1. Download vechile sales data -> https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv

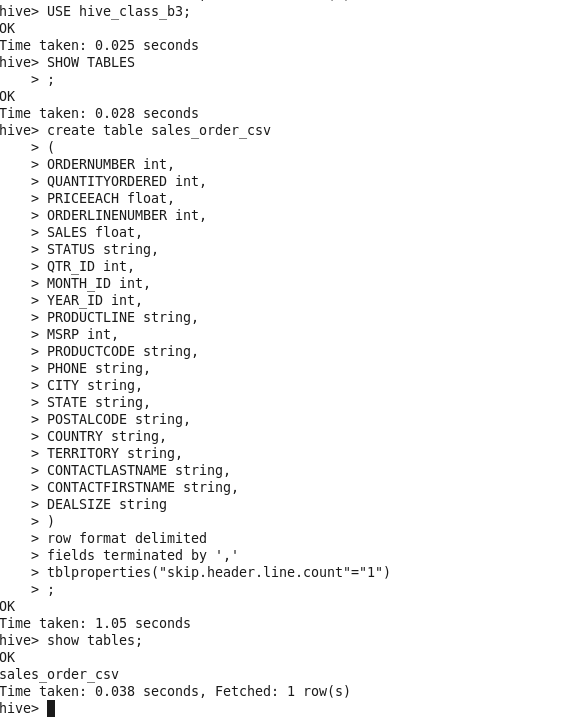
Answer – Data is downloaded to local machine and copied to “/home/cloudera/Prashant” in Cloudera Local File System

2. Store raw data into hdfs location

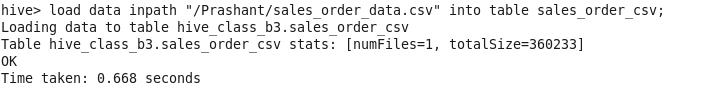
Answer – Data is loaded from Cloudera Local File System to HDFS location ‘

Hadoop fs -copyFromLocal /home/cloudera/Prashant/sales\_order\_data.csv /Prashant/

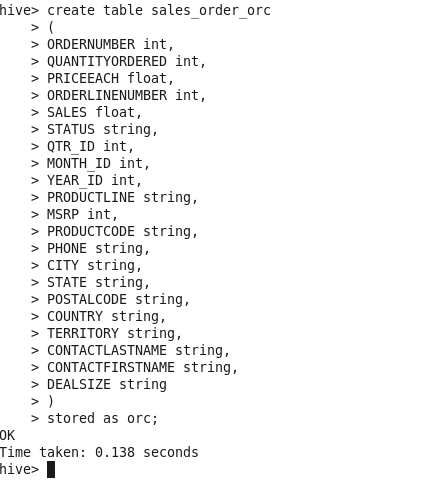
3. Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table



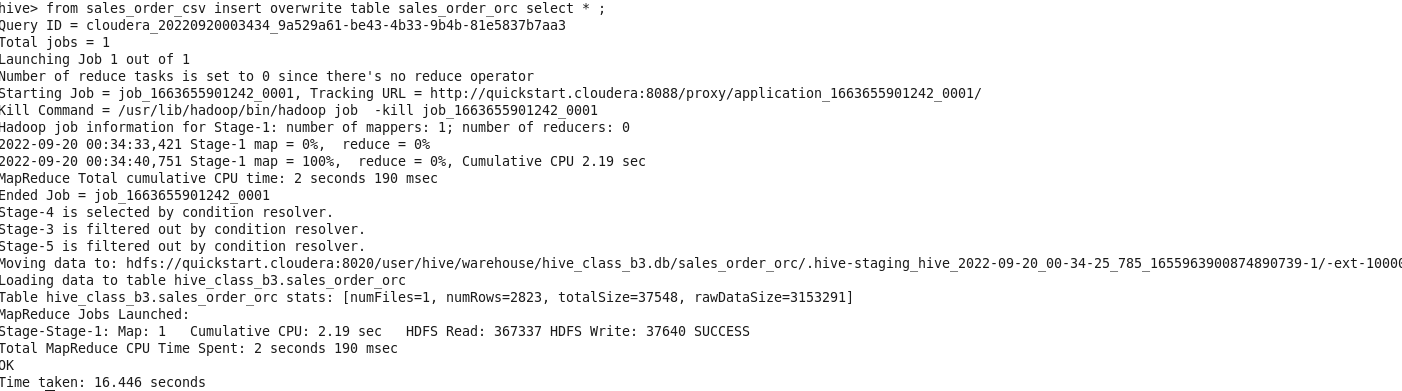
4. Load data from hdfs path into "sales\_order\_csv"



