

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 a PDF file on Canvas. This PDF file should contain the screenshots that are requested below.

Please upload your PDF file on Canvas before **September 5th, 11:59PM ET**.

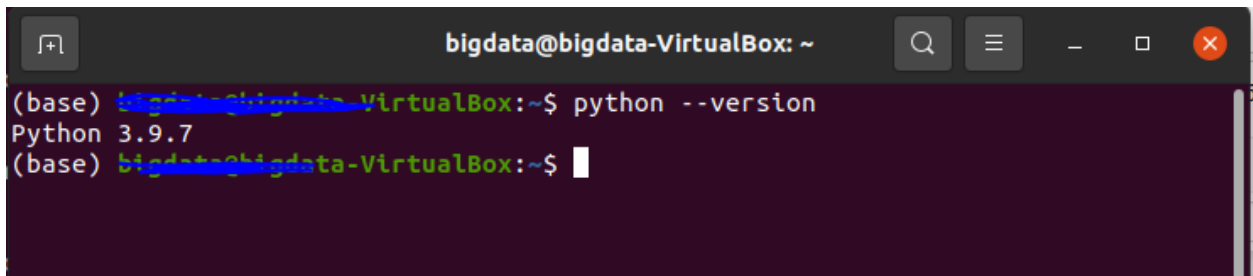
Part-I (70%): Complete Course System Setup

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 use the installation guides to provide general guidance on what you need to install. These guides can be under the Modules page on Canvas.
- All screenshots should have an indication that it was run on your own machine (e.g., your username, your computer information, or your name written in a notepad file that is part of the screenshot taken).
- Sample screenshots are attached below.
- You are required to complete this part of the homework on your local machine.

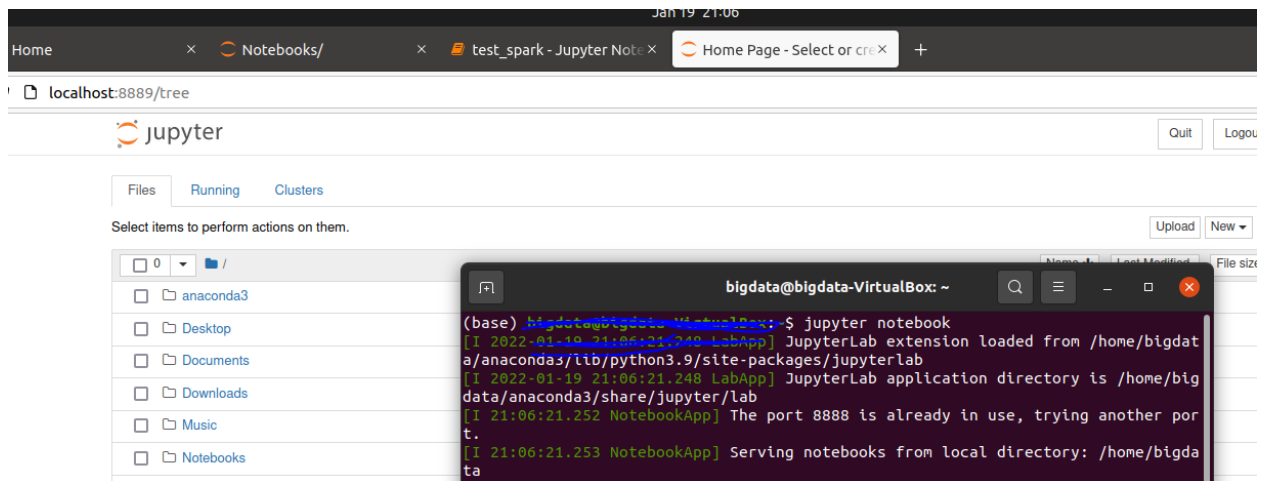
You need to get the following applications to run:

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.

