CONNECTING TO JUPYTER NOTEBOOK (… RUNNING ON THE VM …)

DS 203

Overview

- In the earlier steps you have created the VM in the cloud and connected to it using the following programs: PuTTy / ssh, WinSCP / scp, DBeaver
- The VM also has **Jupyter Notebook** and **SPARK** installed on it:
- This deck outlines the procedures and tools to:
  - Start Jupyter Notebook on the VM
  - Create a network **tunnel** between your local computer and the remote VM
  - Connect to the Jupyter Notebook using a browser running on your local computer
  - Use SPARK from within a Python Notebook.

Pre-requisites …

From here onwards this document assumes the following:

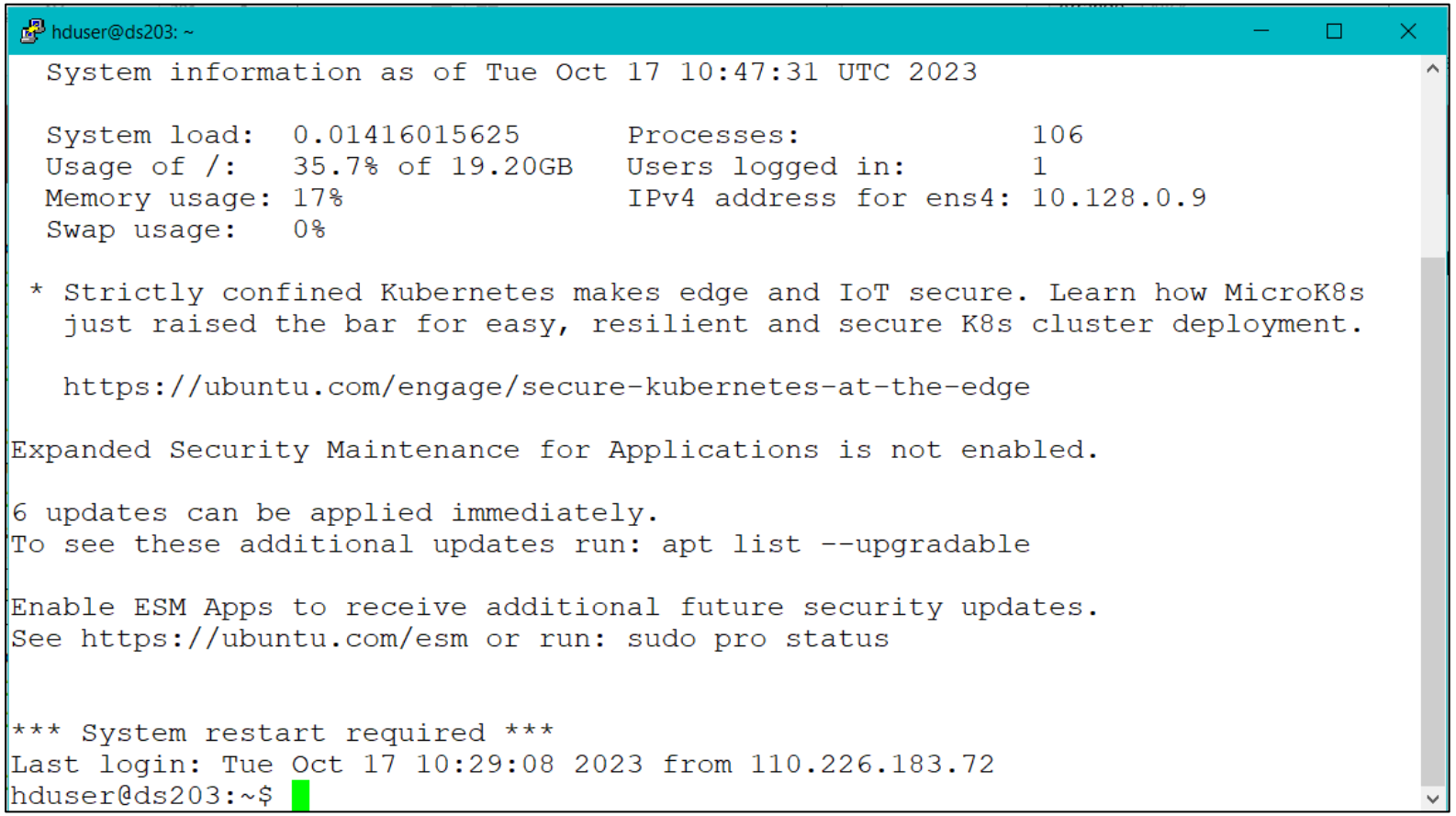
- That the VM is ‘up and running’
- That you know the public IP of the VM
- That you have downloaded the following files from Moodle:
  - **ds203-azure-vm-rsa** (mac OS, Linux, Unix)
  - **ds203-azure-vm.ppk** (Windows)
- That you have successfully made connections to the VM using PuTTY / ssh (Please refer to the document **03-Starting the VM** …)

