

Homework-1: System Setup and GitHub Skill Review

In this homework, you are expected to run and connect all the applications that are required for this class. In addition, you will get familiar with GitHub which will constitute the foundation for submitting future homework assignments.

Your homework submission should be uploaded to Gradescope as a PDF. This PDF file should contain the screenshots requested below.

Please upload your PDF file to Gradescope by **September 11, 11:59PM ET**.

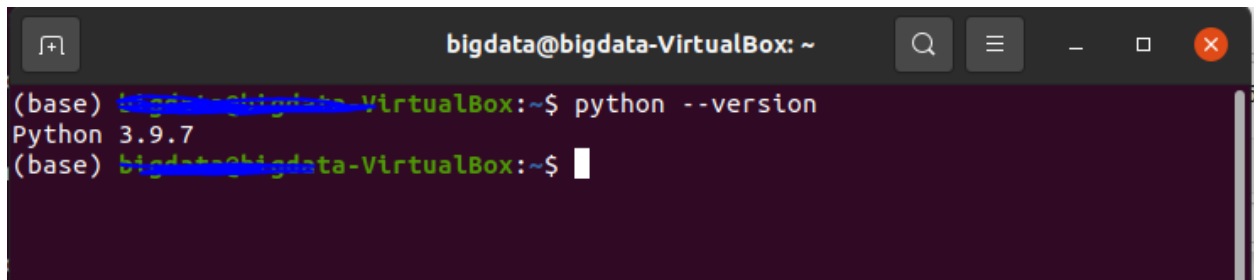
Part-I (70%): Complete Course System Setup

In this part, you are expected to run several applications and make sure they work correctly. You are asked to submit a screenshot for every application to show that it's working properly.

- You may refer to the installation guides (available under the Modules page on Canvas) for high-level instructions on installations.
- **Each screenshot must clearly show that it was taken on your own machine (for example, include your username, computer information, or your name typed in a Notepad file within the screenshot).**
- Sample screenshots are provided below.
- Don't use cloud-hosted [virtual machines](#) for the software installation.

You need to get the following applications to run:

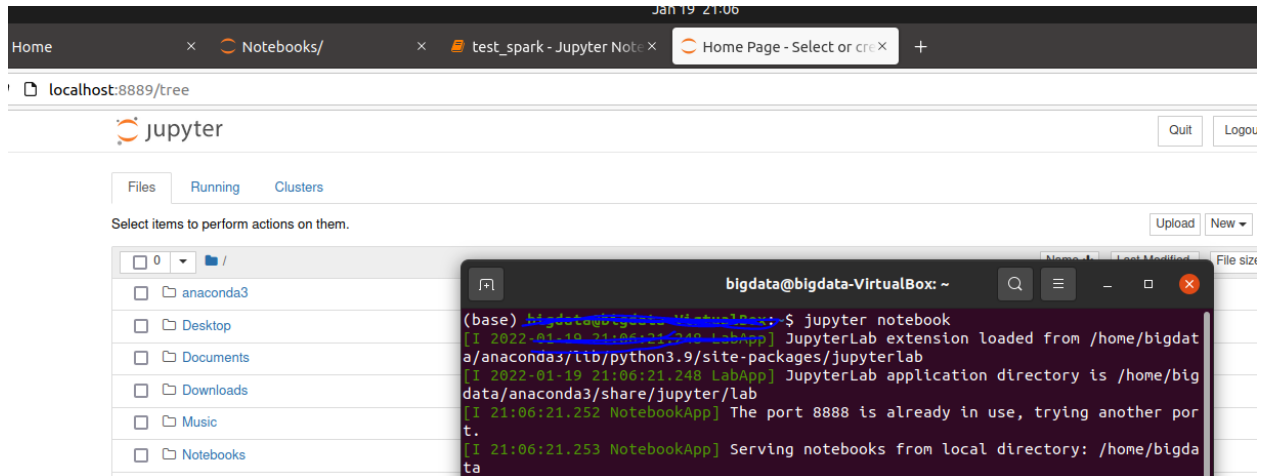
1. [Anaconda Python](#) 3.x. Provide a screenshot showing the version (username was omitted to avoid reusability). If you have Python 3.x already, you don't need to install Anaconda.

