Objective – To create partitions on table and execute aggregate queries.

Tried doing partition two ways-

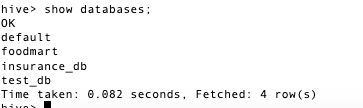
**1.Partition based on gender**

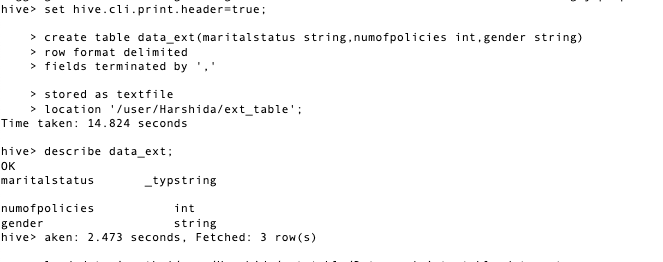
**2.Partition based on education**

**1.Partiton based on gender**

Moved Data1.csv from /home/hdfs to /user/Harshida/ext\_table and the permissions of the file were changed using command **hdfs dfs -chmod 777 /user/Harshida/ext\_table/Data.csv**

Logged into hive- and created the test\_db and created the external\_table

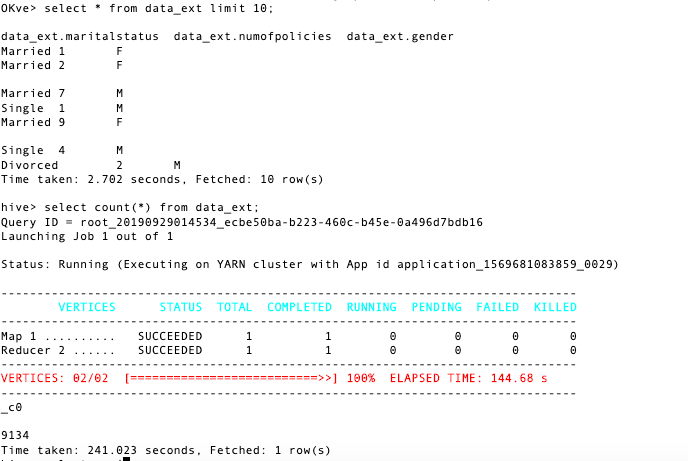


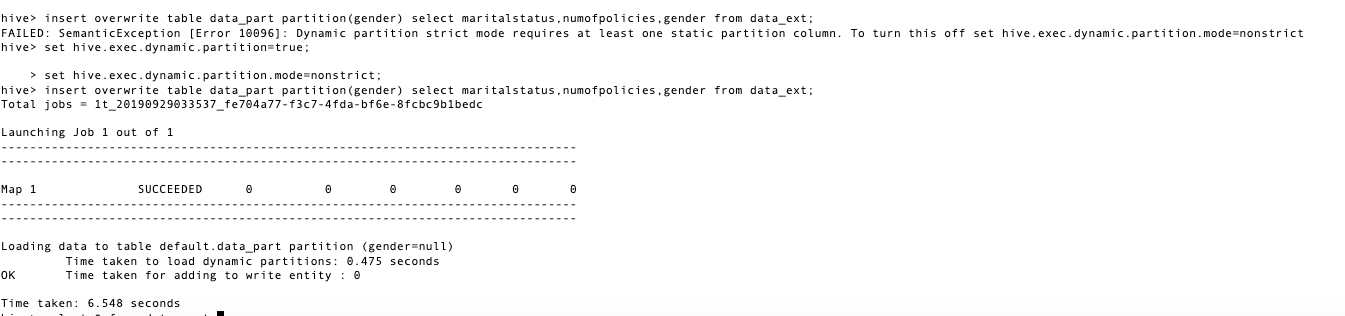


2.Load data from HDFS path into data\_ext table by using command

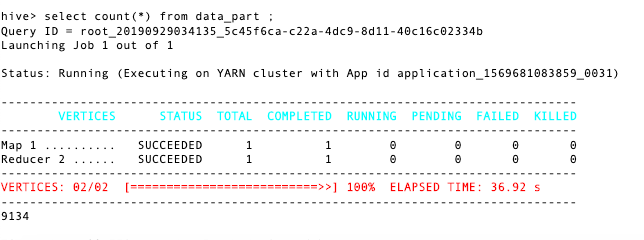
**Load data inpath ‘/user/Harshida/ext\_table/Data.csv’ into table data\_ext ;**