In case of difficulties …

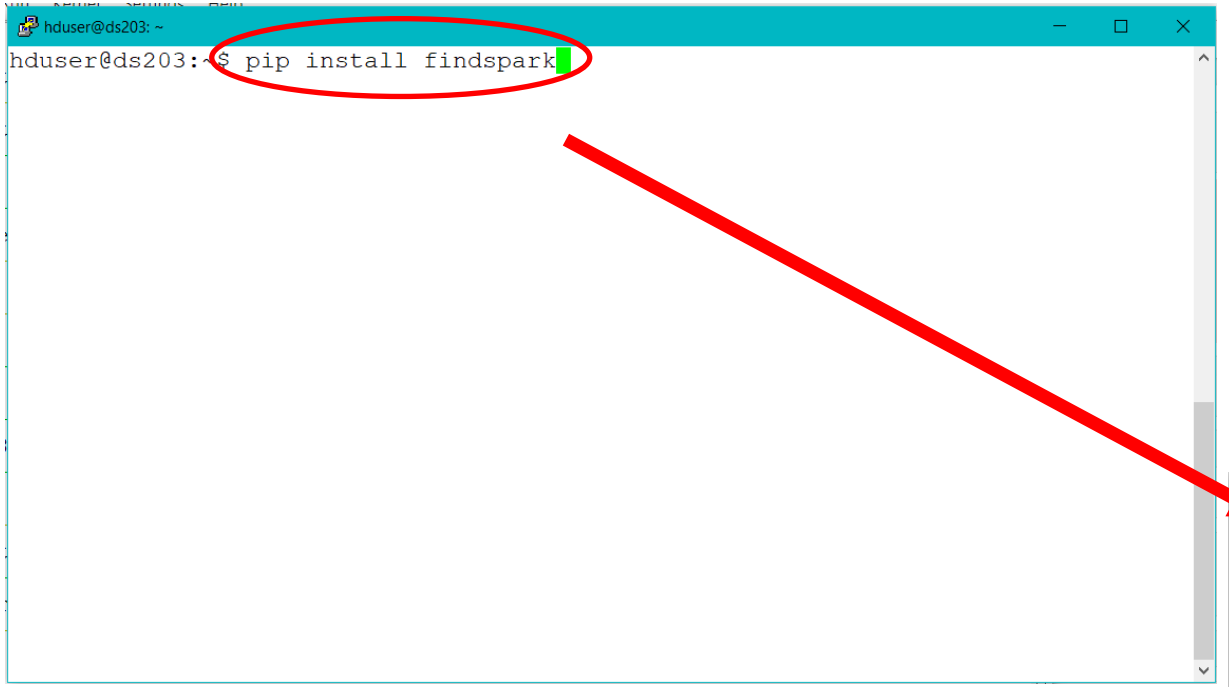
- Log your issues in the Moodle Forum **Queries and Discussions** and a member of the TA team will respond and guide you.

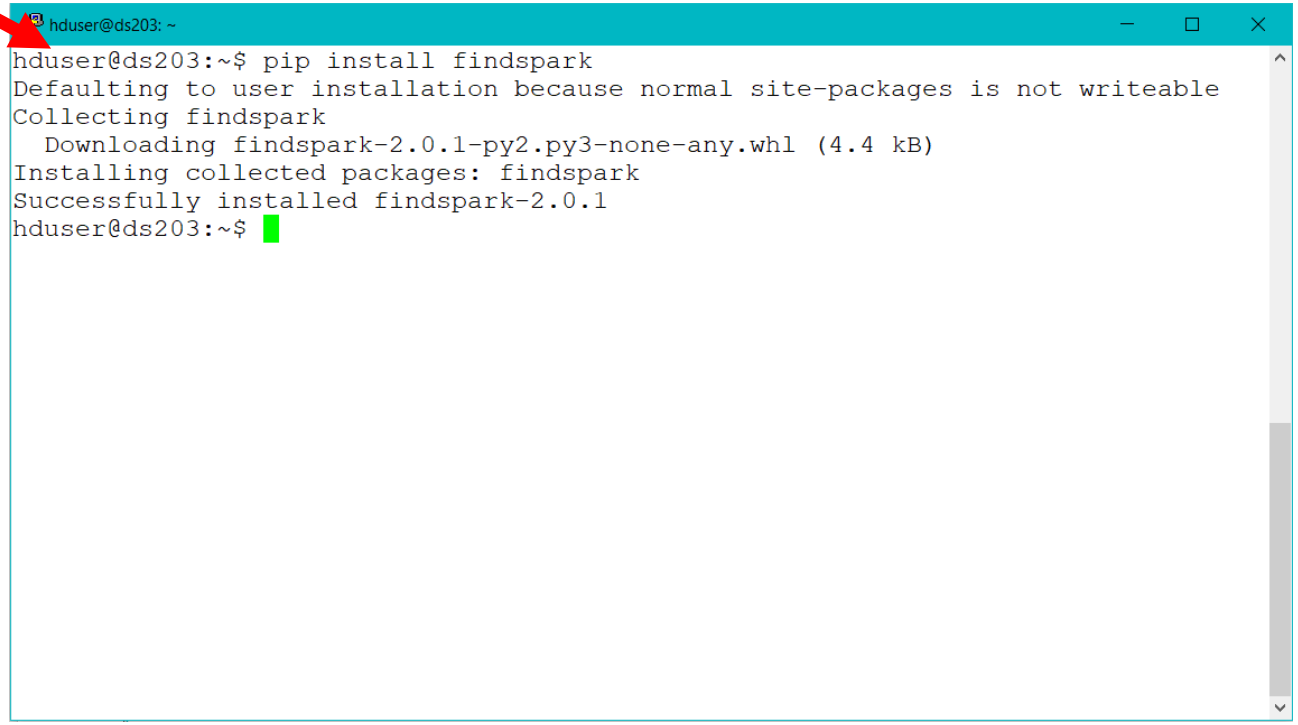
Make a connection to the VM

Using the appropriate terminal program (see 03-Starting the VM and connecting to it.pdf), login as **hduser**

