**Big Data Engineering in the Cloud**

2nd – 4th March 2020

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](mailto:juliewe@sussex.ac.uk)

Pre-course setup and exercises

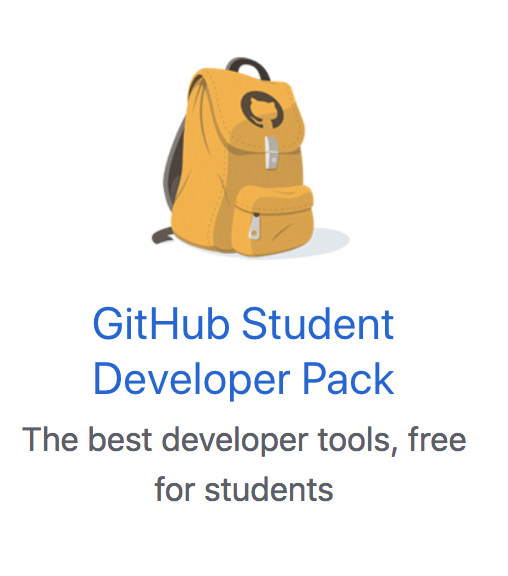
In order to make the course as successful as possible, please follow the following pre-course instructions. This will mean that you can come along to the course well-prepared for the three days.

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 Anaconda
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.

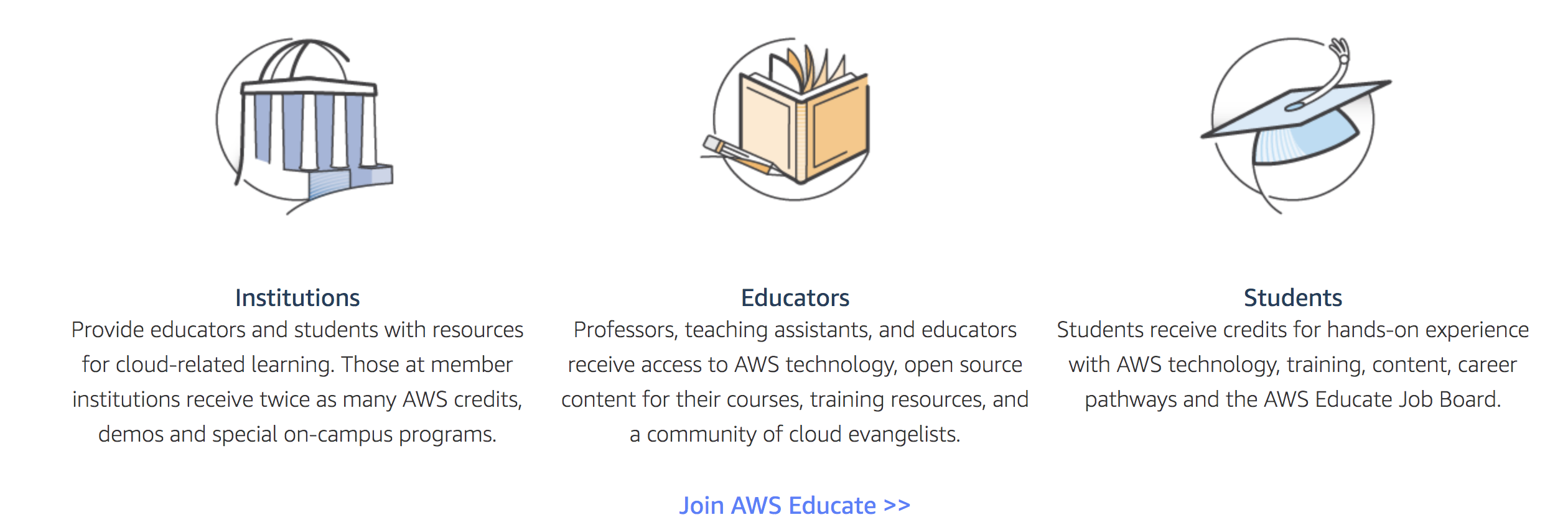
These exercises will ensure that everyone starting the course has successfully installed the virtual machine and has an Amazon AWS account ready to use. In addition, we should all understand how to use lambdas, which are an important part of the Apache Spark approach that will be used in the course.