**Once loaded the table looks like below**

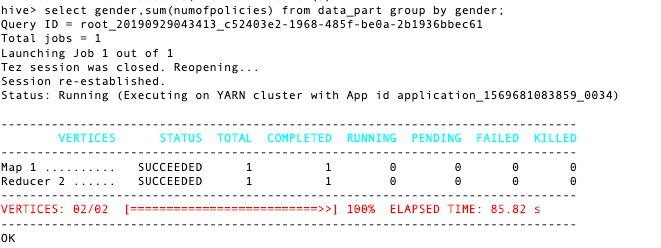


3.Creating another table data\_part which is partitioned by gender-**Dynamic partitioning**

This query takes data from data\_ext and creates a new table data\_part with the given partition column.

4.Applied count (\*) aggregate query on the table created which gave successful result.

Select query on the partition-

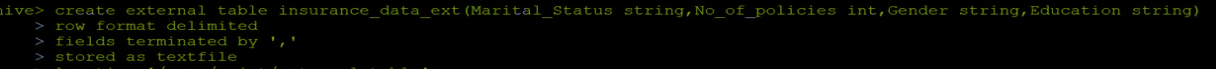


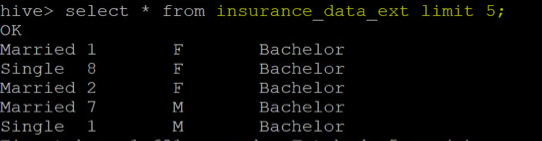


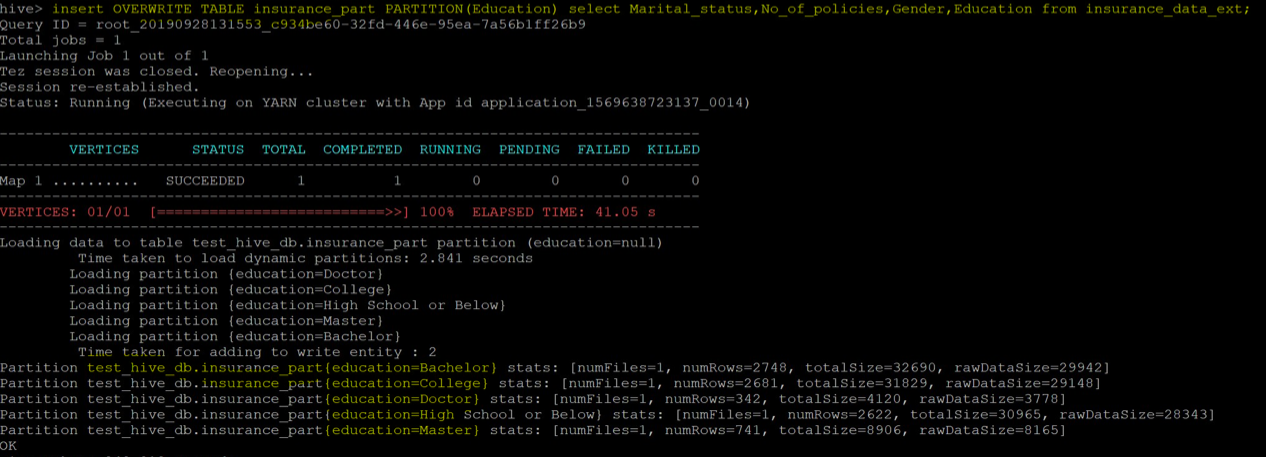
**2.Partition based on education-**

**The Insurance\_data1.csv file is moved to HDFS and all the permissions are given.**

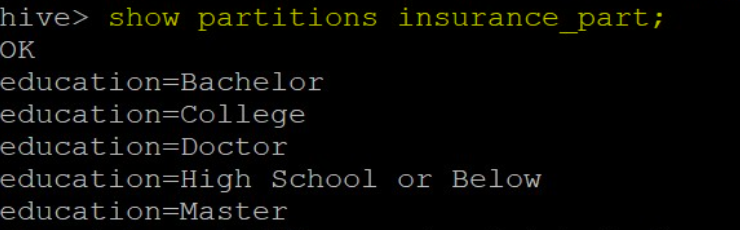
**Logged into hive to create external table-**

**Using test\_db;**

****

** Creating new table Insurance\_part by creating partitioning on education column.**

**Partitions created-**

****

Select query on the partitioned table-