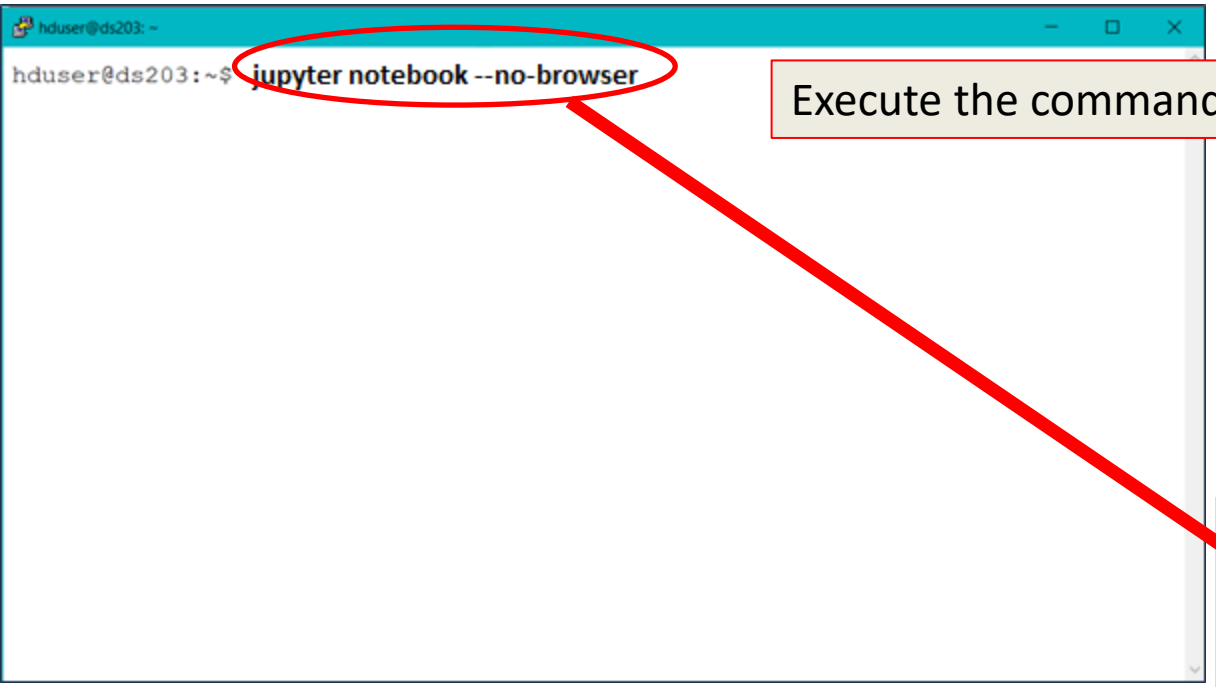


Install the Python package **findspark** (one time activity …)