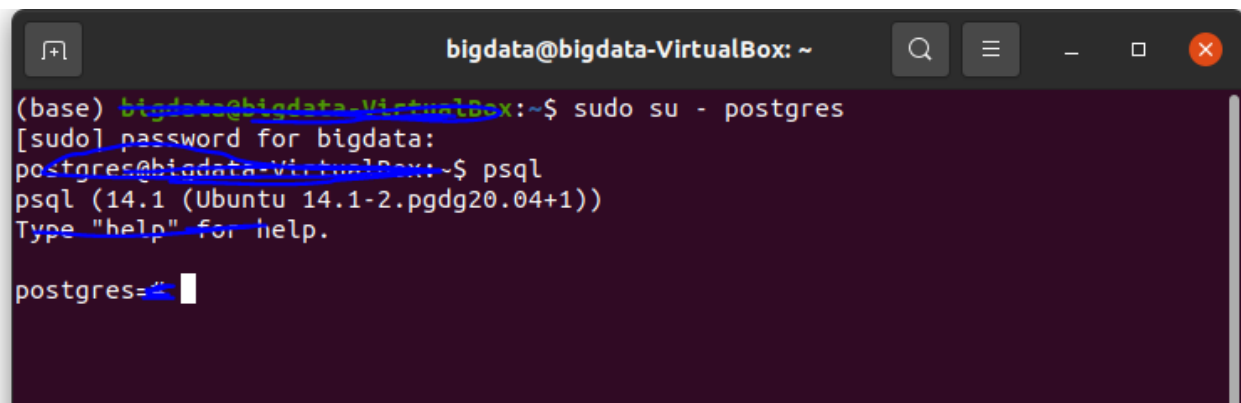
A terminal window with a dark purple background. The title bar at the top reads 'bigdata@bigdata-VirtualBox: ~'. The terminal shows a command prompt '(base) bigdata@bigdata-VirtualBox:~\$' followed by the command 'python --version'. The output is 'Python 3.9.7'. The prompt then returns to '(base) bigdata@bigdata-VirtualBox:~\$' with a cursor. The window has standard Linux window controls (minimize, maximize, close) on the right.

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

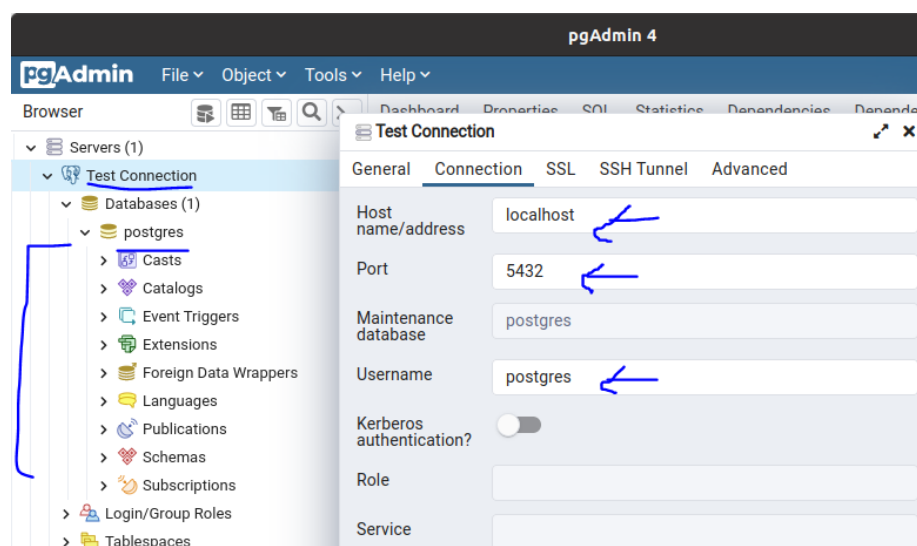
2. Jupyter. Make sure you are able to open and run jupyter notebooks.



3. PostgreSQL DB. Post a screenshot for connecting to the database via the terminal/shell.



4. PgAdmin4. PgAdmin4 is a GUI for PostgreSQL DBMS. Post a screenshot showing a successful connection from PgAdmin4 to PostgreSQL DBMS



5. Apache Spark. We will use Apache Spark for big data processing and machine learning purposes. Run the test_spark notebook file (posted on the homework page on Canvas). Provide a screenshot of successful run of the notebook.