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

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

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



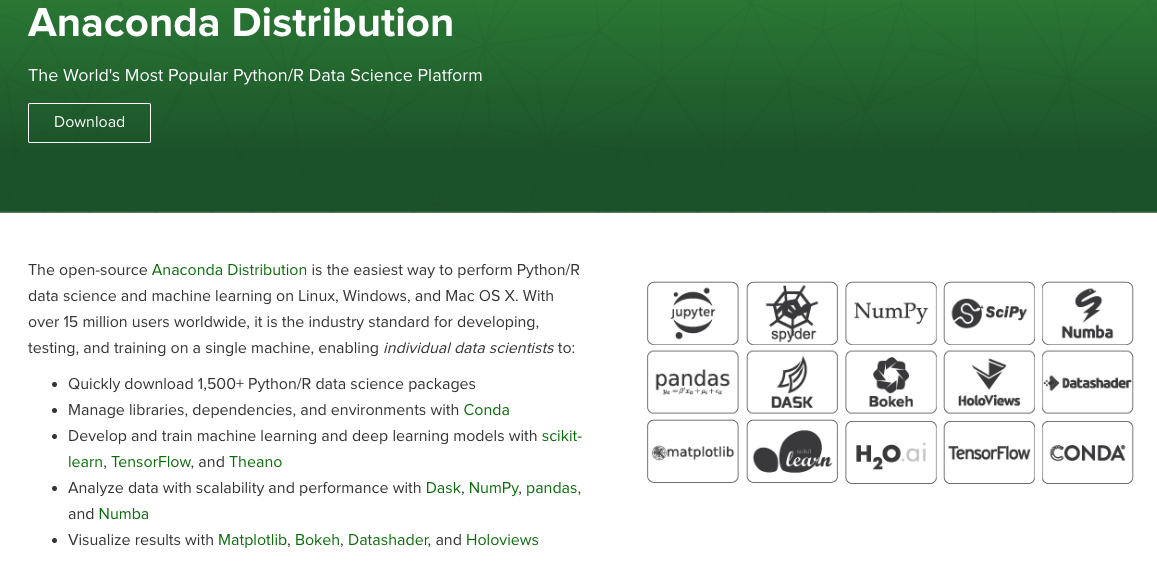
Click on the **unique link** and you should see something like:  
  
Click on Join AWS Educate, select Student and follow the instructions. Eventually (usually within 1 week) you should be approved and receive free credit towards AWS services.

**Note** – there appears to be some difference between the services that can be accessed by students on a “Starter Account” (which does not require a credit card to set up) and students on a standard student account. I would recommend you go for a standard student account (which gives you $100 free credit before your credit card is charged) rather than a starter account – then shut your account down if you want to avoid any risk of charges.

1. **Downloading and installing Anaconda**In the exercises, we will be using Python3.7 including various libraries which are all available in the Anaconda distribution.