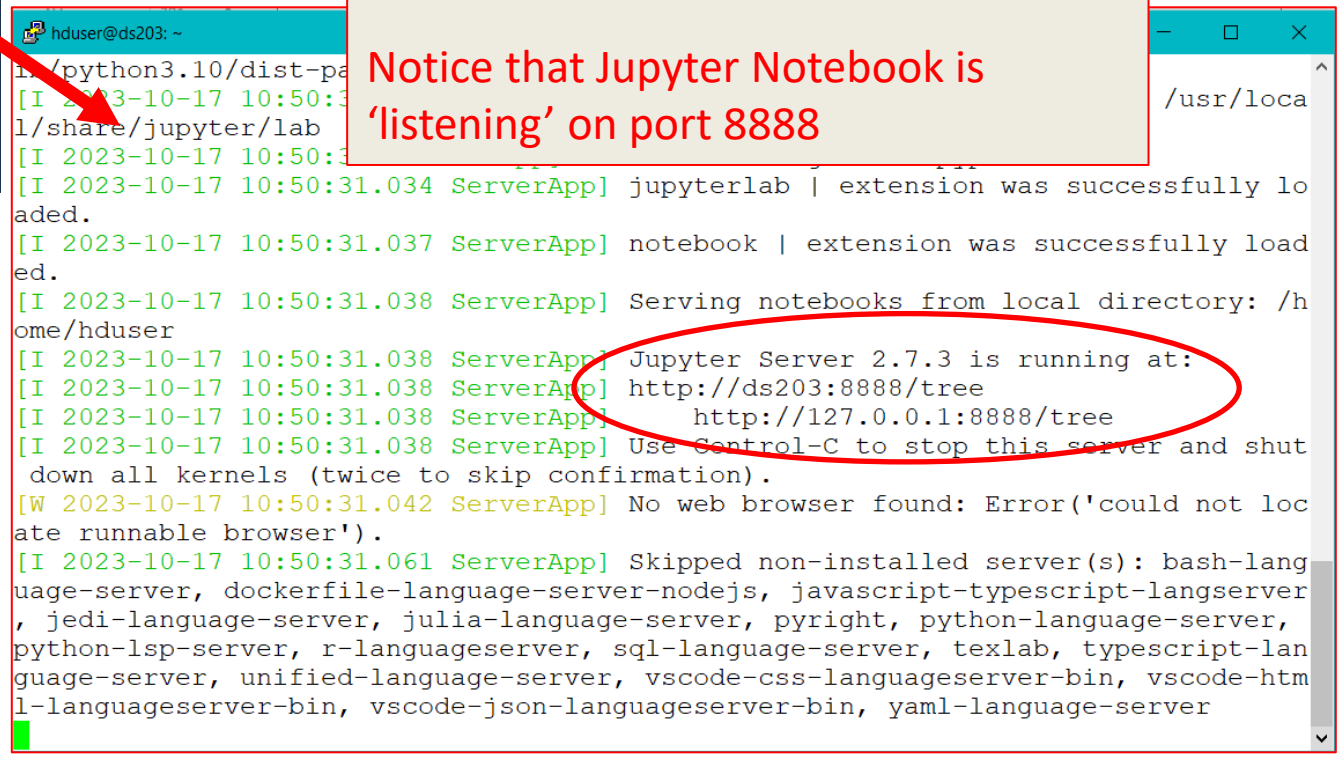
Start **jupyter notebook** on the VM



Execute the command **jupyter notebook --no-browser** on the VM

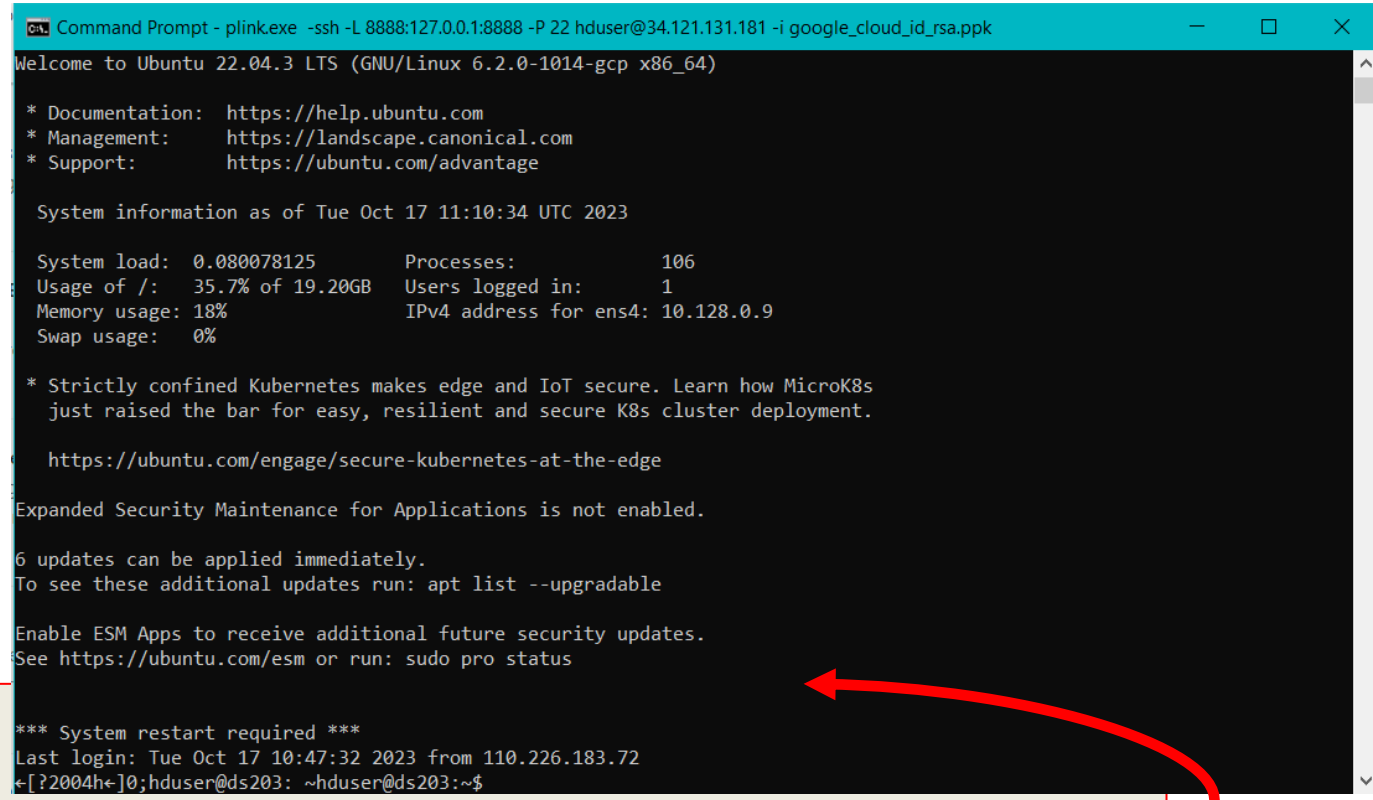
Jupyter Notebooks starts running in terminal mode – ready to take external connections …

Notice that Jupyter Notebook is ‘listening’ on port 8888



Creating a **tunnel** to connect to Jupyter Notebook (**Windows**)

- Our goal is to access this VM based Jupyter Notebook using a browser running on the local computer.
- To accomplish this goal, we have to create a **tunnel** from 'port 8888' of the local computer to 'port 8888' of the VM … as explained in this and the next slide:



Computers running **Windows** OS

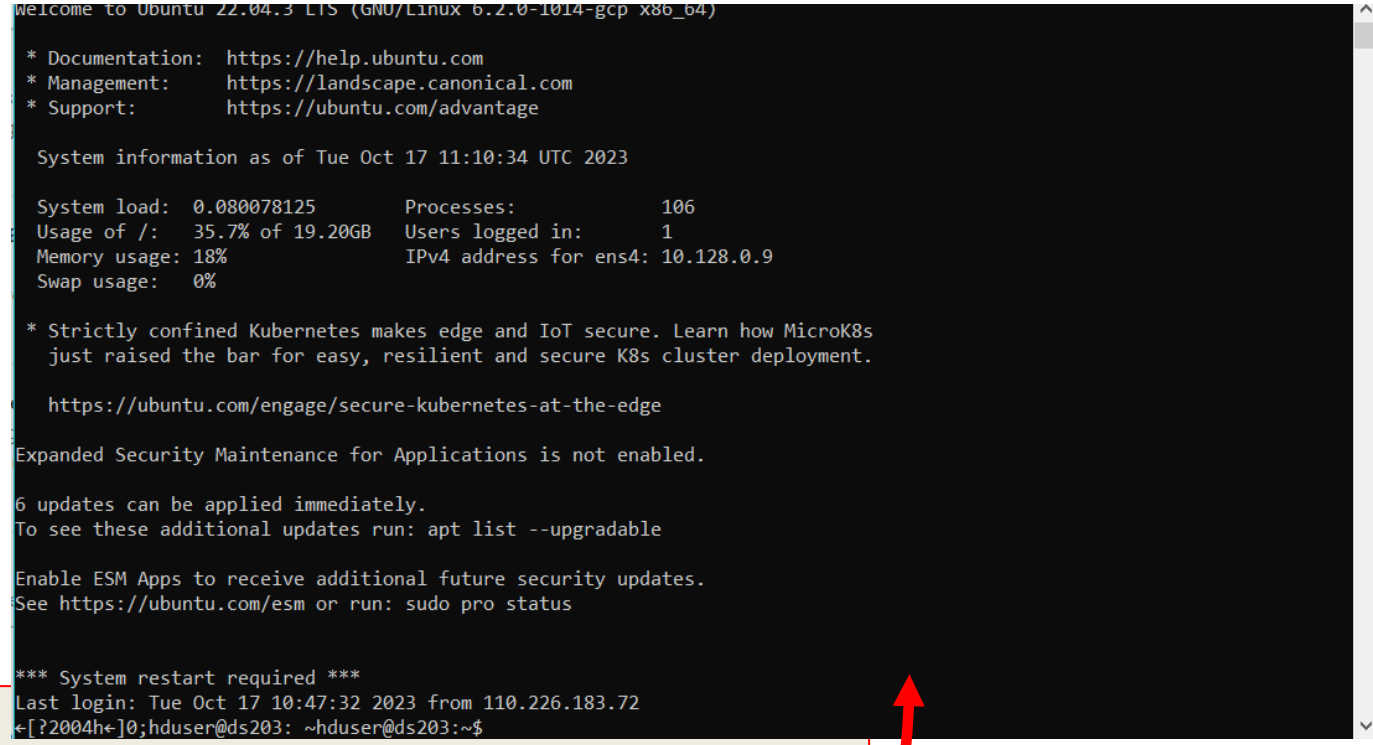
1. Open a command window
2. In this command window, execute the following command:
   - **plink.exe** -ssh -L 8888:127.0.0.1:8888 -P 22 hduser@**X.Y.Z.W** -i <type_full_path_to_the_file>\ds203-azure-vm.ppk
   - For example: plink.exe -ssh 8888:127.0.0.1:8888 -P 22 hduser@X.Y.Z.W -i c:\tools\putty\ds203-azure-vm.ppk
   - (Note: **X.Y.Z.W should be replaced with the public IP of the VM)**