5. Create an internal hive table which will store data in ORC format "sales\_order\_orc"

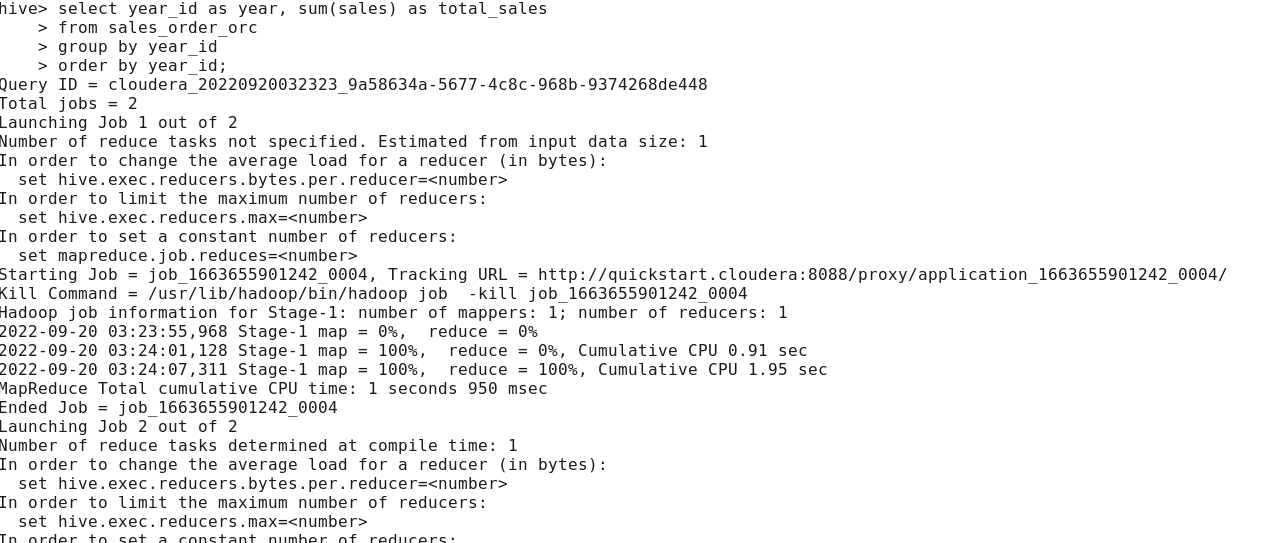


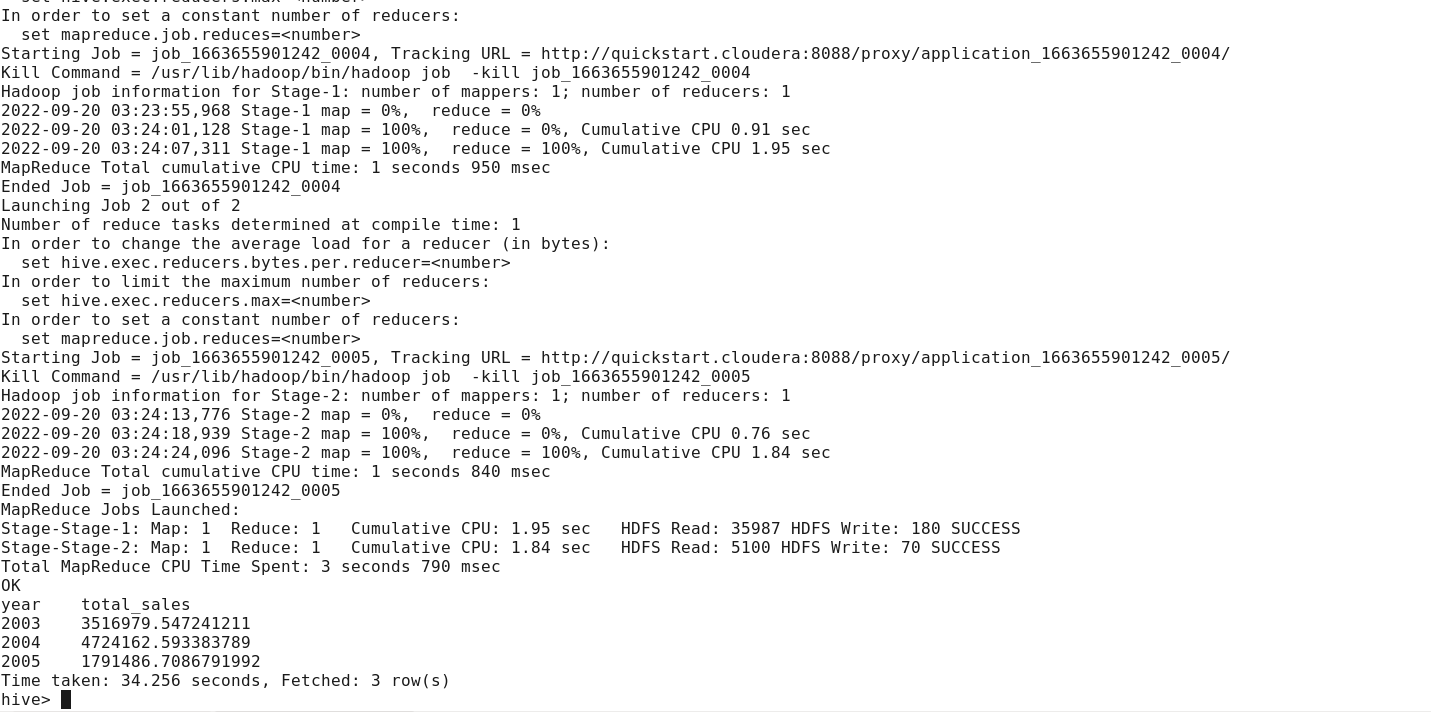
6. Load data from "sales\_order\_csv" into "sales\_order\_orc"



Perform below menioned queries on "sales\_order\_orc" table :

