

# COMP 3610 - Big Data Analytics

The University of The West Indies, St. Augustine

## Environment Setup

January 18, 2022

## 1 System Requirements

For this course you will be executing most of your labs and assignments on your personal computer. We recommend that your computer meet the following minimum specifications in order to install the Anaconda distribution:

- Dual Core CPU
- 6GB RAM
- 30GB Hard Disk/SSD Free Space (SSDs are strongly recommended\*)

In the event of your computer not following these minimum specifications, your running/loading times may be significantly longer. You can opt to use Google Colab, a product from Google Research. Note: We will be using Google Colab for certain labs.

## 2 Google Colab

Colab allows anyone to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. The free tier should be more than fine but you should pay attention to some particulars mentioned in the following link: <https://colab.research.google.com/signup>

### 2.1 Anaconda Installation

<https://medium.com/data-professor/how-to-install-conda-on-google-colab-e7bbf9036f76>

The following Colab notebook contains the code required to install Anaconda and relevant packages.

### 2.2 Anaconda Verification

You can try importing some common packages like pandas, numpy and seaborn.

## 3 Personal Computer

Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.

### 3.1 Anaconda Installation

The steps to install Anaconda vary depending on your operating system. You can click the following link and then select the installation instructions for your operating system: <https://docs.anaconda.com/anaconda/install/>.

### 3.2 Anaconda Verification

In order to confirm that Anaconda is working properly, you can make use of either:

- Anaconda Navigator (GUI)
- conda (Terminal)

*Using Terminal:* Enter **conda list**. If Anaconda is installed and working, this will display a list of installed packages and their versions.

An alternative is to try the following. Enter the command **python**. This command runs the Python shell. If Anaconda is installed and working, the version information it displays when it starts up will include “Anaconda”. To exit the Python shell, enter the command **quit()**.

*Using Anaconda Navigator:* Open Anaconda Navigator. If Anaconda is installed properly, Anaconda Navigator will open.

## 4 Conclusion

- Finally, you can execute the command: **jupyter notebook** via the command line or open Jupyter Notebook from the Anaconda Navigator menu.
- Why not use pip? <https://www.anaconda.com/blog/understanding-conda-and-pip>
- Something we will be exploring later on: <https://rapids.ai/>