(This tunnel connects port 8888 of the local computer to port 8888 of the remote VM)

- If the above command does not succeed (mostly due to multiple programs trying to access port 8888), try using **port 47** instead of port 8888
- **plink.exe** -ssh -L **47**:127.0.0.1:8888 -P 22 hduser@**X.Y.Z.W** -i <type_full_path_to_the_file>\ds203-azure-vm.ppk

Creating a **tunnel** to connect to Jupyter Notebook (**Linux / Mac OS**)

- Our goal is to access this Jupyter Notebook using a browser running on the local computer.
- To accomplish this goal, we have to create a **tunnel** from the local computer to the VM …

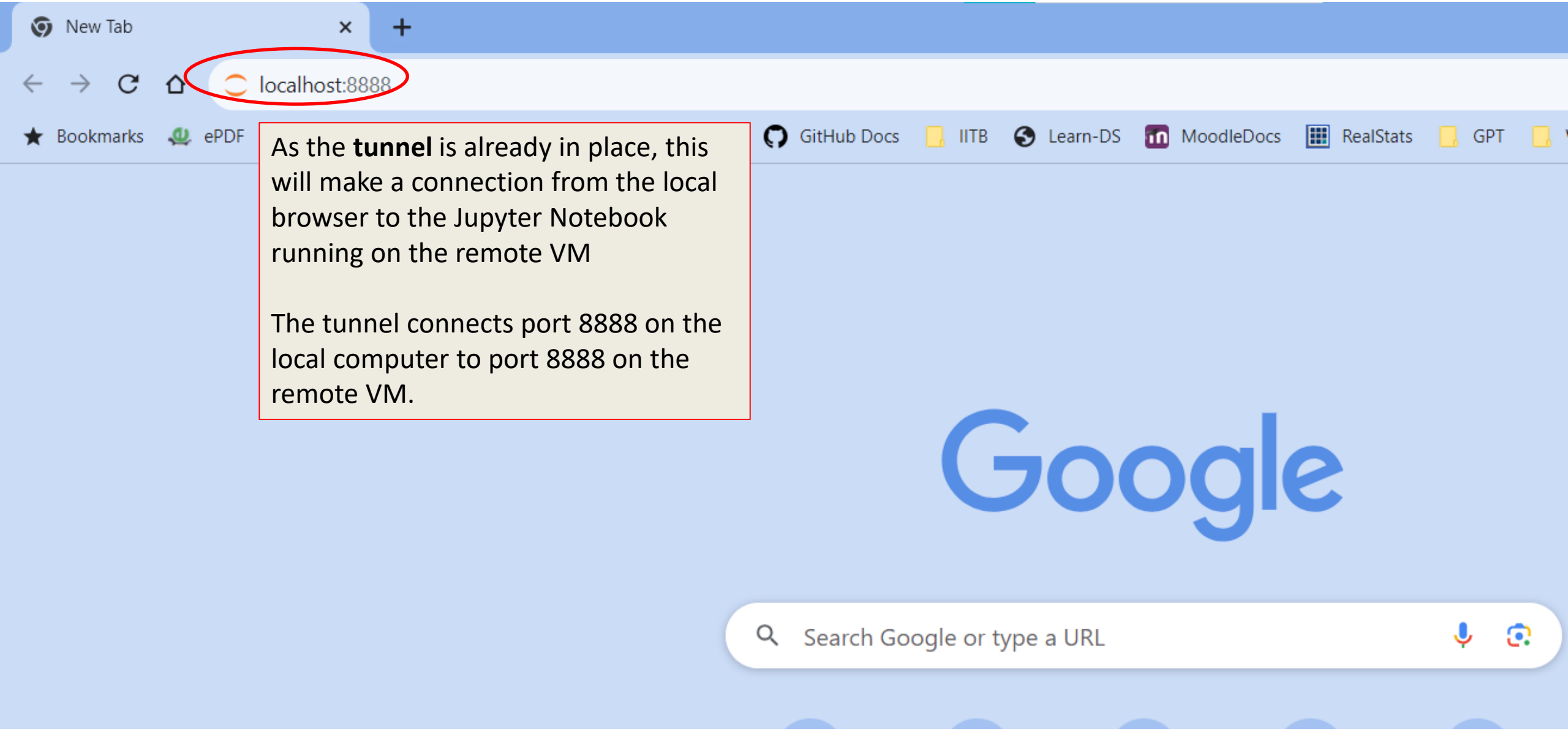


Computers running **Unix / Linux / Mac OS**

1. Open a terminal window
2. Execute the following command:
   - `ssh -L 8888:127.0.0.1:8888 -i <full_path>/ds203-azure-vm-rsa hduser@`**`X.Y.Z.W`**
   - For example: `ssh -L 8888:127.0.0.1:8888 -i /home/rajani/ds203-azure-vm-rsa hduser@`**`X.Y.Z.W`**
   - (Note: **X.Y.Z.W should be replaced with the public IP of the VM)**

