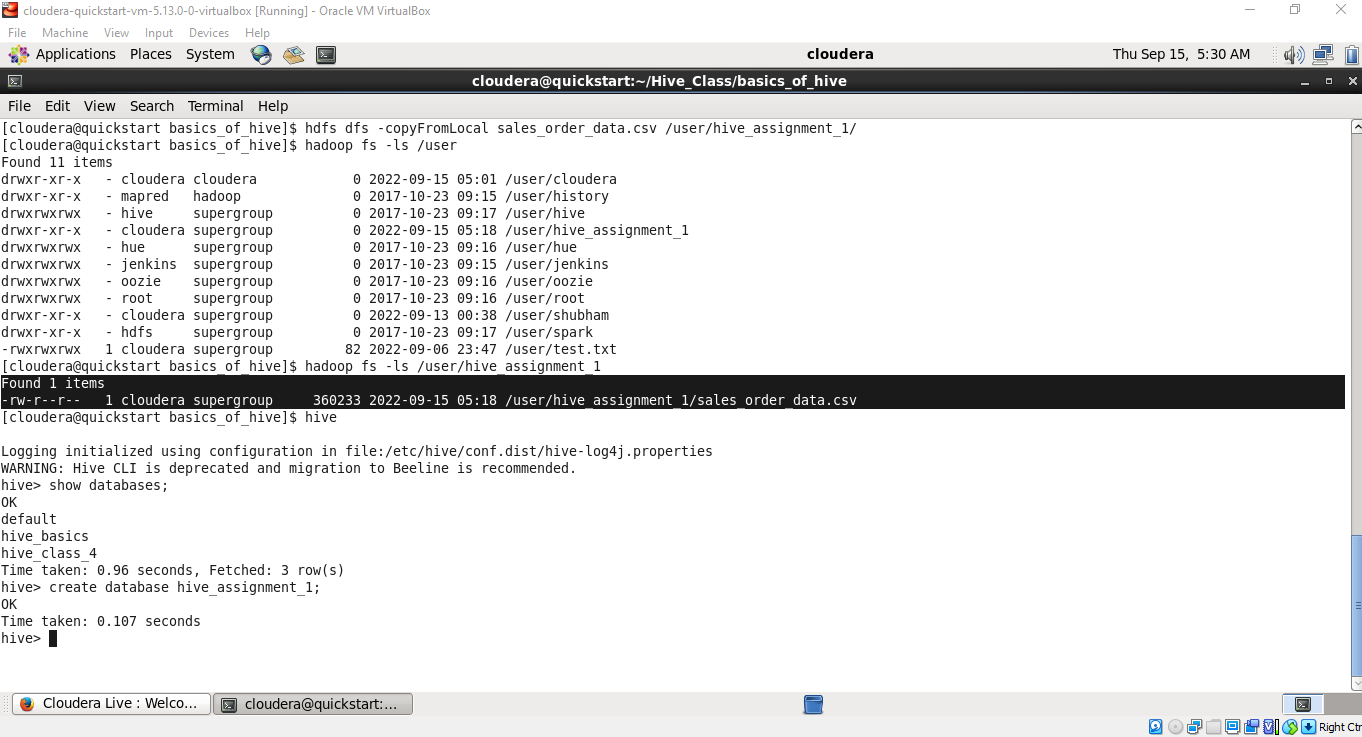
HIVE CHALLENGE 1

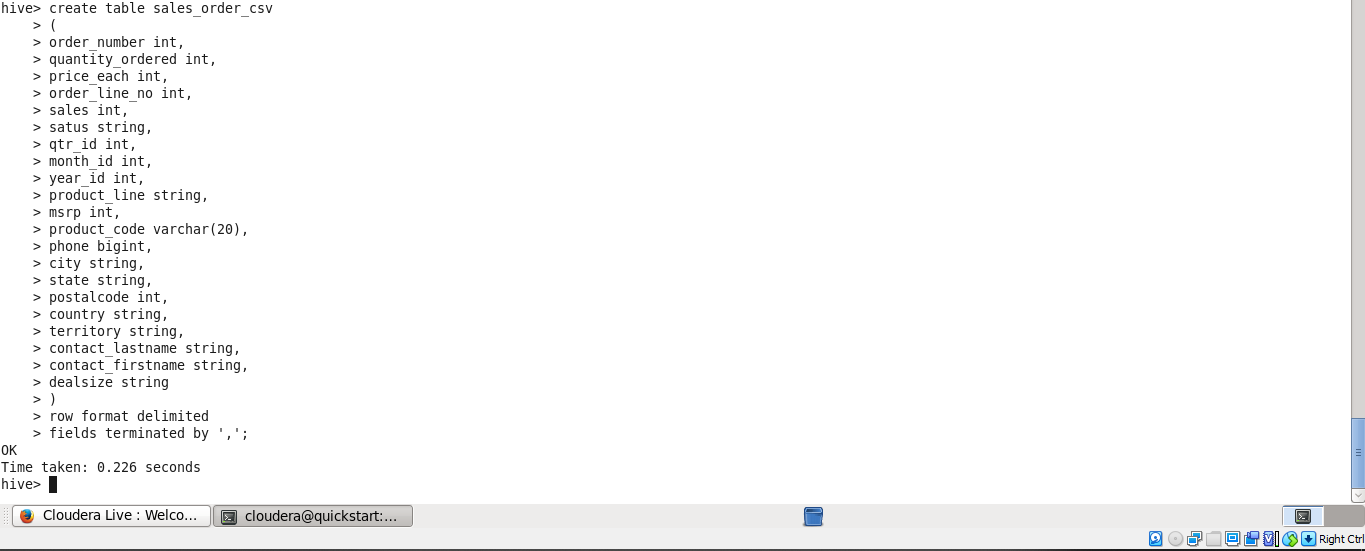
1. Download vehicle sales data ->

https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv

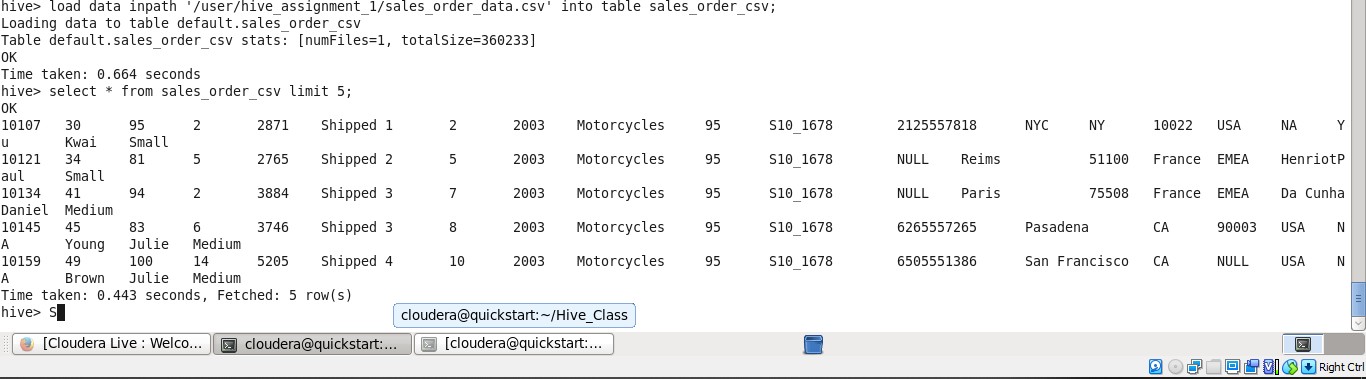
2. Store raw data into hdfs location



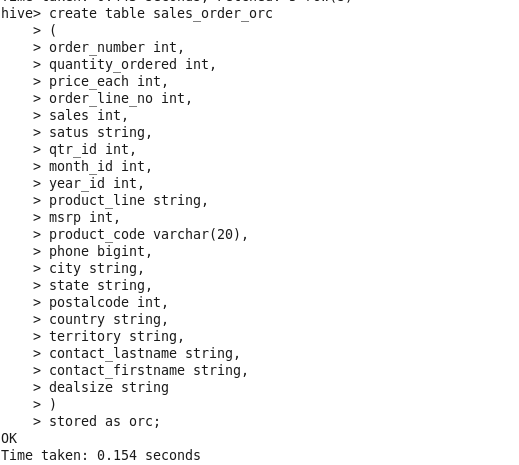
3. Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table