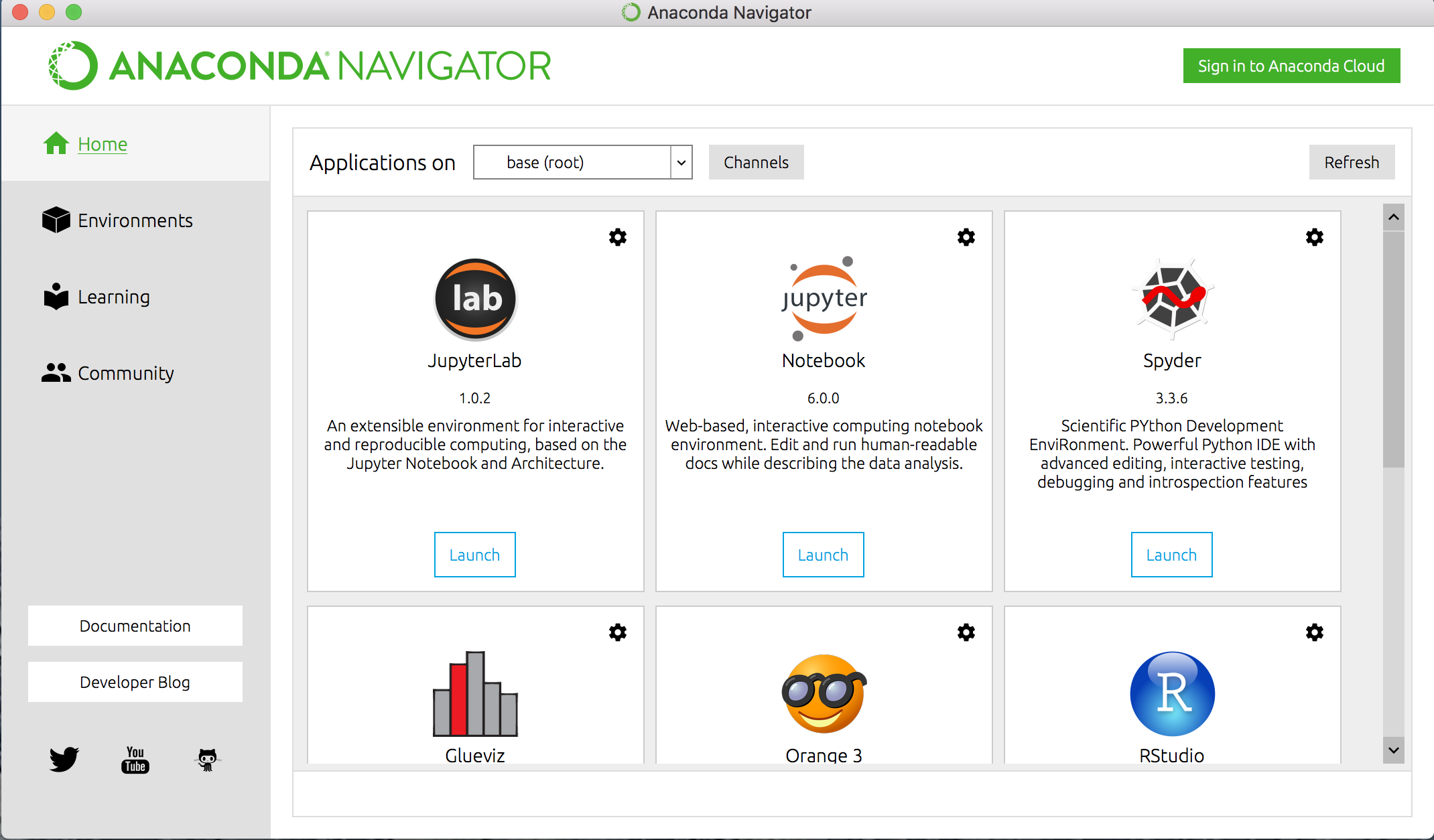
First download and install Anaconda from