(This tunnel connects port 8888 of the local computer to port 8888 of the remote VM)

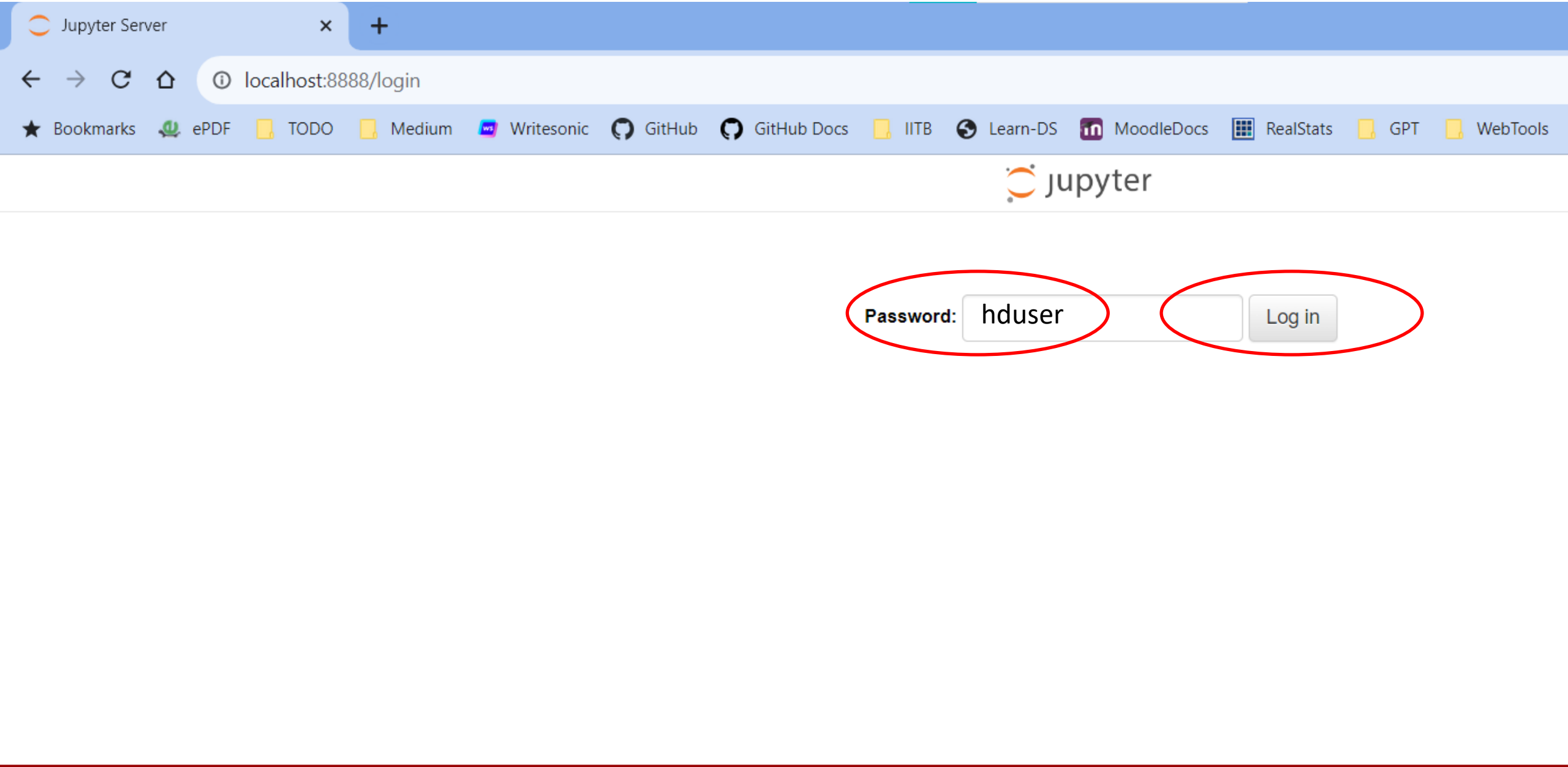
Open a browser windows on your local computer …

