Big Data Engineering in the Cloud

March 4th-6th 2019
Old Thorns Hotel

Instructors will be **Julie Weeds** and **Simon Wibberley**

If you have any pre-course questions please contact Julie: juliewe@sussex.ac.uk

Pre-course setup and exercises

In order to make the course as successful as possible, please follow the following pre-course instructions. This will mean that you can come along to the course well-prepared for the three days.

The pre-course setup and exercises consists of four main activities:

1) Setting up a Github Education account and AWS Educate account to get free credit in the cloud.

2) Downloading and installing the Virtual Machine that will be used during the course.

3) Doing a simple exercise to learn about Jupyter, Python and Lambdas

4) Doing a simple exercise to start a cloud server and test it out.

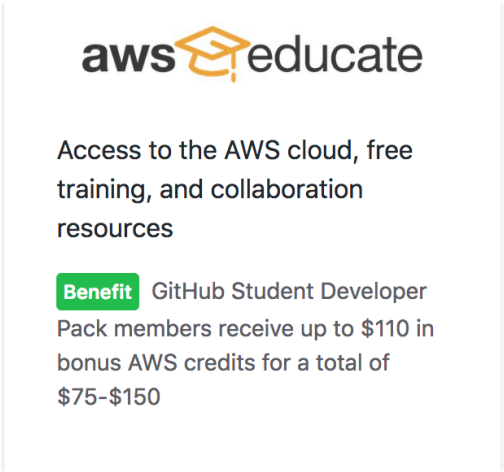
These exercises will ensure that everyone starting the course has successfully installed the virtual machine and has an Amazon AWS account ready to use. In addition, we should all understand how to use lambdas, which are an important part of the Apache Spark approach that will be used in the course.

1. **Github Student Pack and AWS Educate**

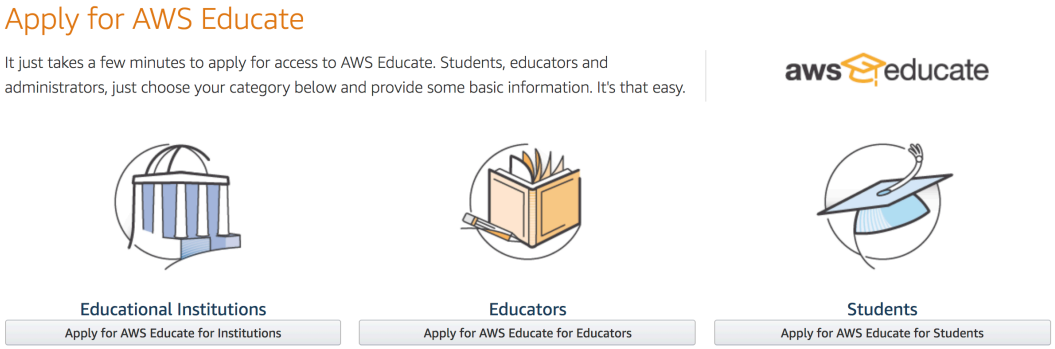
   Go to https://education.github.com/

GitHub Student Developer Pack
The best developer tools, free for students

Follow the instructions to sign up. Once you have signed up you should see a page of offers including the AWS Educate one.



Click on the **unique link** and you should see something like:

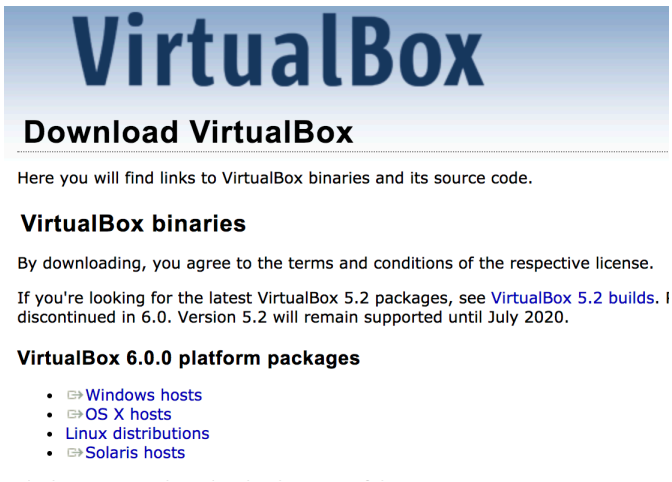


Click on Student and follow the instructions. Eventually (usually within 1 week) you should be approved and receive free credit towards AWS services.

2. Downloading and installing the Virtual Machine

The Virtual Machine is designed to run under a free system called VirtualBox.

First download and install VirtualBox from
https://www.virtualbox.org/wiki/Downloads



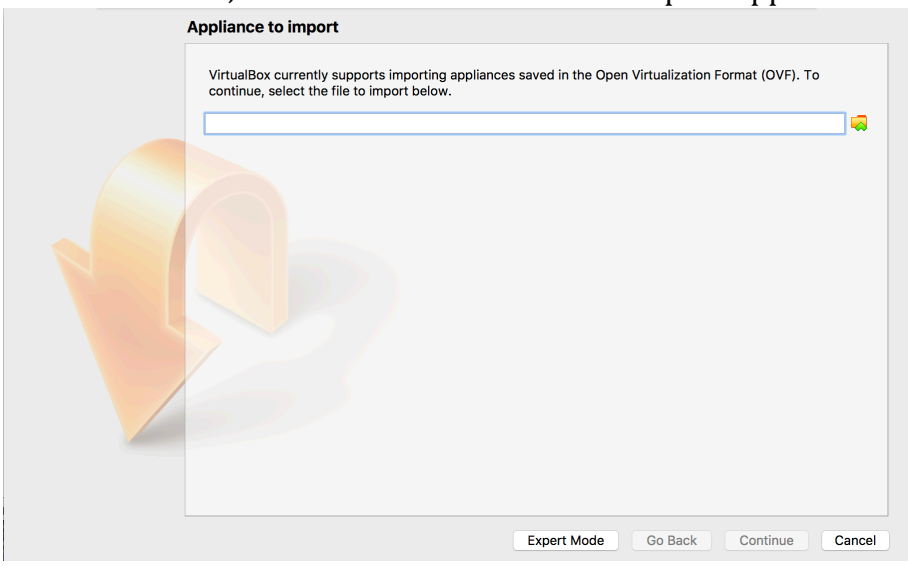
Choose the platform package that is suitable for the operating system on your laptop. You may need to update your security settings ('allow software extensions signed by Oracle') to get a successful installation.