A screenshot of a terminal window titled 'bigdata@bigdata-VirtualBox: ~'. The terminal shows a command prompt '(base) bigdata@bigdata-VirtualBox:~\$' followed by the command 'python --version'. The output of the command is 'Python 3.9.7'. The prompt then returns to '(base) bigdata@bigdata-VirtualBox:~\$' with a cursor. The terminal window has standard Linux window controls (minimize, maximize, close) and a search icon in the top right corner.

```
bigdata@bigdata-VirtualBox: ~  
(base) bigdata@bigdata-VirtualBox:~$ python --version  
Python 3.9.7  
(base) bigdata@bigdata-VirtualBox:~$
```

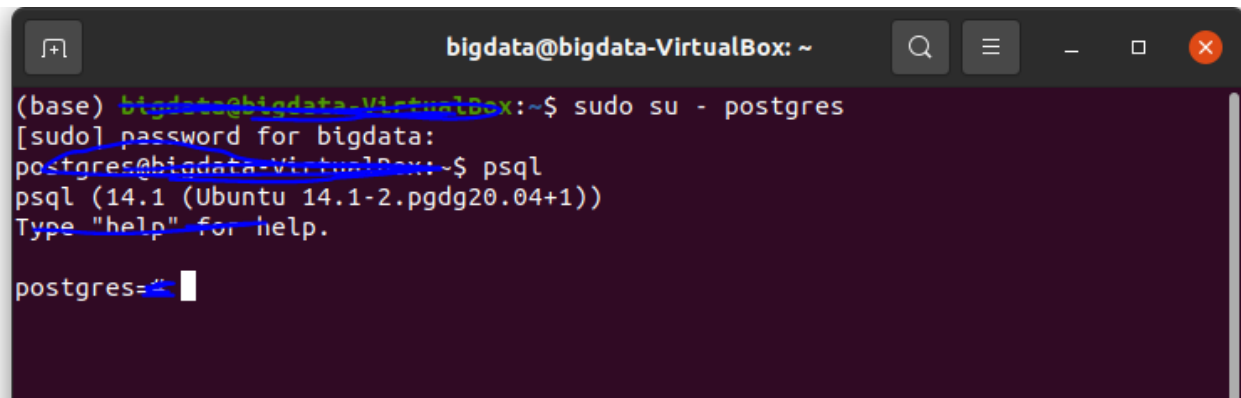
2. [Jupyter](#). Make sure you can open and run jupyter notebooks.

Note: In my screenshot, I launched Jupyter Notebook from the terminal. Your method may be different. Be sure to clearly show and explain how you are launching Jupyter Notebook in your own screenshot. You are welcome to use more than one screenshot if needed.

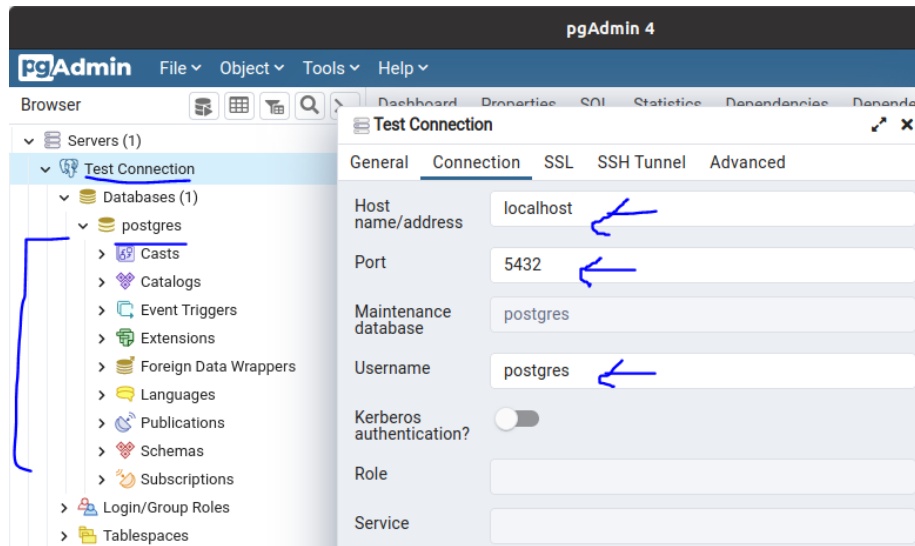


3. [PostgreSQL](#). Submit a screenshot showing how you connect to the database via the terminal or shell.

