

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:

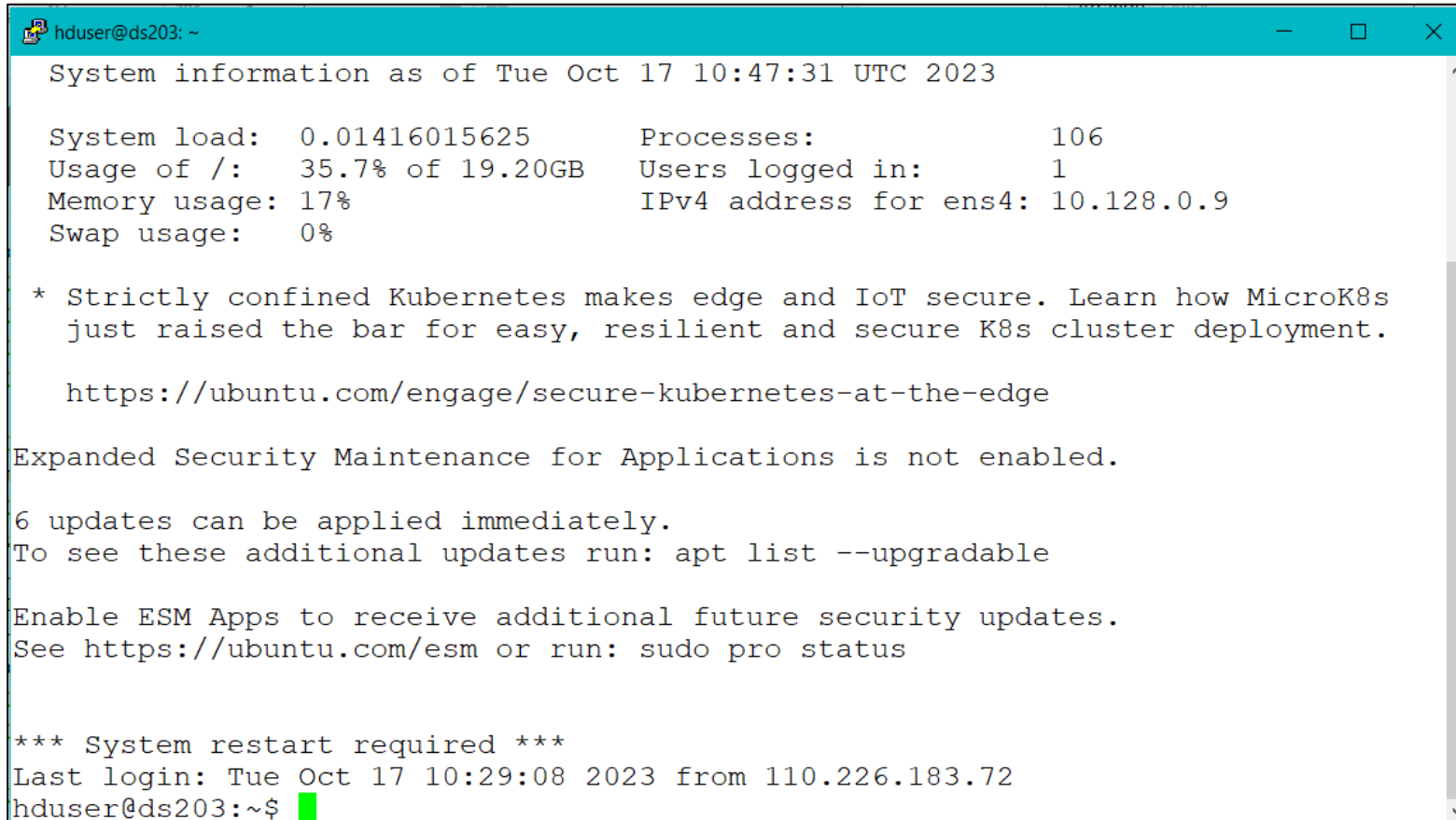
- That the VM is 'up and running'
- That you know the public IP of the VM
- That you have downloaded the following files from Moodle:
 - **ds203-azure-vm-rsa** (mac OS, Linux, Unix)
 - **ds203-azure-vm.ppk** (Windows)
- That you have successfully made connections to the VM using PuTTY / ssh (**Please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



```
hduser@ds203: ~  
System information as of Tue Oct 17 10:47:31 UTC 2023  
  
System load:  0.01416015625      Processes:            106  
Usage of /:   35.7% of 19.20GB   Users logged in:     1  
Memory usage: 17%               IPv4 address for ens4: 10.128.0.9  
Swap usage:   0%  
  
* Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s  
  just raised the bar for easy, resilient and secure K8s cluster deployment.  
  
  https://ubuntu.com/engage/secure-kubernetes-at-the-edge  
  
Expanded Security Maintenance for Applications is not enabled.  
  
6 updates can be applied immediately.  
To see these additional updates run: apt list --upgradable  
  
Enable ESM Apps to receive additional future security updates.  
See https://ubuntu.com/esm or run: sudo pro status  
  
*** System restart required ***  
Last login: Tue Oct 17 10:29:08 2023 from 110.226.183.72  
hduser@ds203:~$
```