Start VirtualBox and it should look like this:

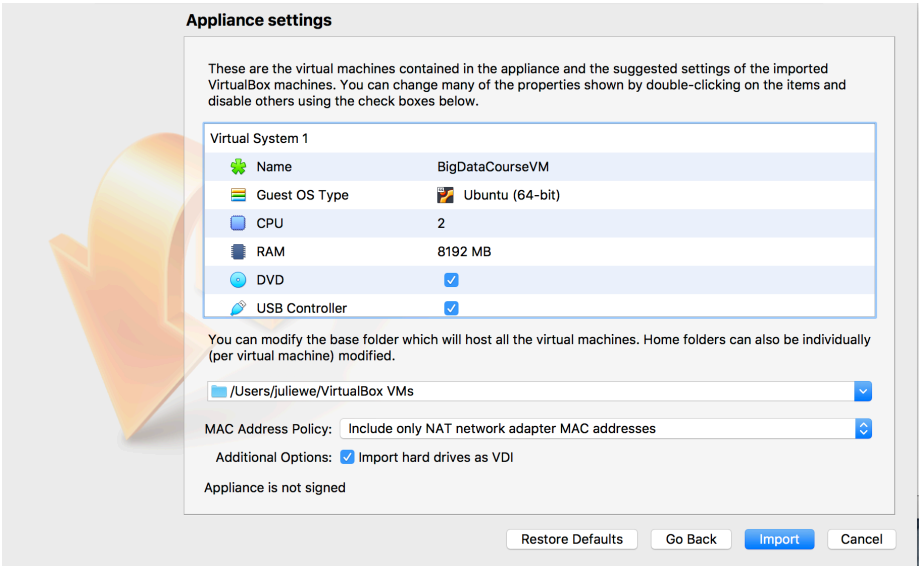
Next, download the Virtual Machine from:
https://s3-eu-west-1.amazonaws.com/bdec/BigDataCourseVM.ova

Its a large file download so it may take some time depending on your internet connection.

From VirtualBox, choose the menu item File->Import Appliance.



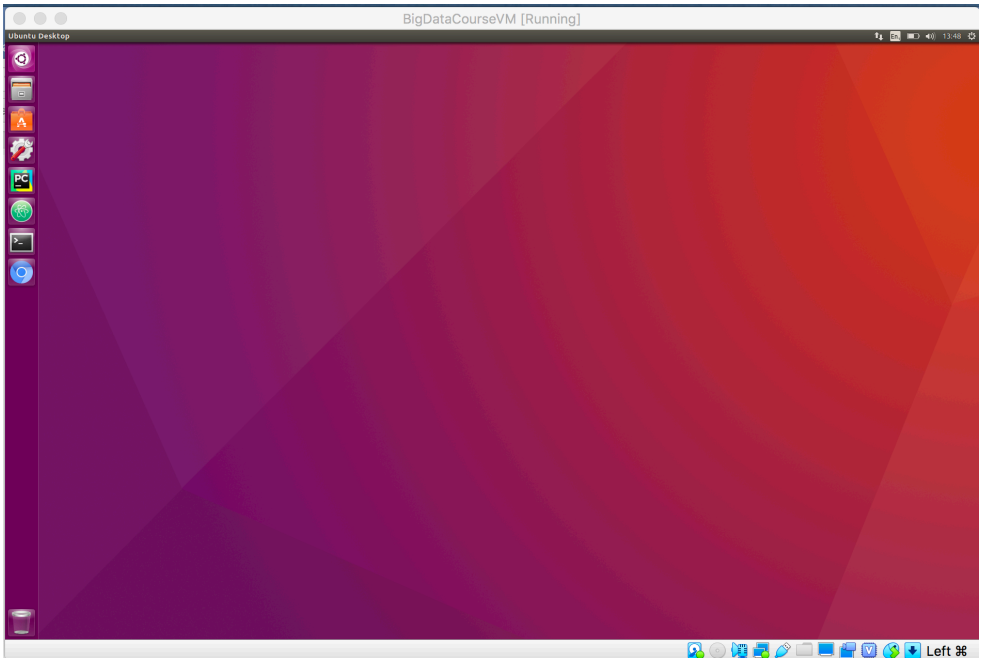
Click on the small file icon, and select the download `BigDataCourseVM.ova` file.
The screen should now look like:



Click **Import**. It will take a few minutes.

Your VirtualBox window should now look like:

Double click on the BigDataCourseVM logo and your VM should start. You should see this:



The VM should start without requiring a login/password, but sometimes you may need them. The username and password are **big/big**.

3. Python Lambdas exercise

The exercise is provided in a separate file: 00-pre-python-lambdas.pdf

4. Amazon exercise

Please use the VM to run this exercise (it has the AWS CLI installed and also we will need the files that are created in your VM during the week).

The exercise is provided in a separate file: pre-amazon-ec2-getstarted.pdf