<https://www.anaconda.com/distribution/>



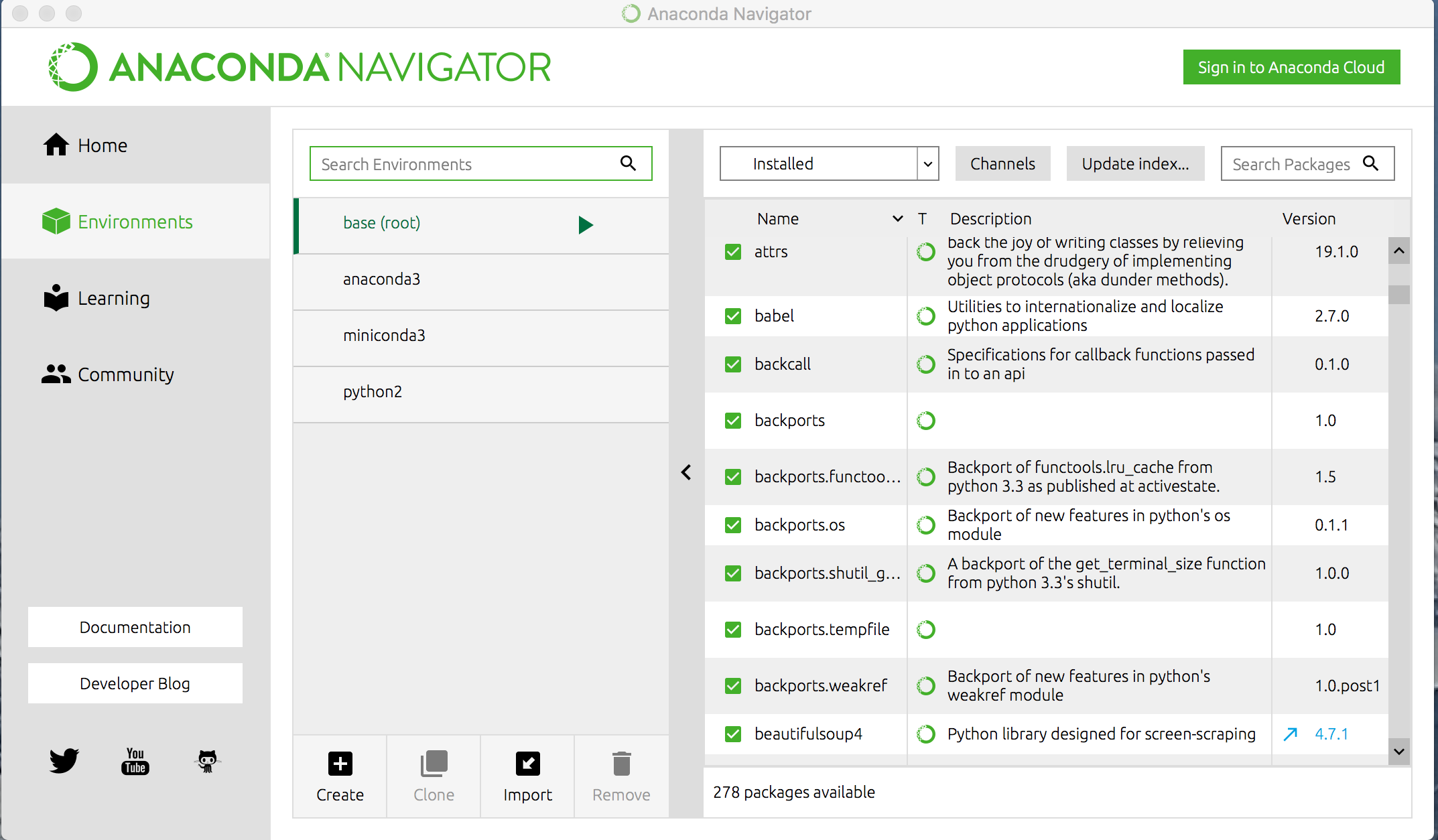
Choose the platform package that is suitable for the operating system on your laptop. You may need to update your security settings (‘allow software extensions signed by Oracle’) to get a successful installation.

Once you have installed Anaconda, you should be able to launch Anaconda Navigator from the Start or Applications menu:

****

From here, you can launch jupyter notebook, which is the programming environment we will be using during the course.

Also, if you click on Environments, you can see what libraries are installed and what other libraries are available to be installed.



1. **Python Lambdas exercise**  
   The exercise is provided in a separate file: 00-pre-python-lambdas.pdf
2. **Amazon exercise**

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