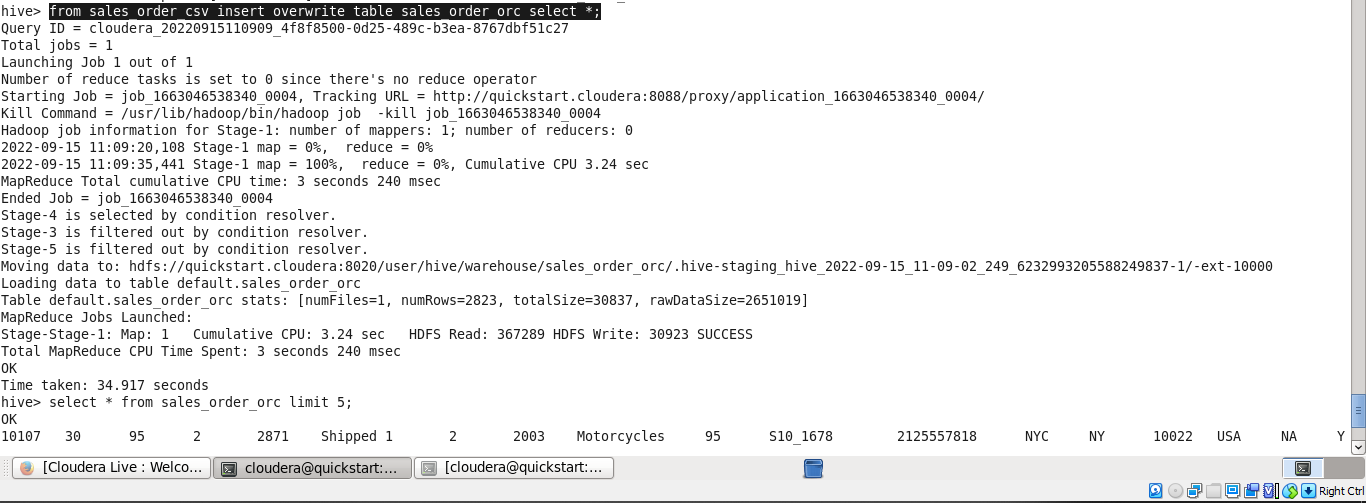
4. Load data from hdfs path into "sales\_order\_csv"