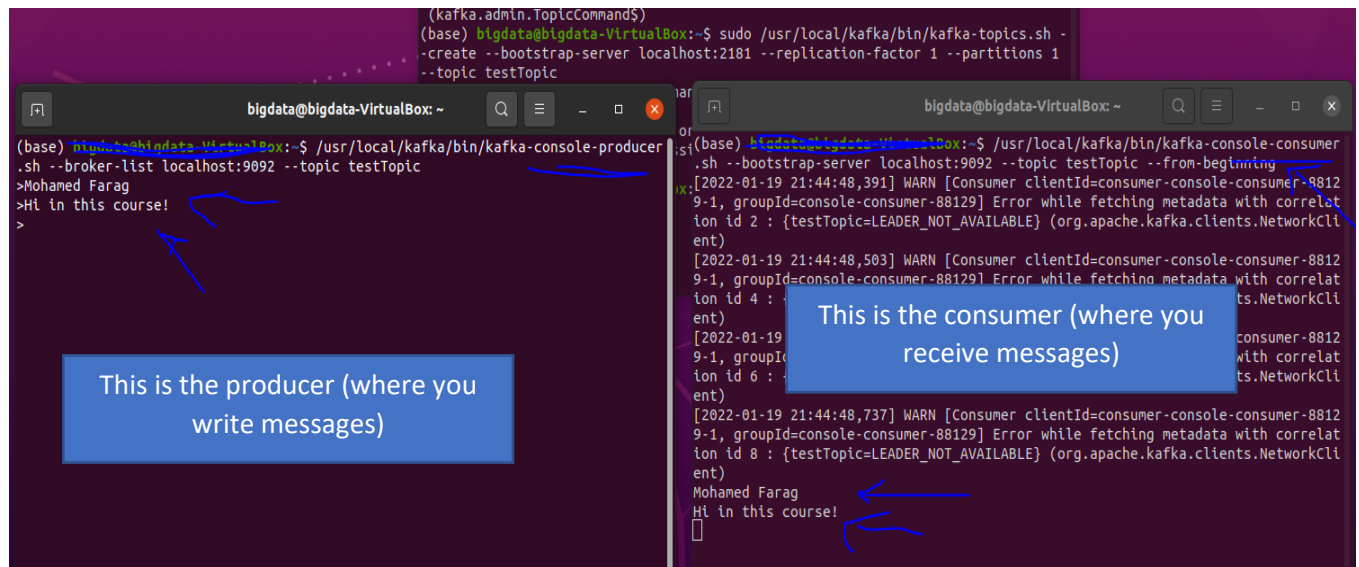
The screenshot shows a Jupyter Notebook interface for 'test_spark'. The notebook has four code cells. The first cell contains 'sc' and the output is 'SparkContext'. The second cell contains 'sc.master' and the output is 'local[*]'. The third cell contains 'from pyspark.sql import SparkSession'. The fourth cell contains 'spark = SparkSession.builder \\.master("local[*]") \\.appName("GenericAppName") \\.config("spark.executor.memory", "3b082").getOrCreate()' and the output is 'To FE1...'. A blue arrow points from the first cell to the terminal window. The terminal window shows the output of the 'sparknotebook' command, indicating that the Jupyter Notebook 6.4.5 is running at 'http://localhost:8888/?token=c203cd86d1f81f2c4cde255c18874824069ac1989243b082'.

6. TensorFlow. We will use TensorFlow for machine learning modeling. Post a screenshot of running test_tensorflow notebook (posted on the homework page on Canvas). Provide a screenshot of successful run of the notebook.

