**Apache Cassandra + Python**

Assuming Apache Cassandra and Python installed. Please refer to the installation document if not installed.

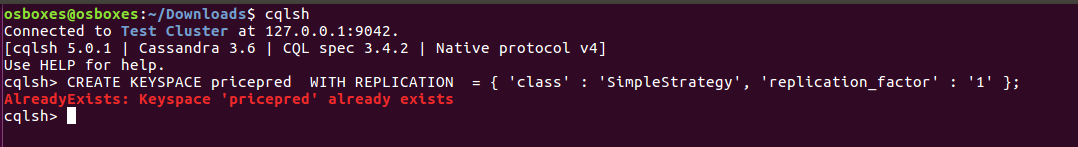
Apache Cassandra is an open-source distributed database system.

We will try to import the sample dataset from CSV to Cassandra.

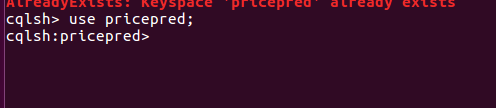
**CASSANDRA Queries**

CREATE KEYSPACE pricepred WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication\_factor' : '1' };

Note: I have already created the keyspace, that’s the error.



use pricepred;



CREATE TABLE prices (

rank2014 int,

city varchar,

state varchar,

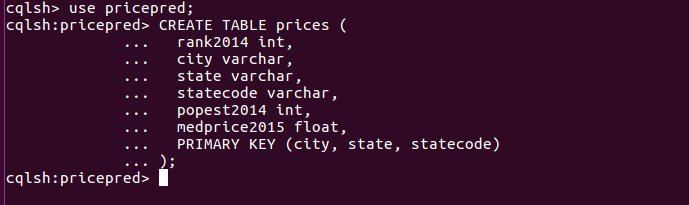
statecode varchar,

popest2014 int,

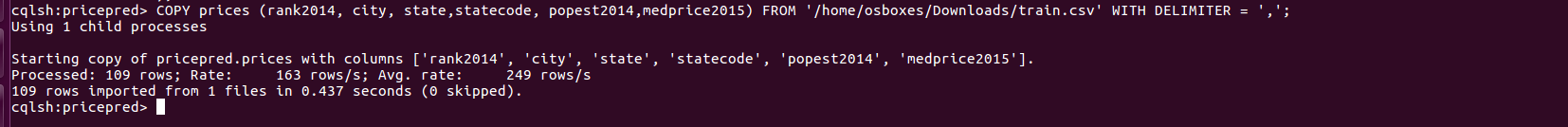
medprice2015 float,

PRIMARY KEY (city, state, statecode)

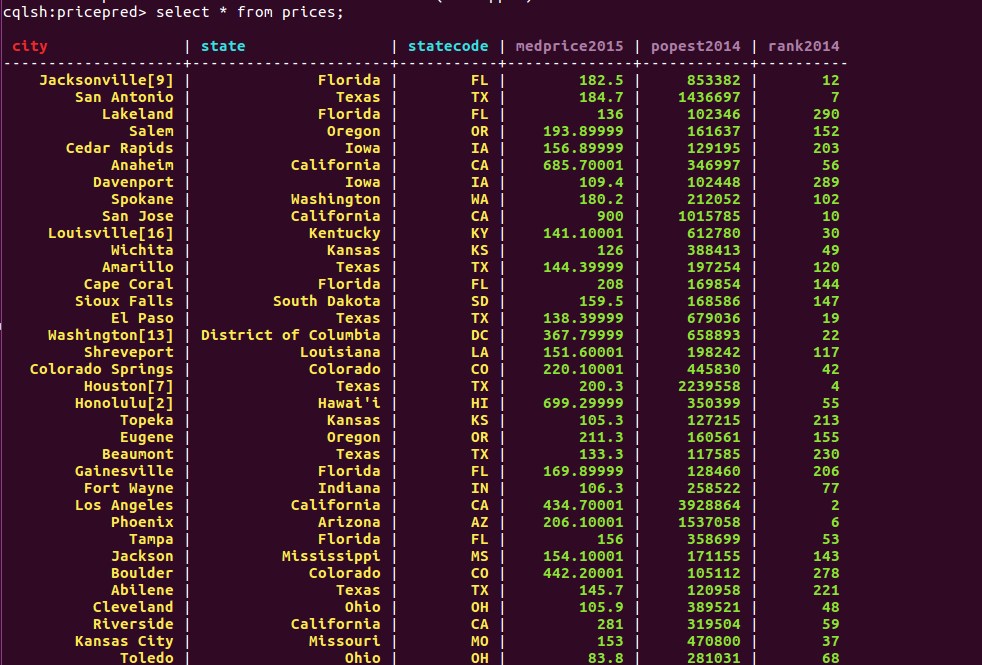
);



COPY prices (rank2014, city, state,statecode, popest2014,medprice2015) FROM '/home/osboxes/Downloads/train.csv' WITH DELIMITER = ',';



select \* from prices;



Interfacing with Cassandra using Python is made possible using the Python client driver that you can pip install:

# pip install cassandra-driver

We will use ipython notebook, it’s very good for Exploratory Data Analysis, troubleshooting and sharing the notebooks.

In Python you can simply execute Cassandra Query Language (CQL) queries and receive the results in an iterator. CQL is syntactically very similar to SQL; however it misses join, group by and some other operations.

Please refer to the **Prediction of House Prices - Cassandra – Python.ipynb** for further retrieving of data from Cassandra, converting to dataframe and running machine-learning algorithm (Regression for this example dataset).

To access and work on any ipython notebook, we need to execute below command.

#ipython notebook <notebookname>

Or

#jupyter notebook <notebookname>