Install the Python package findspark (one time activity ...)

```
hduser@ds203: ~  
hduser@ds203:~$ pip install findspark
```

```
hduser@ds203: ~  
hduser@ds203:~$ pip install findspark  
Defaulting to user installation because normal site-packages is not writeable  
Collecting findspark  
  Downloading findspark-2.0.1-py2.py3-none-any.whl (4.4 kB)  
Installing collected packages: findspark  
Successfully installed findspark-2.0.1  
hduser@ds203:~$
```

Start jupyter notebook on the VM

```
hduser@ds203: ~  
hduser@ds203:~$ jupyter notebook --no-browser
```

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

```
hduser@ds203: ~  
I /python3.10/dist-pa  
[I 2023-10-17 10:50:31.034 ServerApp] jupyterlab | extension was successfully lo  
ad ed.  
[I 2023-10-17 10:50:31.037 ServerApp] notebook | extension was successfully load  
ed.  
[I 2023-10-17 10:50:31.038 ServerApp] Serving notebooks from local directory: /h  
ome/hduser  
[I 2023-10-17 10:50:31.038 ServerApp] Jupyter Server 2.7.3 is running at:  
[I 2023-10-17 10:50:31.038 ServerApp] http://ds203:8888/tree  
[I 2023-10-17 10:50:31.038 ServerApp] http://127.0.0.1:8888/tree  
[I 2023-10-17 10:50:31.038 ServerApp] Use Control-C to stop this server and shut  
down all kernels (twice to skip confirmation).  
[W 2023-10-17 10:50:31.042 ServerApp] No web browser found: Error('could not loc  
ate runnable browser').  
[I 2023-10-17 10:50:31.061 ServerApp] Skipped non-installed server(s): bash-lang  
uage-server, dockerfile-language-server-nodejs, javascript-typescript-langserver  
, jedi-language-server, julia-language-server, pyright, python-language-server,  
python-lsp-server, r-languageserver, sql-language-server, texlab, typescript-lan  
guage-server, unified-language-server, vscode-css-languageserver-bin, vscode-htm  
l-languageserver-bin, vscode-json-languageserver-bin, yaml-language-server
```

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:

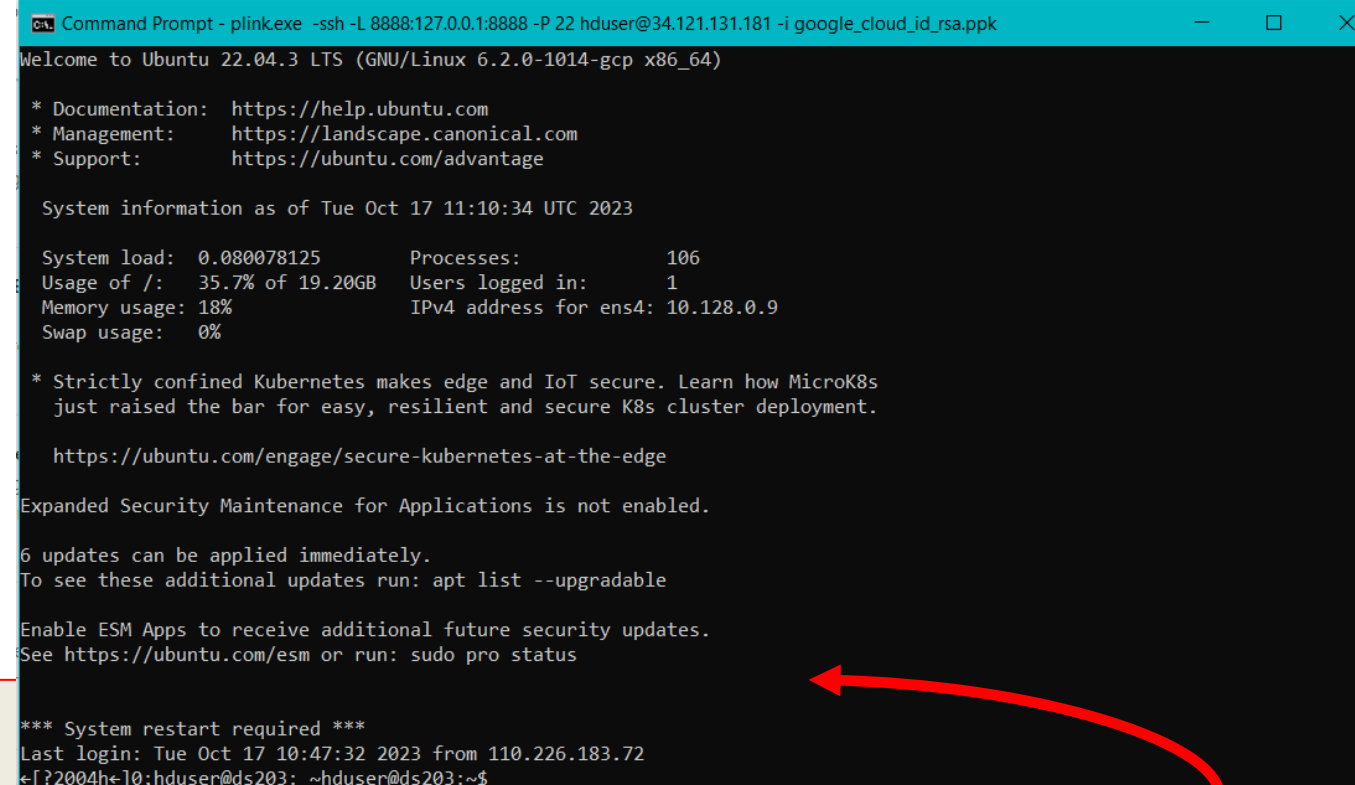


Computers running **Windows** OS

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



```
CA Command Prompt - plink.exe -ssh -L 8888:127.0.0.1:8888 -P 22 hduser@34.121.131.181 -i google_cloud_id_rsa.ppk
Welcome to Ubuntu 22.04.3 LTS (GNU/Linux 6.2.0-1014-gcp x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Tue Oct 17 11:10:34 UTC 2023

System load:  0.080078125      Processes:            106
Usage of /:   35.7% of 19.20GB Users logged in:          1
Memory usage: 18%             IPv4 address for ens4: 10.128.0.9
Swap usage:   0%

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
   just raised the bar for easy, resilient and secure K8s cluster deployment.

https://ubuntu.com/engage/secure-kubernetes-at-the-edge

Expanded Security Maintenance for Applications is not enabled.

6 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

*** System restart required ***
Last login: Tue Oct 17 10:47:32 2023 from 110.226.183.72
[?2004h+]0;hduser@ds203: ~hduser@ds203:~$
```