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

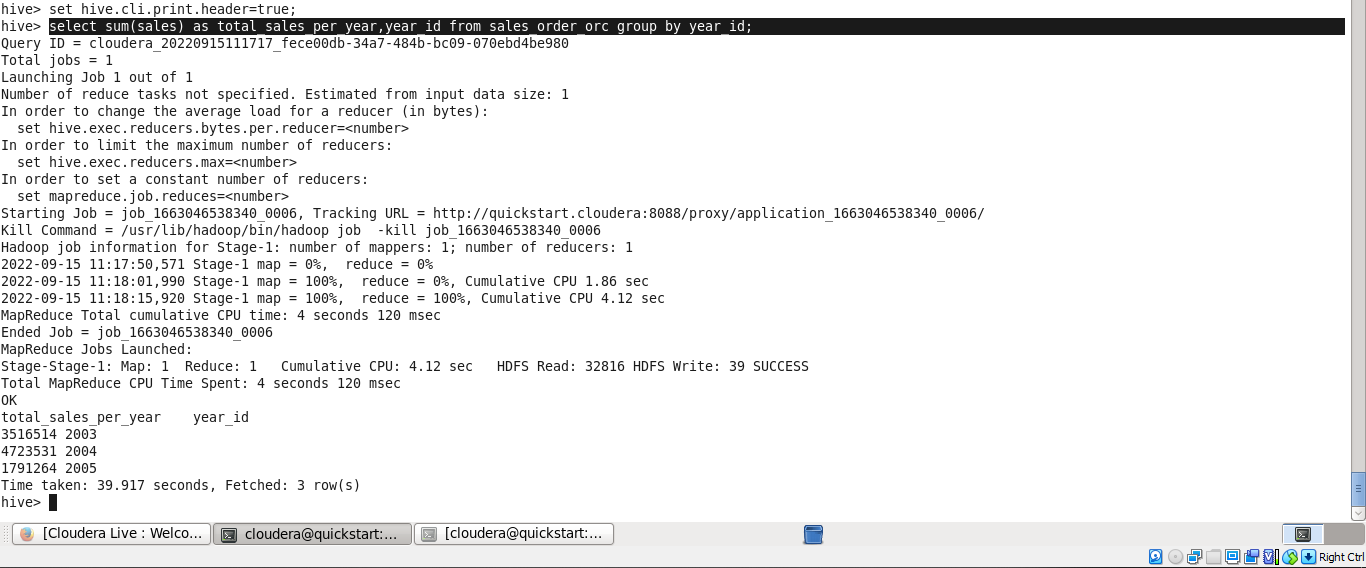


6. Load data from "sales\_order\_csv" into "sales\_order\_orc"

