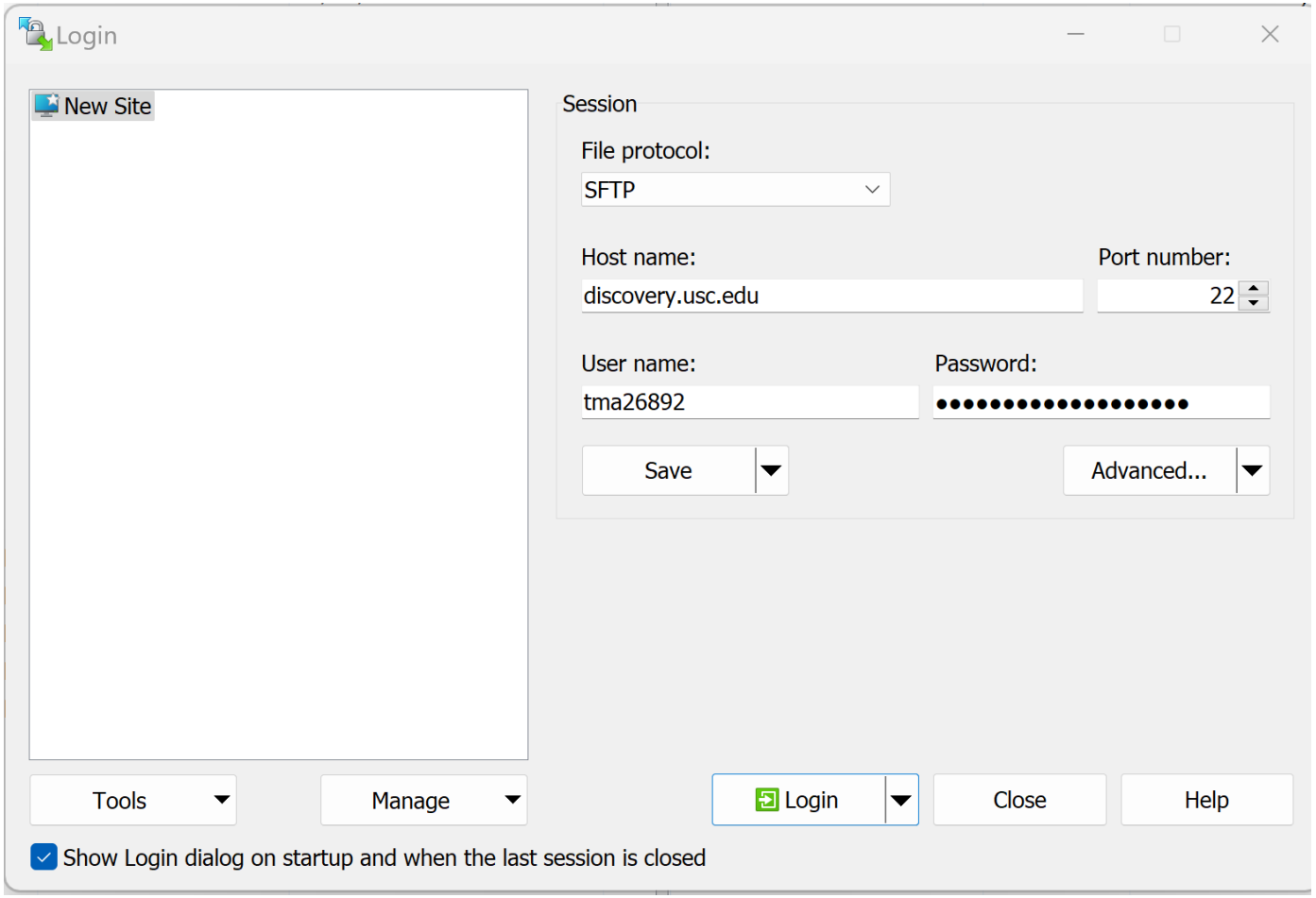
USC Viterbi School of Engineering

- WinSCP is an open source free SFTP client, FTP client, WebDAV client, S3 client and SCP client and file manager for Windows. Its main function is file transfer between a local and a remote computer. Beyond this, WinSCP offers scripting and basic file manager functionality.
- https://winscp.net/eng/docs/introduction



