Big Data Management:

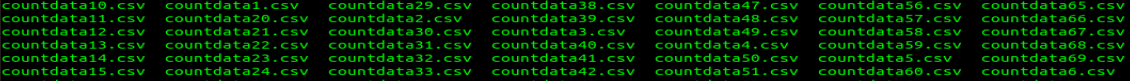
# (Assignment 3)

# **Description:**

This assignment involves development of real-time views on traffic data using pyspark streaming and writing the real-time views to Cassandra tables.

**Live traffic Emulation using Python:**

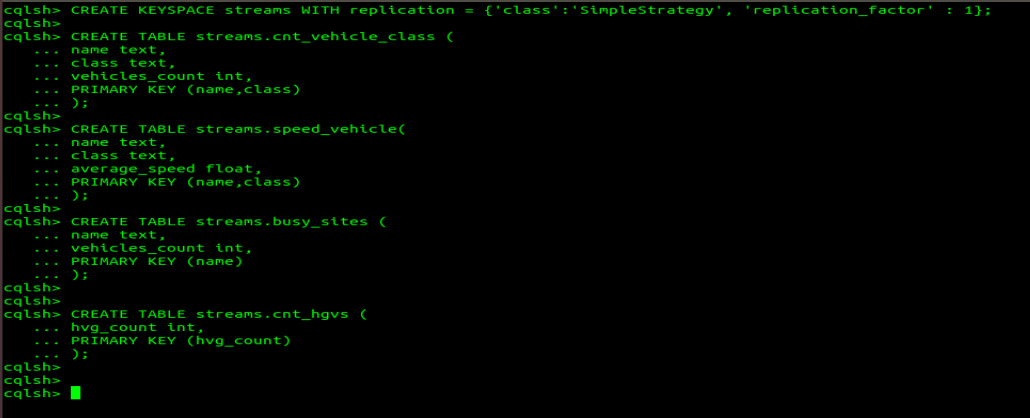
From the per-vehicle-records file containing traffic data live traffic streaming are generated every 5 seconds with names like countdata1, countdata2…etc.



# **Data tables creation in Cassandra:**

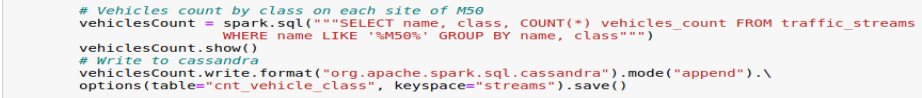
Queries in the file DDL.txt are used to create key-space and required tables for this assignment.

PFB snapshot of creation of data tables in Cassandra to store real-time views.

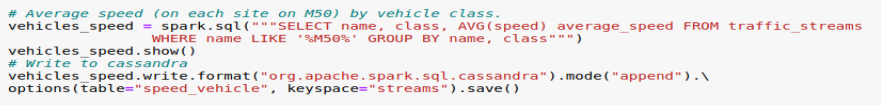


# **Computation of Real-time views using pyspark streaming and writing to Cassandra tables:**

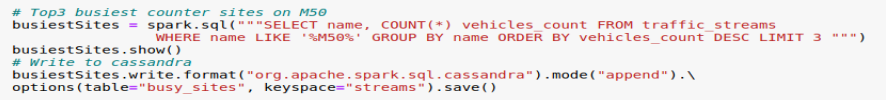
* Count of vehicles on M50 by class.



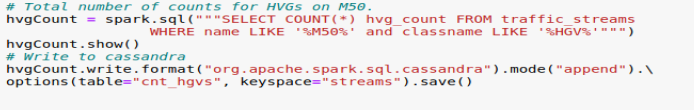
* Average speed of vehicles on M50 by vehicle class.



* Top 3 busiest counter sites on M50.

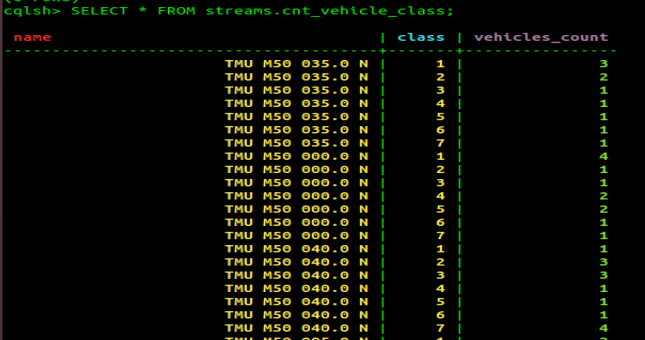


* Total number of counts for HGVs on M50.

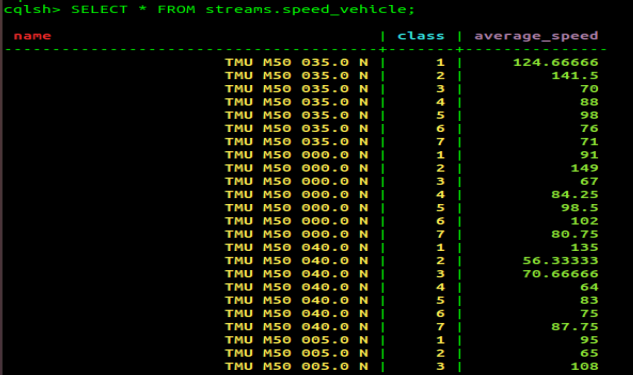


**Real-time views in Cassandra tables:**

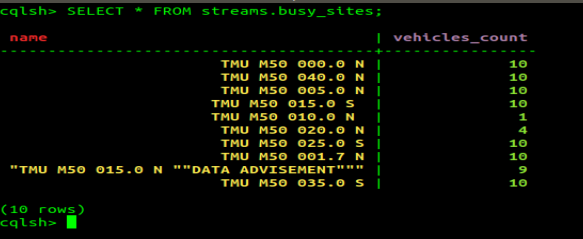
* Count of vehicles on M50 by class.



* Average speed of vehicles on M50 by vehicle class.



* Top 3 busiest counter sites on M50.



* Total number of counts for HGVs on M50.