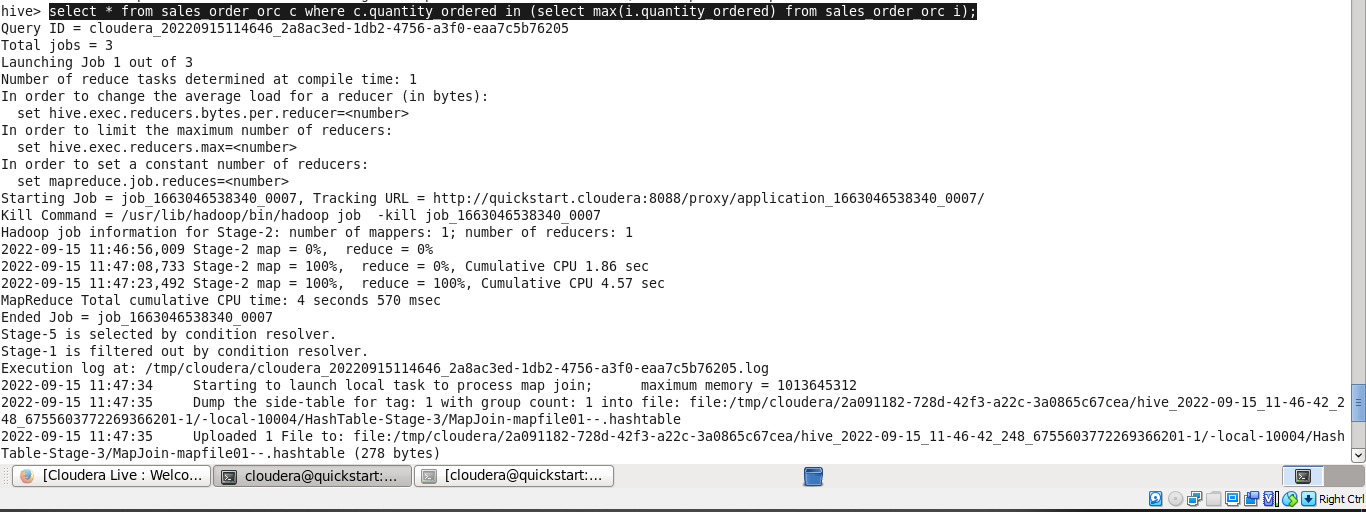


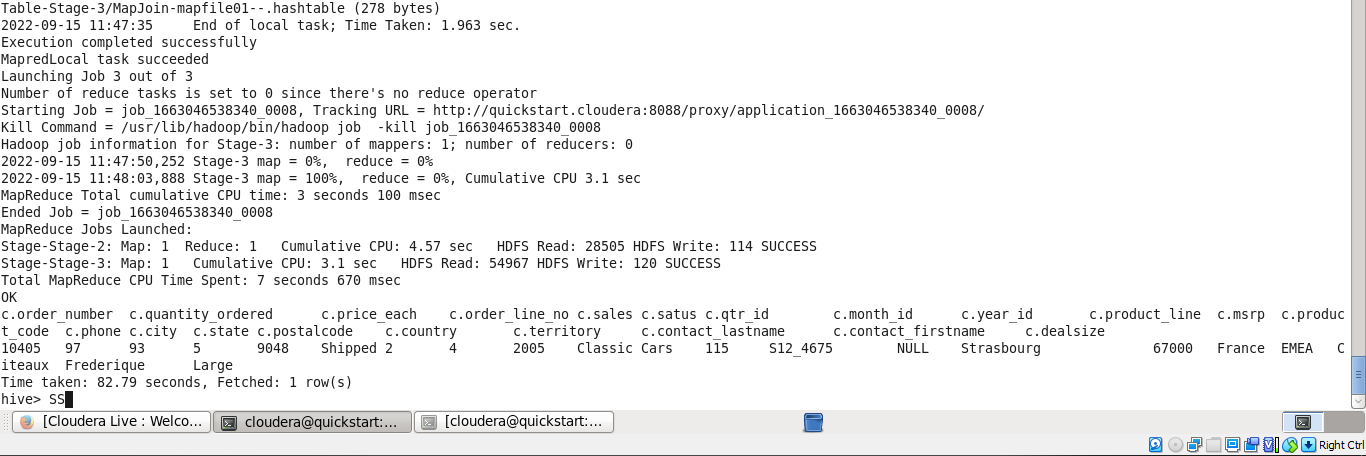
Perform below mentioned queries on "sales\_order\_orc" table :show