Transfer files to HPC-cluster

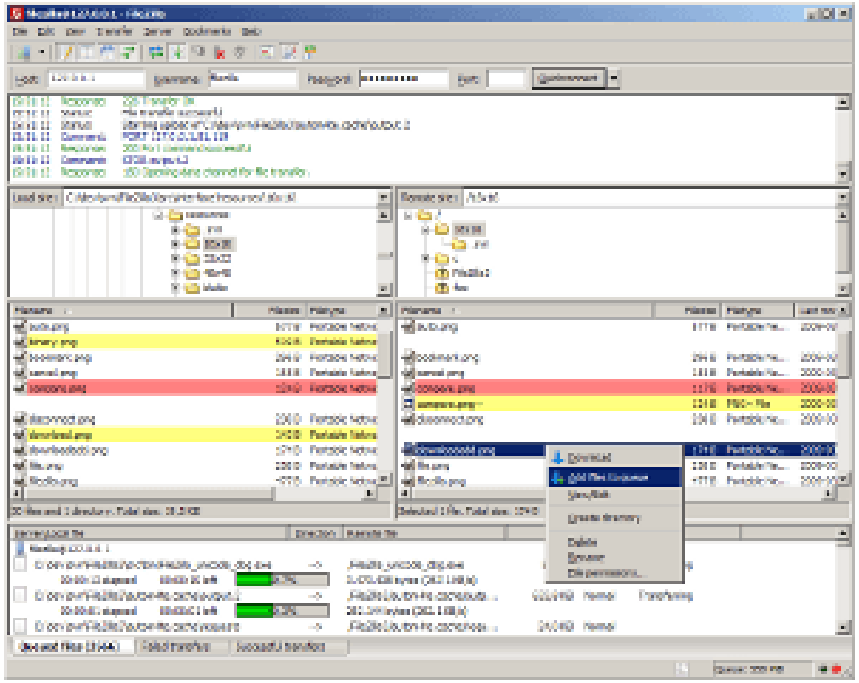
- Install WinSCP
- Setup connection



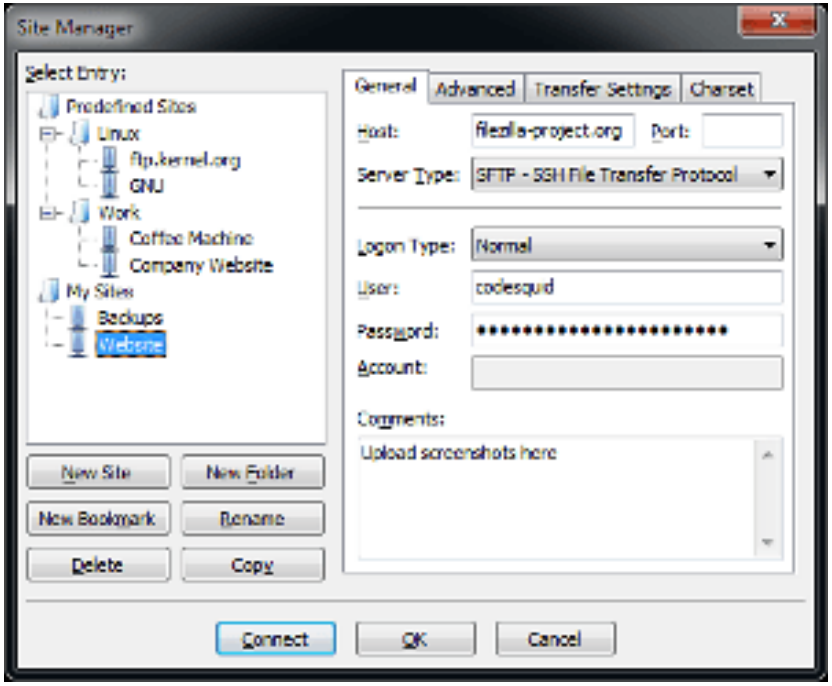
Transfer files to HPC-cluster

- Install FileZilla®, the free FTP solution for all other platforms, e.g., MacOS
- Download: https://filezilla-project.org/
- Setup connection

Main Window



Network Connection



SUBMIT JOB TO HPC

Submit a job to HPC-cluster

- Slurm is an open source, fault-tolerant, and highly scalable **cluster management** and **job scheduling system** for large and small Linux clusters.

- Documentation about Slurm can be found here
  https://slurm.schedmd.com/documentation.html

- Submit a job to Slurm management system:
  $ sbatch test.slurm

- Check the status of your job:
  $ squeue –u <yourloginID>

Slurm file example

```
#!/bin/bash
#SBATCH --account=tma26892_1425
#SBATCH --partition=main
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=4
#SBATCH --mem=16G
#SBATCH --time=2:00:00

echo "“
echo "Starting at `date`“
echo "Running on hosts: $SLURM_NODELIST“
echo "Running on $SLURM_NNODES nodes.“
echo "Running on $SLURM_NPROCS processors.“

# Move to the correct directory
cd /home1/tma26892/code

echo "Current working directory is `pwd`“

# Train & Test the kNN model
python knn-demo.py

# end of the program
echo "“
echo "Program finished with exit code $? at: `date`“
echo ""
```