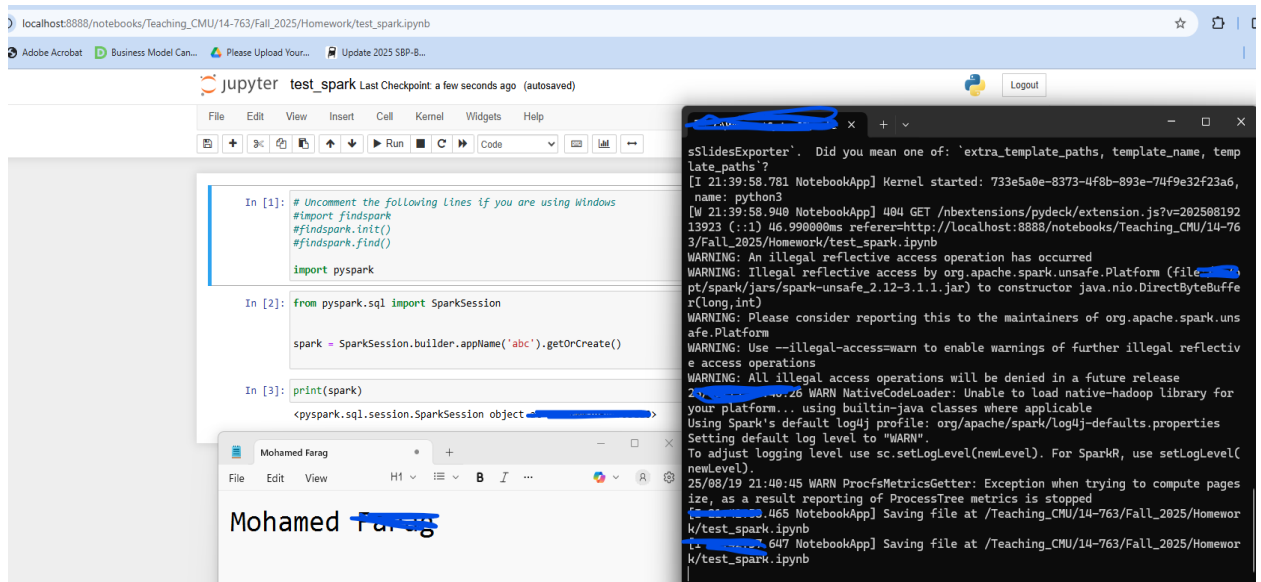
Note: The command you use may differ depending on your operating system. On Windows, you can use SQL Shell (psql) to connect to PostgreSQL



4. PgAdmin4. PgAdmin4 is a Graphical User Interface for PostgreSQL. Post a screenshot showing a successful connection from PgAdmin4 to PostgreSQL.



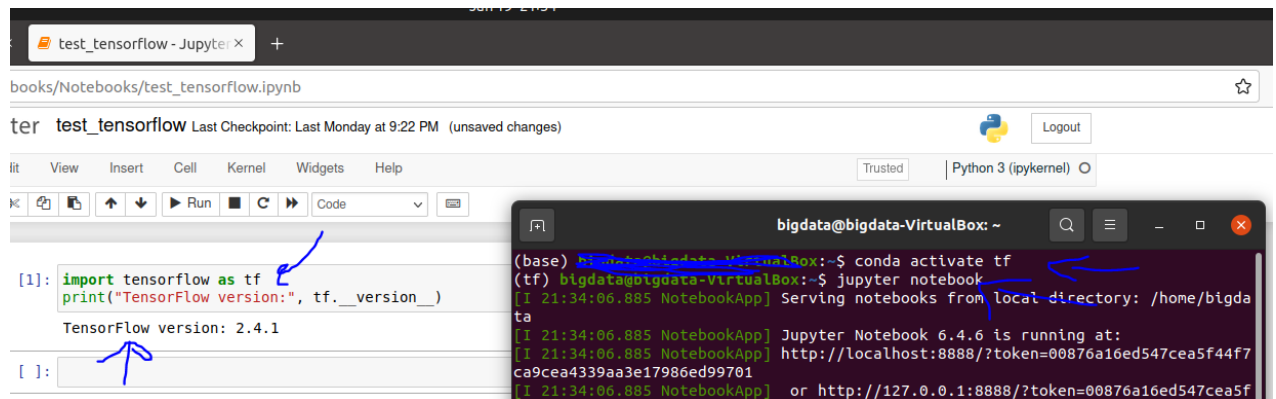
5. [Apache Spark](#). We will use Apache Spark for big data processing and machine learning. Run the test_spark notebook file (posted on the homework page in Canvas) and provide a screenshot of its successful execution. Note: In my screenshot, I launched Spark from the terminal using the command jupyter notebook. Your method may be different. Please make sure to clearly show how you are launching Spark and display the output of the notebook in your screenshot(s). You may include more than one screenshot if needed.



6. [TensorFlow](#). We will use TensorFlow for machine learning modeling. Run the test_tensorflow notebook (posted on the homework page in Canvas) and provide a screenshot showing its successful execution.

Hint:

1. If you can't run TensorFlow in the regular environment, you may consider creating a virtual environment.



The screenshot shows a Jupyter Notebook interface with a code cell containing the following code:

```
[1]: import tensorflow as tf
     print("TensorFlow version:", tf.__version__)
```

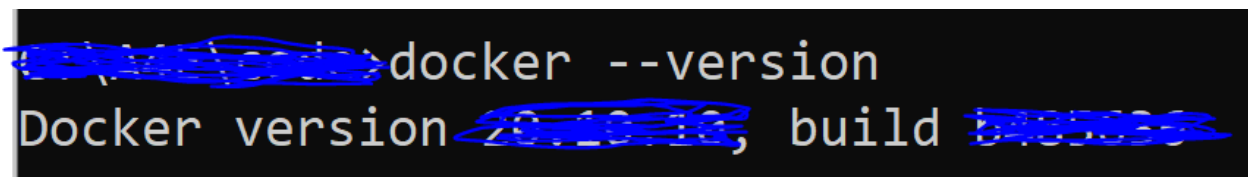
The output of the code cell is:

```
TensorFlow version: 2.4.1
```