a. Calculate total sales per year

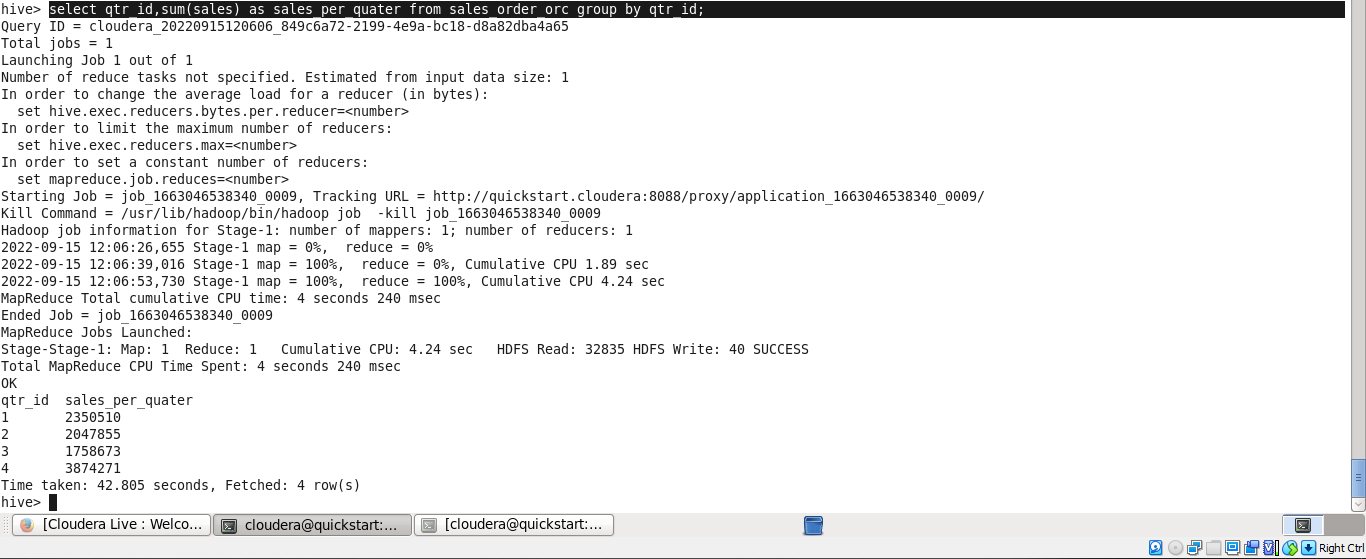


b. Find a product for which maximum orders were placed





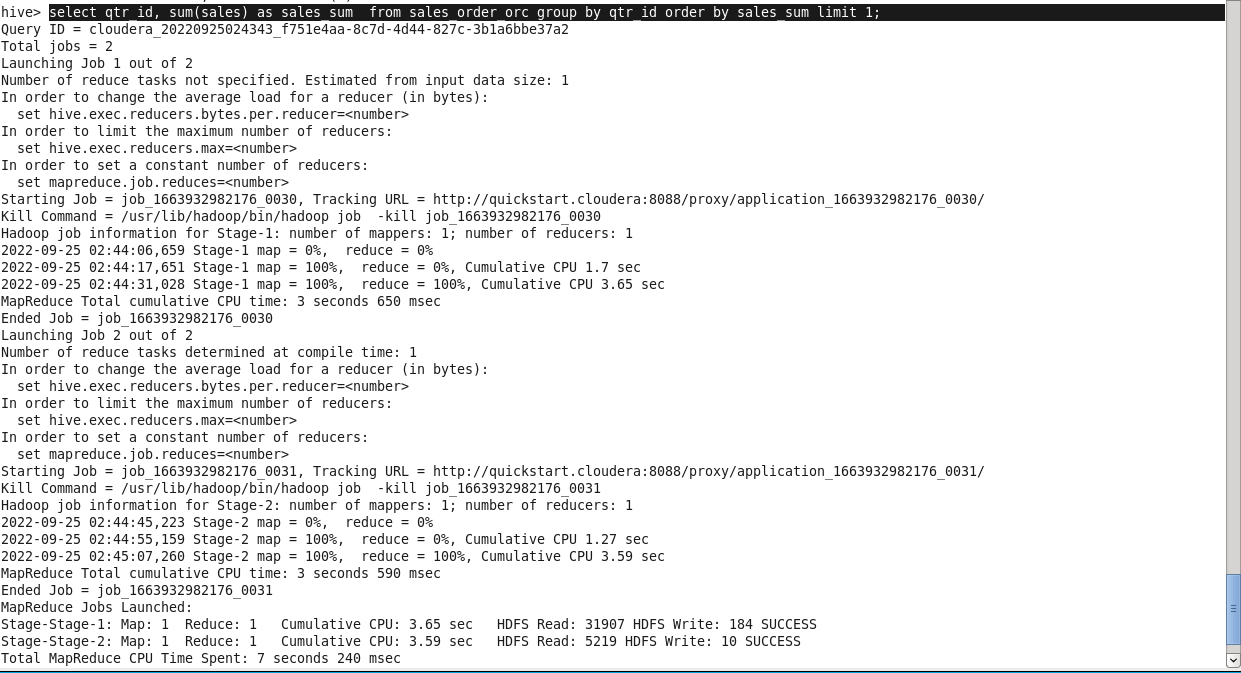
c. Calculate the total sales for each quarter