1. Calculatye total sales per year



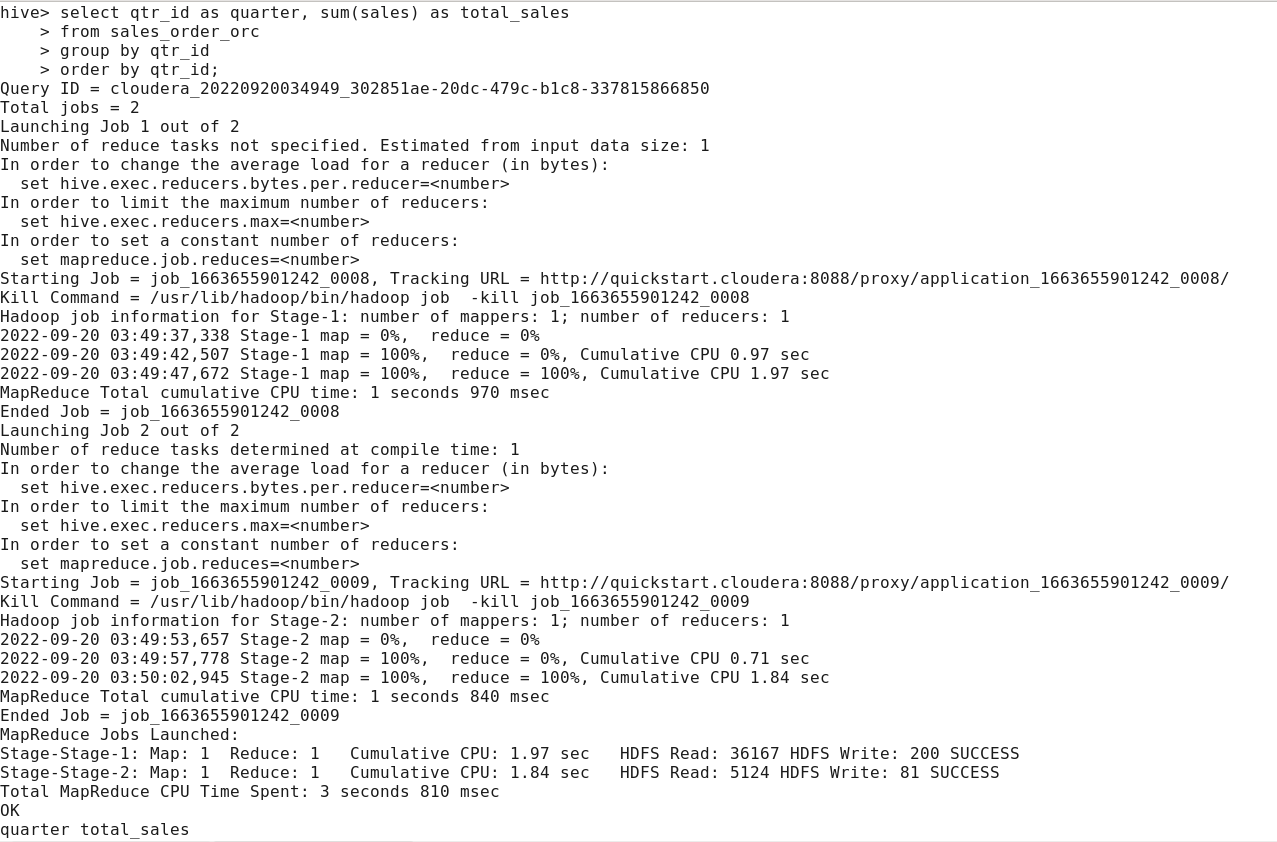


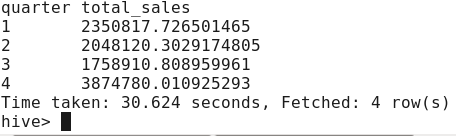
1. Find a product for which maximum orders were placed



