Setting up SPARK development Environment

In order to get started with the development of Apache Spark and to follow along with this course, please follow the following steps:

Go to Oracle VM Virtual Box website (https://www.virtualbox.org/wiki/ Downloads) and download and install it in your system. Its a virtualization software that allows you to run virtual machines in your system. We will be using a pre-configured Virtual Machine from Cloudera and for that, we need to have this software installed.

Once done, go to Cloudera Quickstart VM website (https://www.cloudera.com/ downloads/quickstart_vms/5-13.html), select Virtual Box from the "Select Platform" drop down menu and download the VM. It's going to be approximately 3-4GB in size so it may take sometime to download.

Once downloaded, open Oracle Virtual VM Virtual Box, click File Menu and click Import Appliance. Navigate to the folder where you download the Cloudera Quickstart VM and select the configuration file there.

Once done, the rest of the process is quite straightforward. Just click Next in the displayed wizards to finish the configuration. You don't necessarily need to change the defaults here.

Once installed, the VM will appear in Virtual Box. Select that and click on start. It will start the virtual machine and you will be ready to do development in Apache Spark.