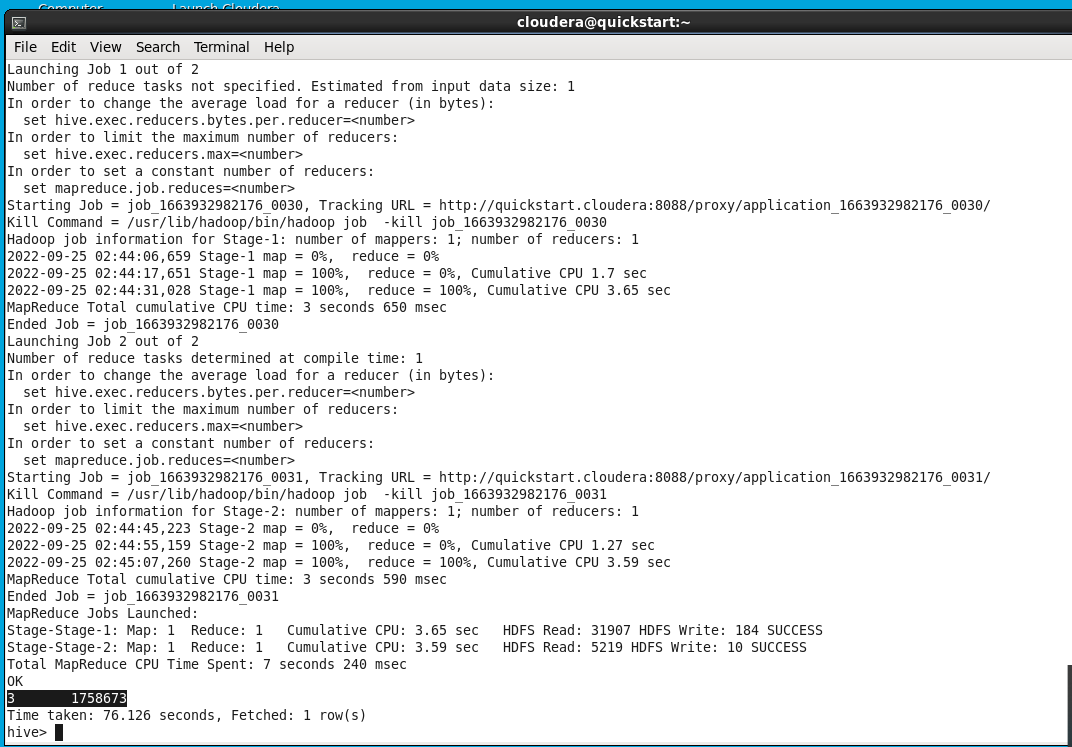


d. In which quarter sales was minimum

Step 1: aggregate the sales group by quarter

Step 2 : find quarter having min sales