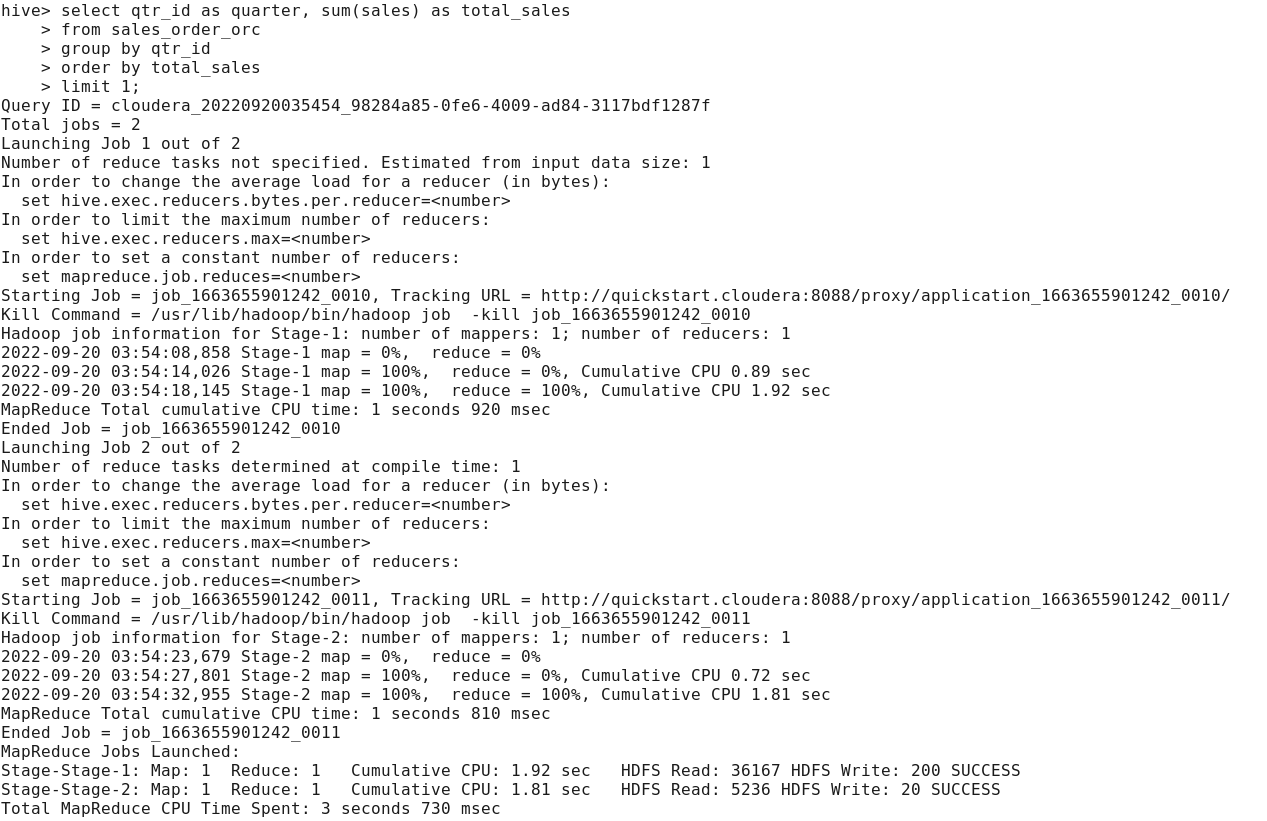


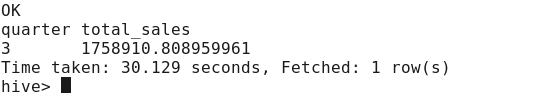
1. Calculate the total sales for each quarter