Overlaid on the notebook is a terminal window showing the following commands and output:

```
(base) bigdata@bigdata-VirtualBox:~$ conda activate tf
(tf) bigdata@bigdata-VirtualBox:~$ jupyter notebook
[I 21:34:06.885 NotebookApp] Serving notebooks from local directory: /home/bigdata
[I 21:34:06.885 NotebookApp] Jupyter Notebook 6.4.6 is running at:
[I 21:34:06.885 NotebookApp] http://localhost:8888/?token=00876a16ed547cea5f44f7ca9cea4339aa3e17986ed99701
[I 21:34:06.885 NotebookApp] or http://127.0.0.1:8888/?token=00876a16ed547cea5f
```

7. [Pytorch](#). We will use PyTorch for building neural networks. In your Jupyter notebook, import PyTorch without errors. Ensure that both the **pyspark** and **torch** packages can be imported within the same notebook. Submit a screenshot showing the successful execution of these imports.
8. [Docker](#). We will use Docker for NoSQL databases. Submit a screenshot that reflects the version of your docker.



The screenshot shows a terminal window with the following command and output:

```
bigdata@bigdata-VirtualBox:~$ docker --version
Docker version 20.10.18, build 53c35bc
```

Part-II (30%): Get familiar with GitHub

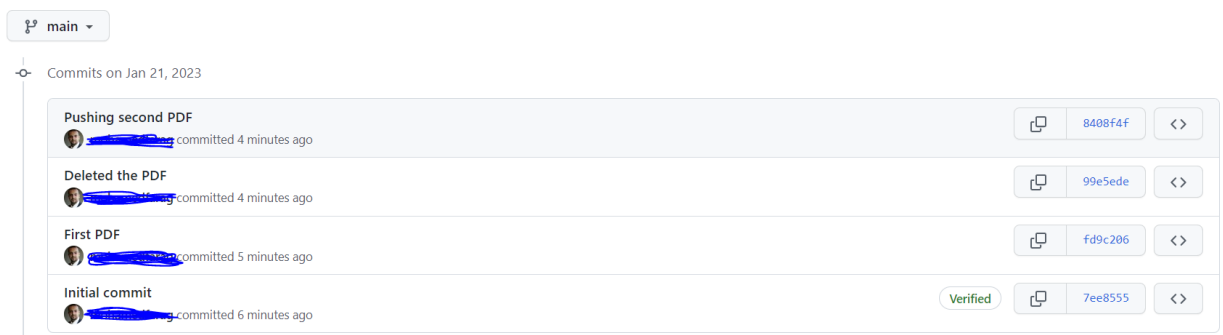
I. Submit screenshots for the completion of ALL the following GitHub labs:

- <https://github.com/skills/introduction-to-github>
- <https://github.com/skills/communicate-using-markdown>

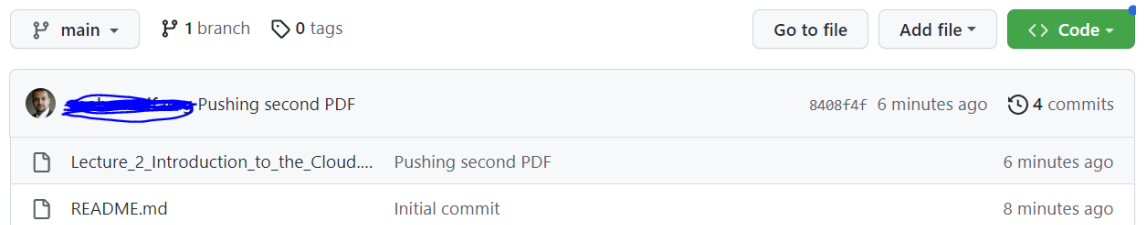
Your screenshots should reflect successful completion of the above labs.

II. Also, complete the following task:

- Create a public GitHub repository named “Test-Toolchains”.
- Include a README file when creating the repository.
- Commit and push the PDF of the first lecture to the repository.
- In a separate commit, delete the first lecture PDF and push the change.
- Commit and push a different PDF file to the repository. Submit the following:
 - a) A screenshot of your GitHub repository history



b) A screenshot of the current view of your repository



c) The URL of your **public** repository

- If you haven't used version control systems before, you can watch the following videos:
 - Version Control Systems Overview: <https://www.youtube.com/embed/ils1aehi3VU>
 - Git and GitHub: <https://www.youtube.com/embed/91ZZv002etU>