19162121031

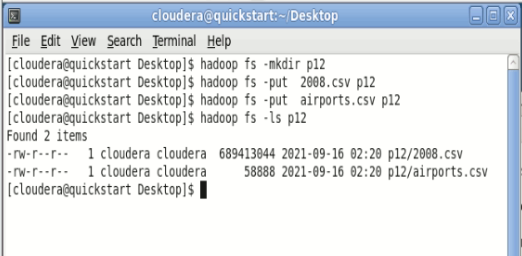
SMIT R PATEL

BDA

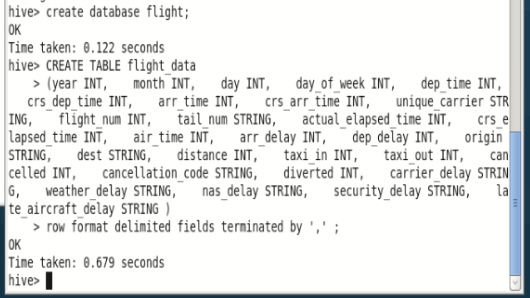
SEM 5

PRACTICAL 12

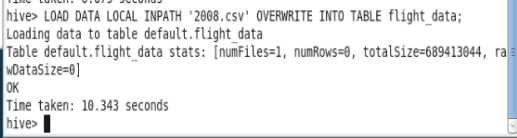
Here, first of all we make directory p12 in Hadoop and put file 2008.csv and airports.csv in p12 directory from Cloudera Desktop :



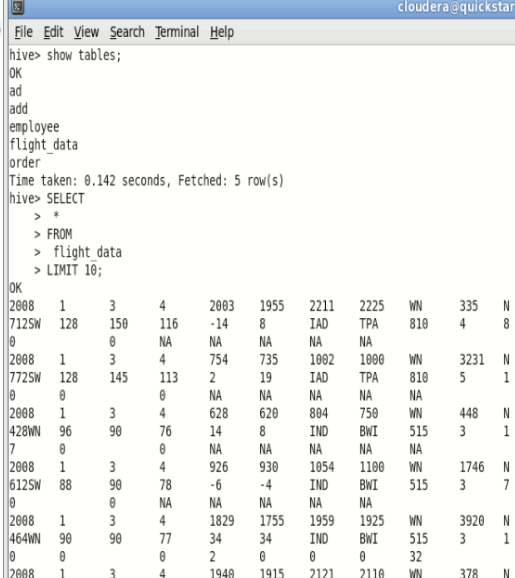
1 Create hive table, flight\_data:



2 Load the data into the table:



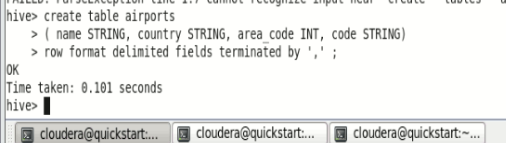
3 Ensure the table got created and loaded fine:



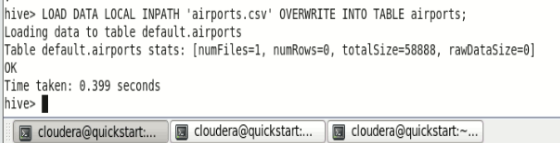
4 Query the table. Find average arrival delay for all flights departing SFO in January:



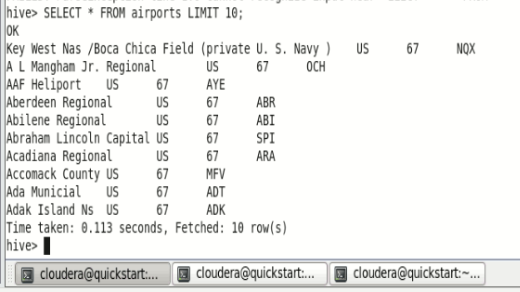
5 On hive shell: create the airports table



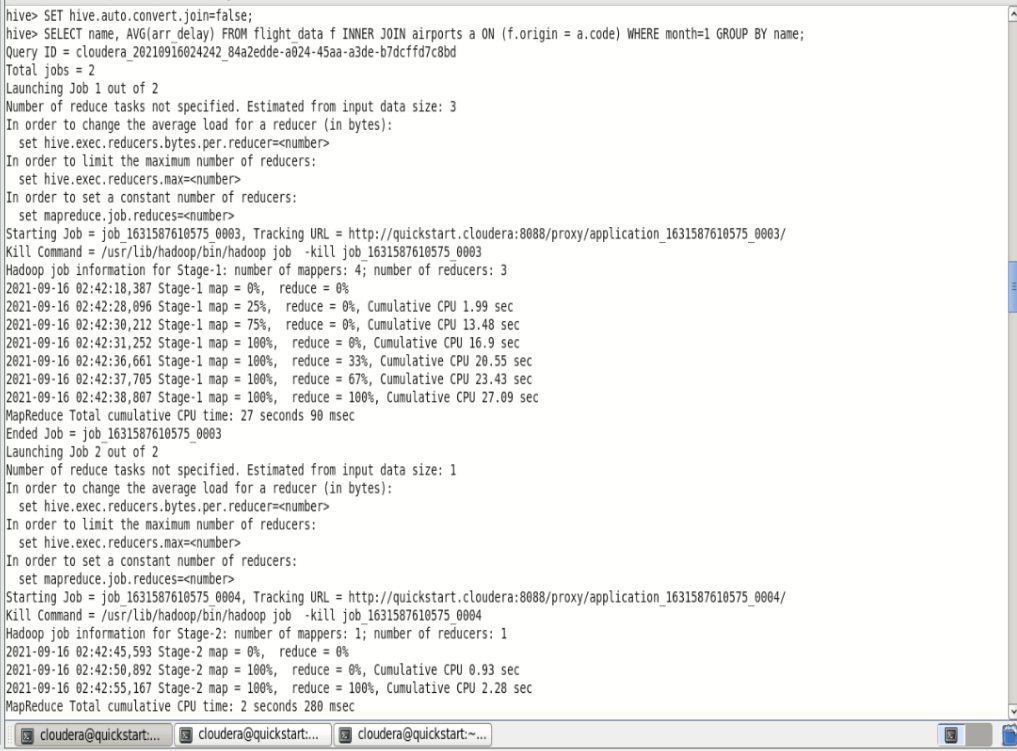
6 Load data into airports table:

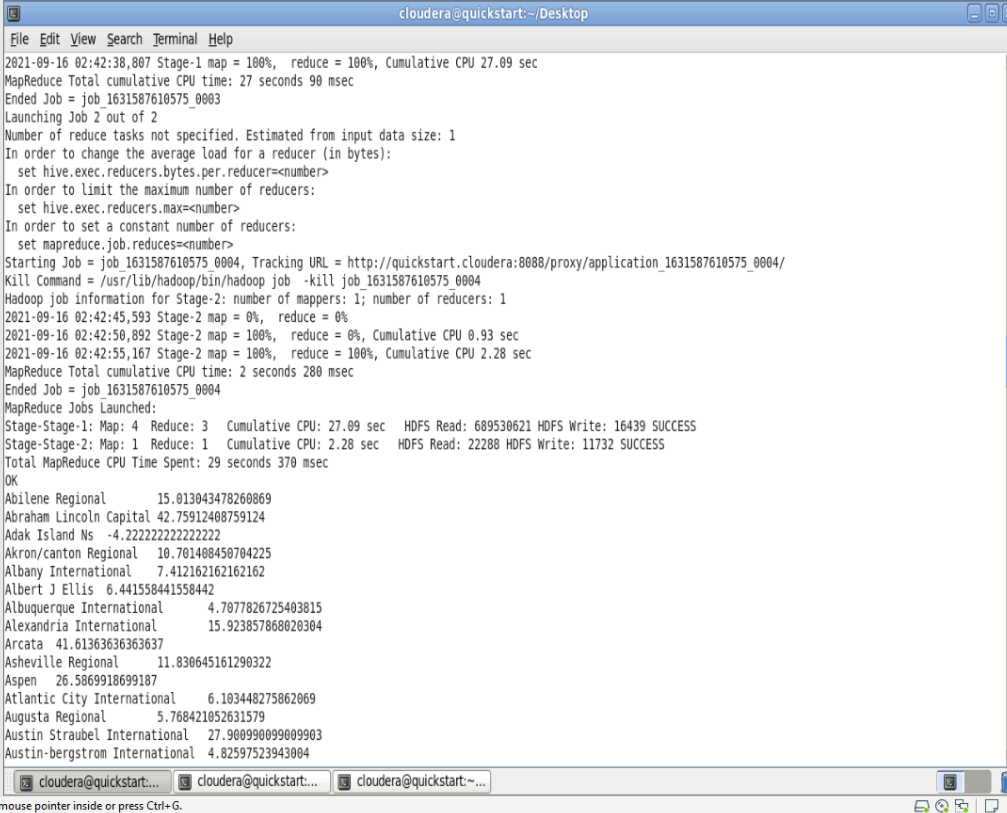


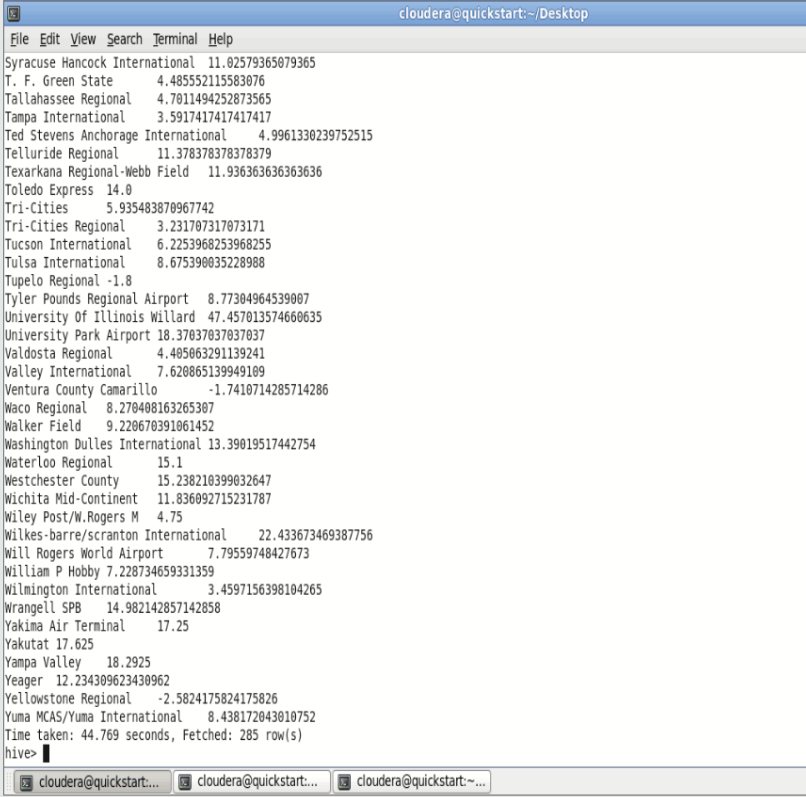
7 On hive shell, list some rows from the airports table:



8 On hive shell: run a join query to find the average delay in January 2008 for each airport and to print out the airport's name:







Extra