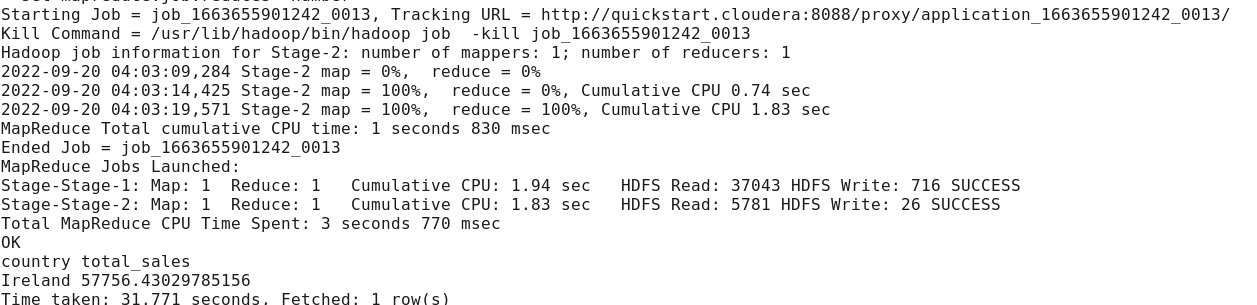
1. In which quarter sales was minimum



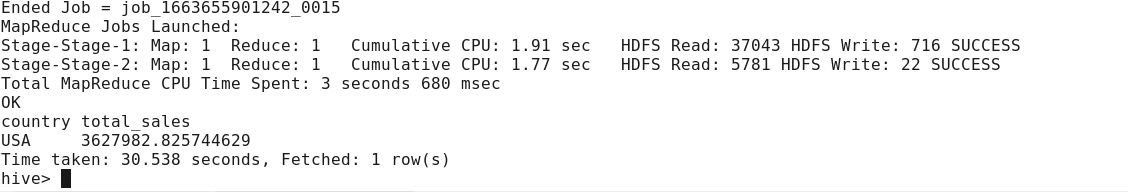


In which country sales was maximum and in which country sales was minimum