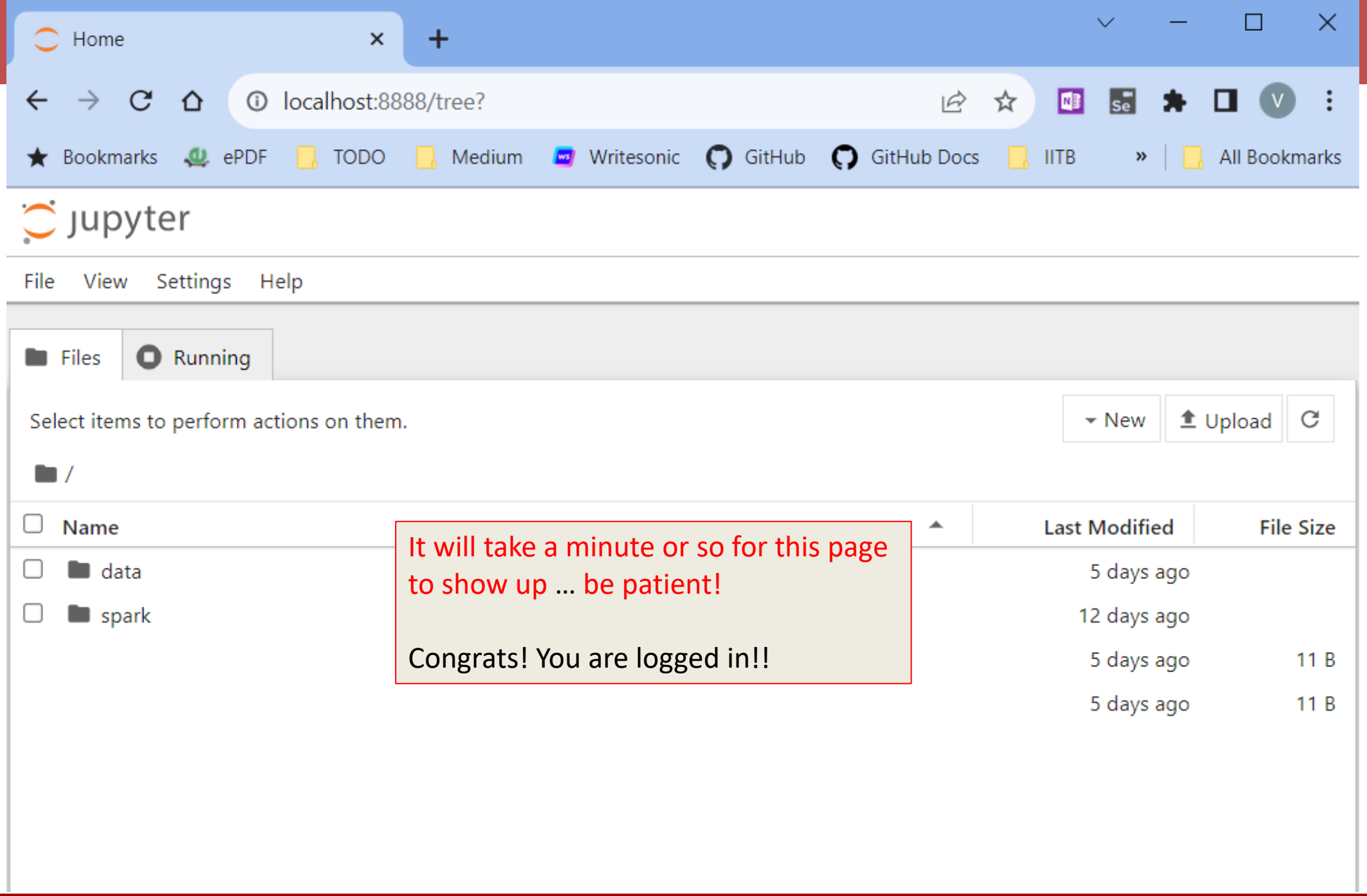


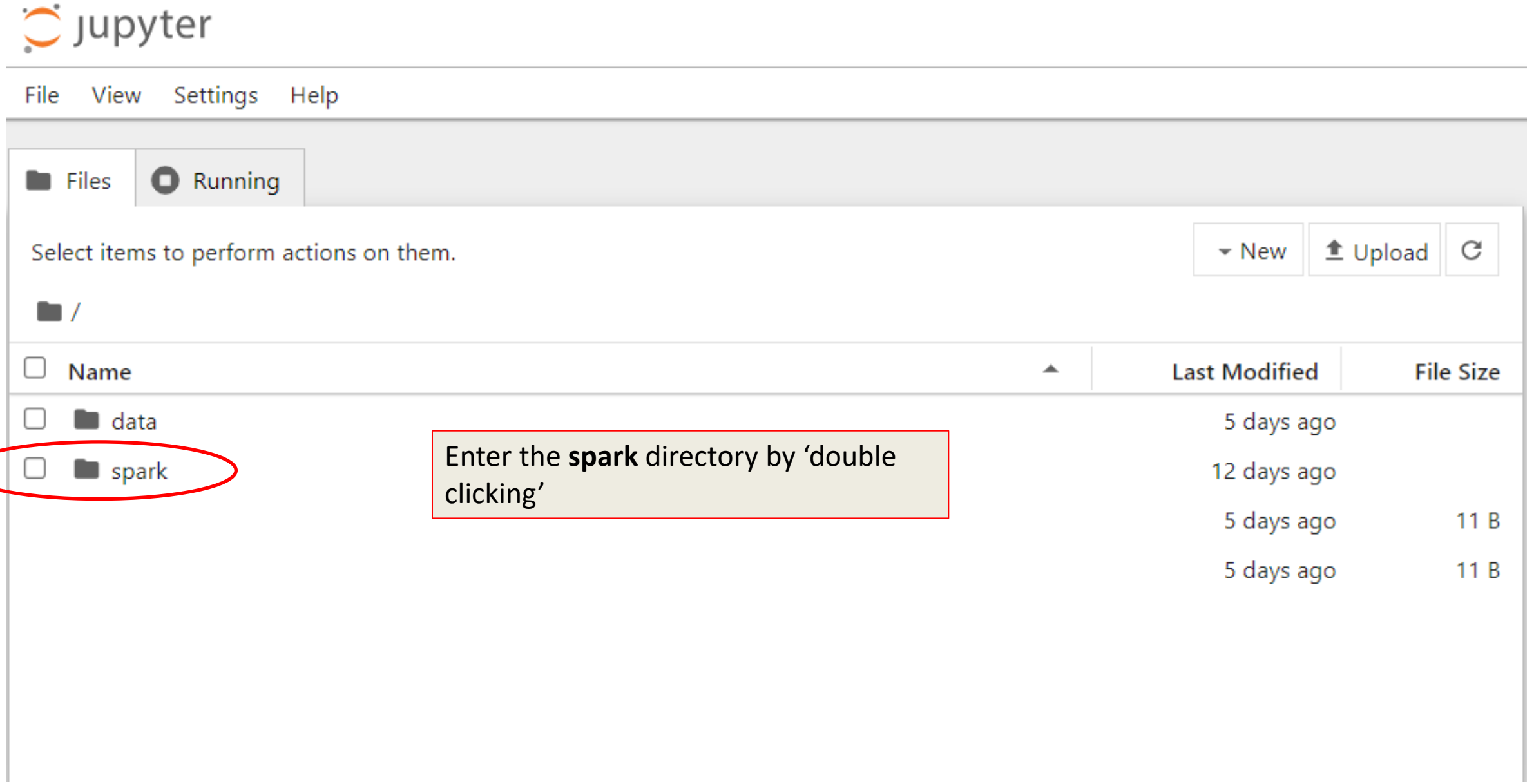
As the **tunnel** is already in place, this will make a connection from the local browser to the Jupyter Notebook running on the remote VM

The tunnel connects port 8888 on the local computer to port 8888 on the remote VM.