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)

```
Welcome to Ubuntu 22.04.3 LTS (GNU/Linux 6.2.0-1014-gcp x86_64)

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:        https://ubuntu.com/advantage

System information as of Tue Oct 17 11:10:34 UTC 2023

System load:  0.080078125      Processes:            106
Usage of /:   35.7% of 19.20GB Users logged in:          1
Memory usage: 18%              IPv4 address for ens4: 10.128.0.9
Swap usage:   0%

* Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
  just raised the bar for easy, resilient and secure K8s cluster deployment.

  https://ubuntu.com/engage/secure-kubernetes-at-the-edge

Expanded Security Maintenance for Applications is not enabled.

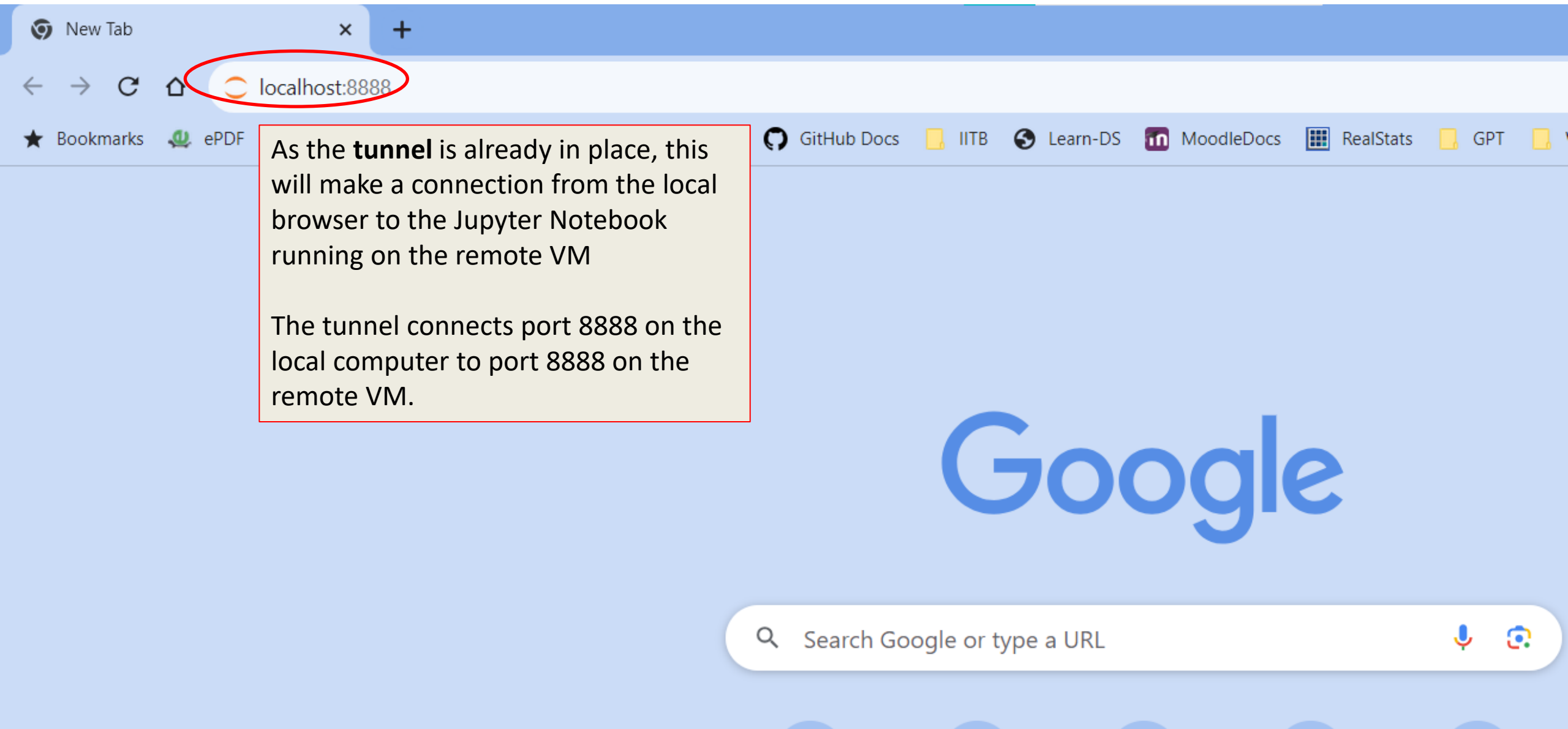
6 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

*** System restart required ***
Last login: Tue Oct 17 10:47:32 2023 from 110.226.183.72
hduser@ds203: ~hduser@ds203:~$
```