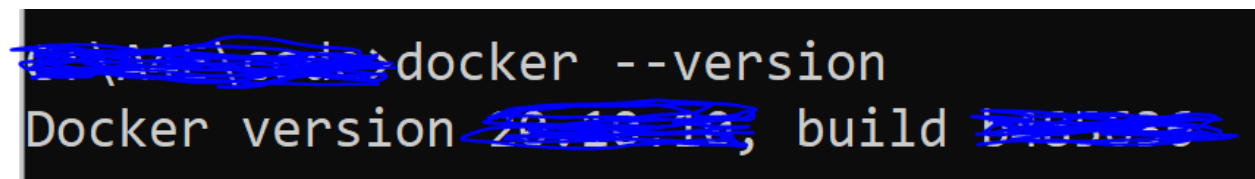
Hint:

1. Make sure your Jupyter notebook is running on the default port, i.e. 8888.
2. If you can't run TensorFlow in the regular environment, you may consider creating a virtual environment.

The screenshot shows a Jupyter Notebook interface for 'test_tensorflow'. The notebook has one code cell containing 'import tensorflow as tf' and 'print("TensorFlow version:", tf.__version__)'. The output is 'TensorFlow version: 2.4.1'. A blue arrow points from the code cell to the terminal window. The terminal window shows the output of the 'conda activate tf' command, indicating that the Jupyter Notebook 6.4.6 is running at 'http://localhost:8888/?token=00876a16ed547cea5f44f7ca9cea4339aa3e17986ed99701'.



7. Pytorch. We will use Pytorch for Neural networks. Import pytorch successfully in your Jupyter notebook without getting an error. Make sure to be able to import **pyspark** and **torch** packages in the same Jupyter notebook.
8. Docker. We will use Docker to try DynamoDB on your local machine. Submit a screenshot that reflects the version of your docker



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

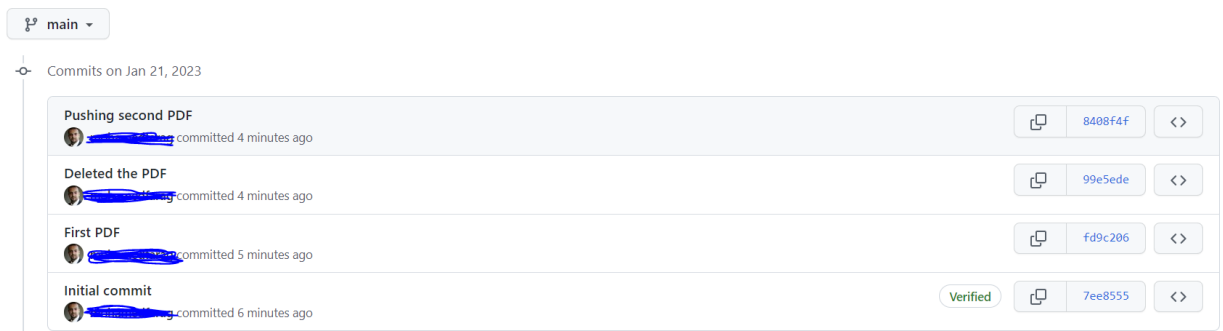
Submit screenshots for the completion of all the following GitHub Skills

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

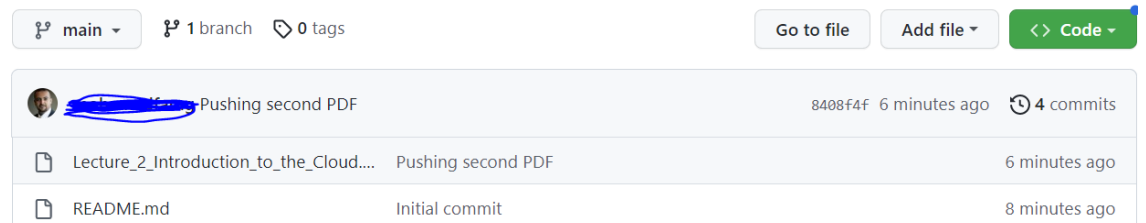
Your screenshot should reflect successful completion of the above labs.

Also, complete the following task:

- Create a Public GitHub Repository with the name “Test-Toolchains”
- Make sure to include ReadMe file in your repository.
- Push the PDF of the first lecture to the repository.
- In a separate commit and push, delete the PDF file of the first lecture
- Push a different PDF to your public repository.
- Submit the following:
 - a) A screenshot of your GitHub repository history



- b) A screenshot of 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>