<u>Preparing the VM for the project</u>

1. We are working in the virtual machine Cloudera-Quickstart-VM-5.12.0-0.
2. We are using Python 2.7.13.
3. Download and install Kafka using parcels in the Cloudera Manager
4. Download and install Anaconda Navigator 4.3.1 via cloudera@quickstart terminal:
   $ conda install -c conda-forge
5. Create in Anaconda Navigator new environment my_root:
   $ conda create -n my_root --clone=/opt/cloudera/parcels/Anaconda
6. Download and Install jupyter-notebook via cloudera@quickstart terminal:
   $ python -m pip install jupyter
7. Download and install kafka library 1.3.5 for python in the new my_root environment:
   $ conda update kafka-python
8. Download, install and update in Anaconda Navigator (**my_root environment**) all the necessary libraries: pandas 0.22.0, numpy 1.13.1, matlibplot 2.0.2, json 2.6.0, seaborn 0.8, scikit-learn 0.19.0
9. Start jupyter-notebook in **my_root environment** via **terminal@bash-4.1$**
   $ jupyter-notebook
10. Start the Cloudera Manager via cloudera@quickstart terminal:
    $ sudo /home/cloudera/cloudera-manager --pause --express --force
11. Start all the services in the Cloudera Manager.
12. Uploading all the necessary datasets (flights.csv, airlines_num.csv, airports.csv) in the jupyter-notebook in **my_root environment**.