Open a browser windows on your local computer ...



Logging into the Jupyter Notebook

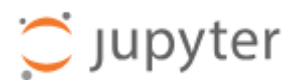
Jupyter Server

localhost:8888/login

Bookmarks ePDF TODO Medium Writesonic GitHub GitHub Docs IITB Learn-DS MoodleDocs RealStats GPT WebTools

jupyter

Password:



File View Settings Help

Files Running

Select items to perform actions on them.

New

Upload



/

☐ Name

☐ data

☐ spark

It will take a minute or so for this page to show up ... be patient!

Congrats! You are logged in!!

Last Modified

File Size

5 days ago

12 days ago

5 days ago

11 B

5 days ago

11 B

 Files  Running

Select items to perform actions on them.

▼ New

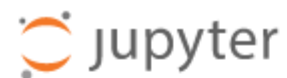
⬆ Upload



<input type="checkbox"/> Name		Last Modified	File Size
<input type="checkbox"/> data		5 days ago	
<input type="checkbox"/> 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 ...










File View Settings Help

Files Running

Rename Delete

/ spark /

☐ Name

- ☐  01-ds203-SPARK-Basics.ipynb
- ☐  02-ds203-RDD-From-File-And-Basic-Operations.ipynb
- ☐  03-ds203-Read-txt-File-into-Dataframe.ipynb
- ☐  04-ds203-Read-CSV-File-into-Dataframe.ipynb
- ☐  05-ds203-Read-CSV-File-into-Dataframe-without-inferSchema.ipynb
- ☐  06-ds203-Read-CSV-and-Run-Linear-Regression.ipynb
- ☐  nsedata.csv

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!

Microsoft Azure

Search resources, services, and docs (G+)

Home > Virtual machines >

ds203-VM
Virtual machine

Search

Connect Start Restart **Stop** Hibernate (preview) Capture Delete Refresh Open in mobile Feedback CLI

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Connect

- Connect
- Bastion

Networking

- Network settings
- Load balancing
- Application security groups
- Network manager

Settings

Essentials

Resource group (move) : [DS203](#)

Status : Running

Location : East US (Zone 1)

Subscription (move) : [Azure for Students](#)

Subscription ID :

Availability zone : 1

Tags (edit) : [Add tags](#)

Operating system : Linux (ubuntu 20.04)

Size : Standard B2s (2 vcpus, 4 GiB memory)

Public IP address : [x.y.z.w](#)

Virtual network/subnet : [DS203-VM1-vnet/default](#)

DNS name : [Not configured](#)

Health state : -

Properties Monitoring Capabilities (7) Recommendations Tutorials

Virtual machine

Computer name	VM3
Operating system	Linux (ubuntu 20.04)
Image publisher	VinayK
Image offer	Course

Networking

Public IP address	x.y.z.w	(Network interf
Public IP address (IPv6)	-	
Private IP address	10.0.0.5	
Private IP address (IPv6)	-	