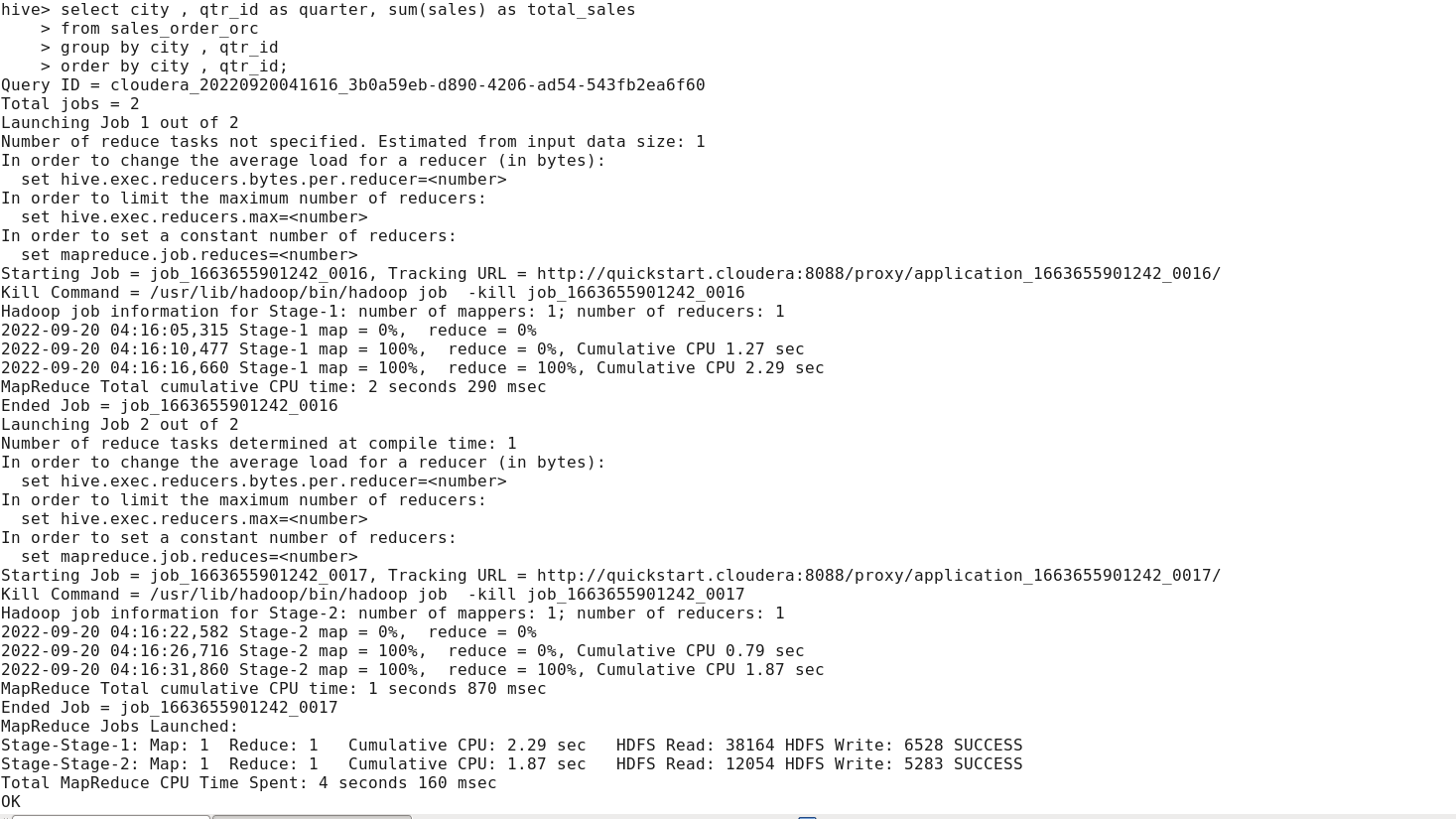


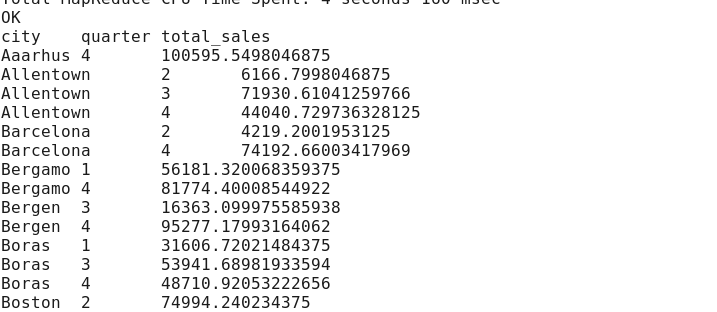






1. Calculate quartelry sales for each city





h. Find a month for each year in which maximum number of quantities were sold