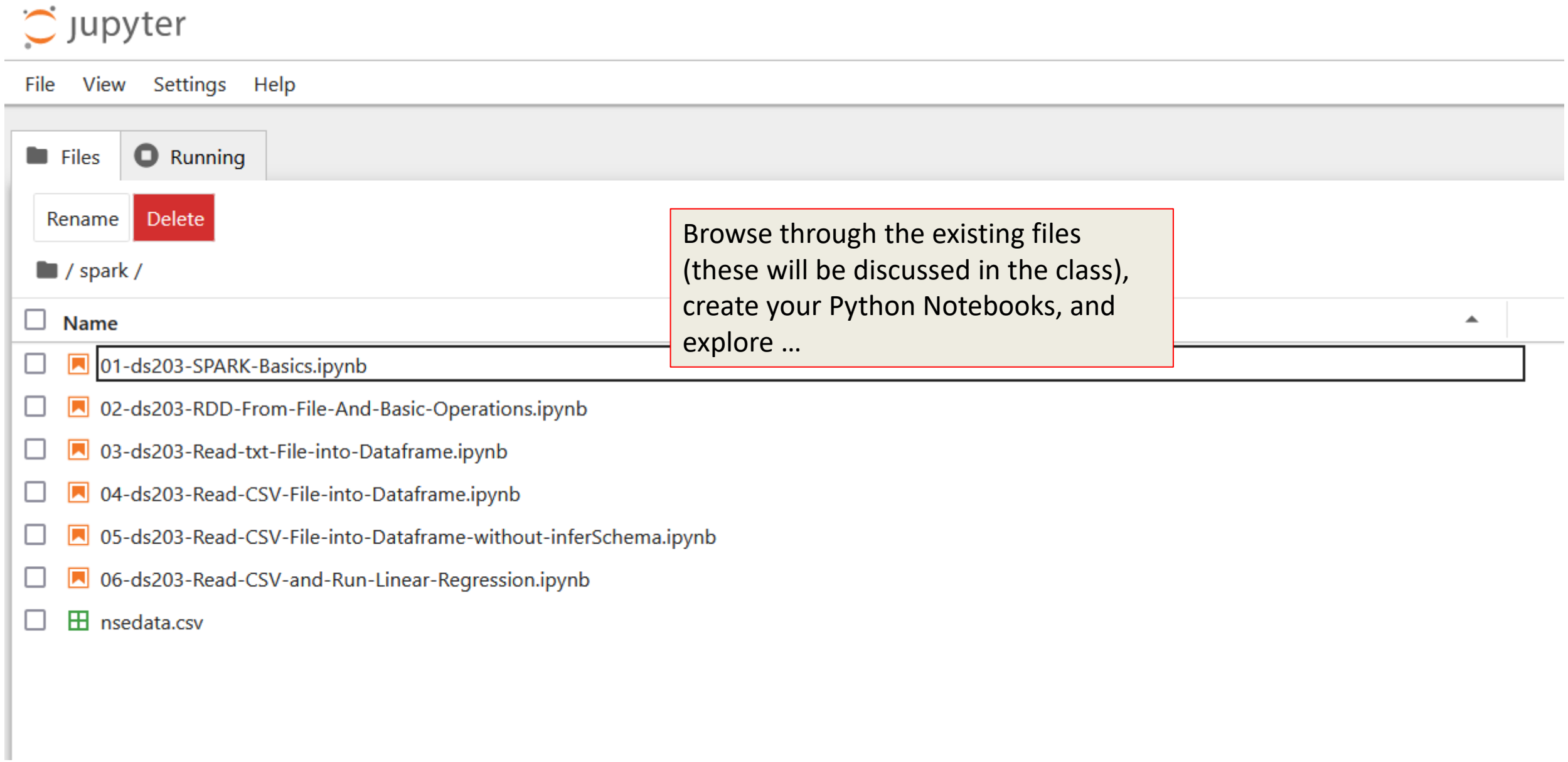
Logging into the Jupyter Notebook




Home
localhost:8888/tree?
Bookmarks
ePDF
TODO
Medium
Writesonic
GitHub
GitHub Docs
IITB
All Bookmarks
jupyter
File
View
Settings
Help
Files
Running
Select items to perform actions on them.
New
Upload
/
Name
Last Modified
File Size
data
5 days ago
spark
12 days ago
5 days ago
11 B
5 days ago
11 B
It will take a minute or so for this page to show up … be patient!
Congrats! You are logged in!!


jupyter
File View Settings Help
Files
Running
Select items to perform actions on them.
New
Upload
/
Name
Last Modified
File Size
data
5 days ago
spark
12 days ago
5 days ago 11 B
5 days ago 11 B
Enter the **spark** directory by 'double clicking'

Browse through the Notebooks …



Browse through the existing files (these will be discussed in the class), create your Python Notebooks, and explore …

Once you are done with your work …

- In the browser, under Jupyter, select File / Shutdown
  - This shuts down the Notebook on the remote VM
- In the command or terminal (where you have set up the tunnel), type **exit** to *close* the tunnel
- Close the PuTTy or ssh terminal that you have opened to the VM
- If you are done with working on the VM, do not forget to SHUT the VM
  - Else … you will run out of your credit sooner … and there will be no re-charge!

IMPORTANT ** IMPORTANT ** IMPORTANT

- Once you are through with your work on / with the VM, be sure to **STOP** it to pause the Billing for this resource!