For Windows & Mac user:

The objective is to upload your python code (.py file) to the server and execute the file. If you already have knowledge on file transfer, you can jump to Step 3 &4.

1. Compile python script using your local computer.
2. Transfer python file (.py) into HADOOP server (Host: fscresearchvm23.fsc.stevens.edu) using FileZilla.

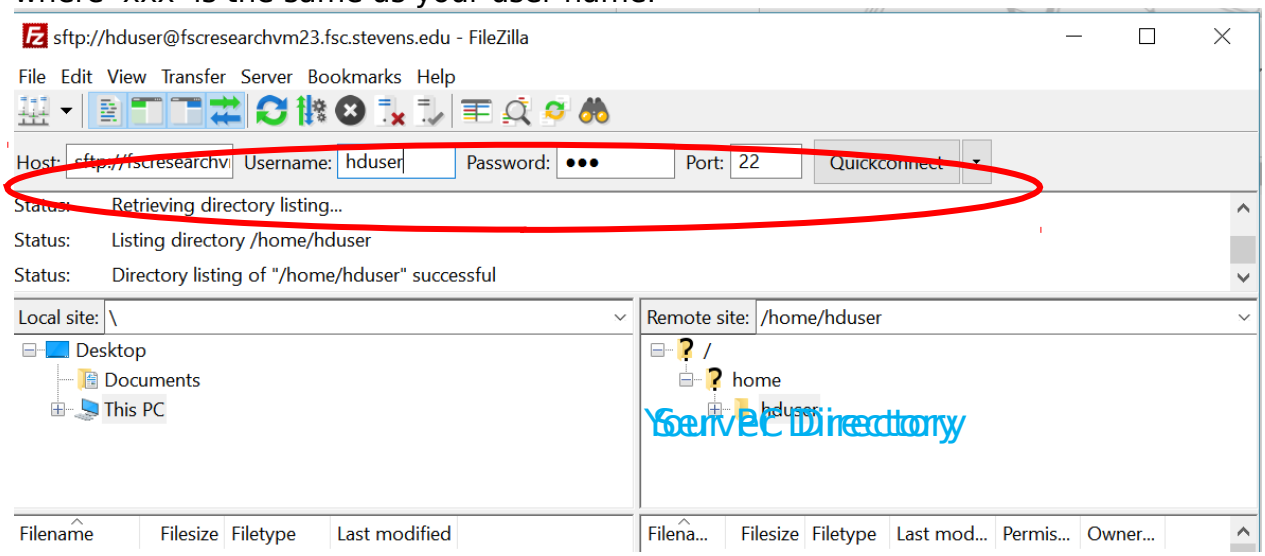
2.1. Download and install FileZilla (windows: https://filezilla-project.org/download.php?platform=win64, Mac: https://filezilla-project.org/download.php?platform=osx)

2.2. Open FileZilla. Fill the information (Host, Username, Password), Port is 22. Click Quickconnect button. (see red circle)

2.3. You are able to see successful information on status panel, and see the files and directories of your local site and remote site. (see blue)

2.4. The listed files are files in the selected directory. For example, in the screenshot, I selected the personal folder called 'hduser'. All files inside hduser folder are listed in the bottom.

2.5. You need to upload your python file into the user's personal folder /home/xxx where 'xxx' is the same as your user name.

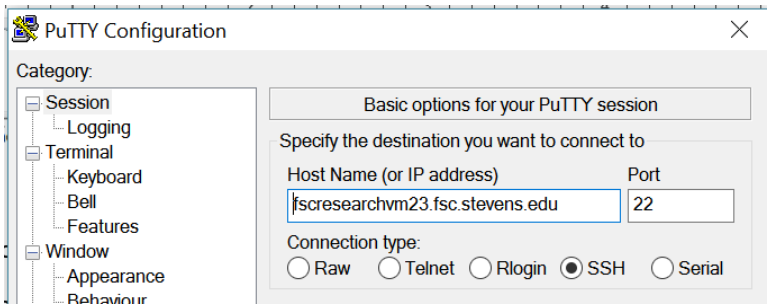


3. Use command to execute python file.

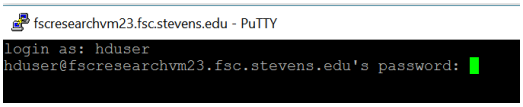
For Windows user:

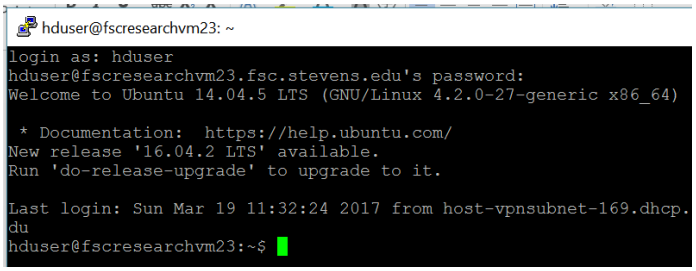
3.1. Download putty (https://the.earth.li/~sgtatham/putty/latest/w64/putty.exe)

3.2. Put Host Name and click Open button on the bottom.



3.3. You need to put user name and password. (Note: When you put password, there is no displayed character. After you typing the password, hit Enter, and it will login for you.)





For mac user:

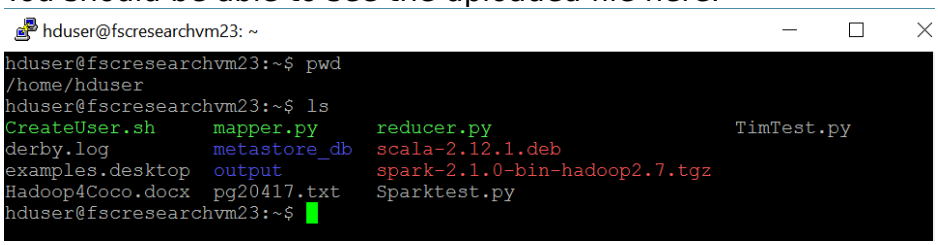
3.1. Open Terminal.

3.2. Type ssh xxx@fscresearchvm23.fsc.stevens.edu, click Enter.

3.3. Input password.

The following 2 commands are the same for both Windows & Mac users

3.4. pwd command will show you current directory. By default, the directory should be your personal folder, where is the place you uploaded your file in previous step. ls command will show you a list of files under current directory. You should be able to see the uploaded file here.



For all users to execute file:

4. Execute python file using spark.

4.1. Type command cd $SPARK_HOME.

4.2. Type command ./bin/spark-submit --master spark://fscresearchvm23.fsc.stevens.edu:7077 /path/filename.py