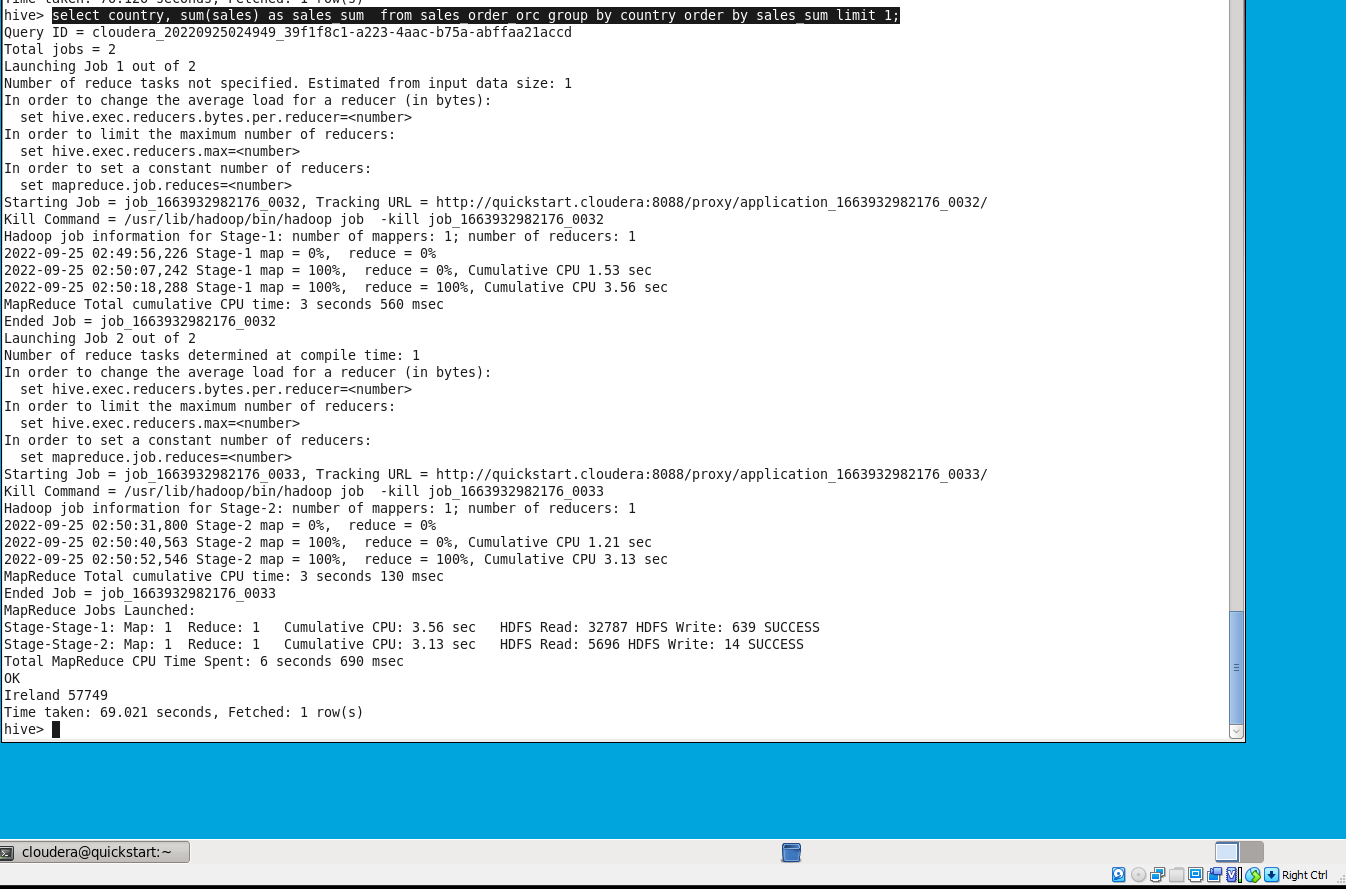


e. In which country sales was maximum and in which country sales was minimum

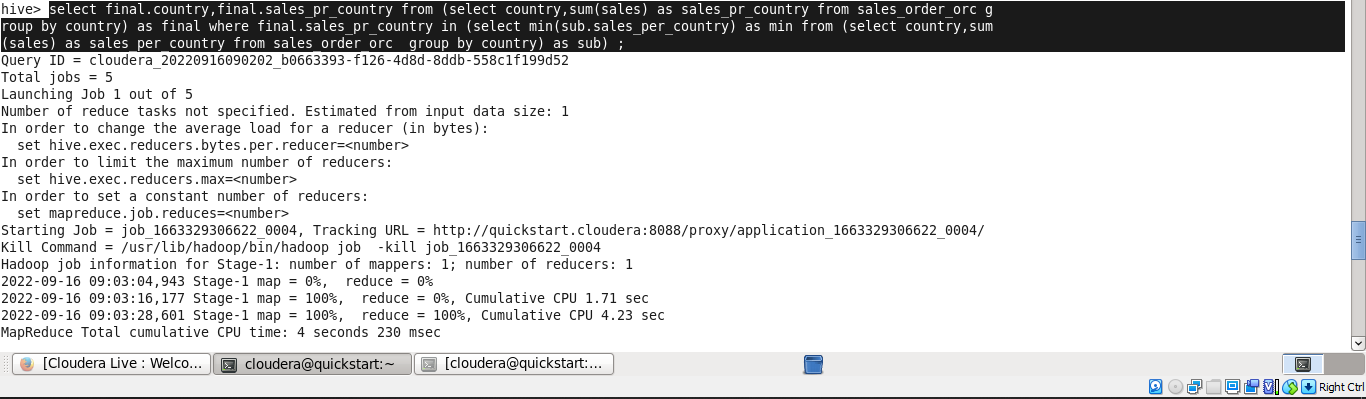
step 1: find sales data group by country

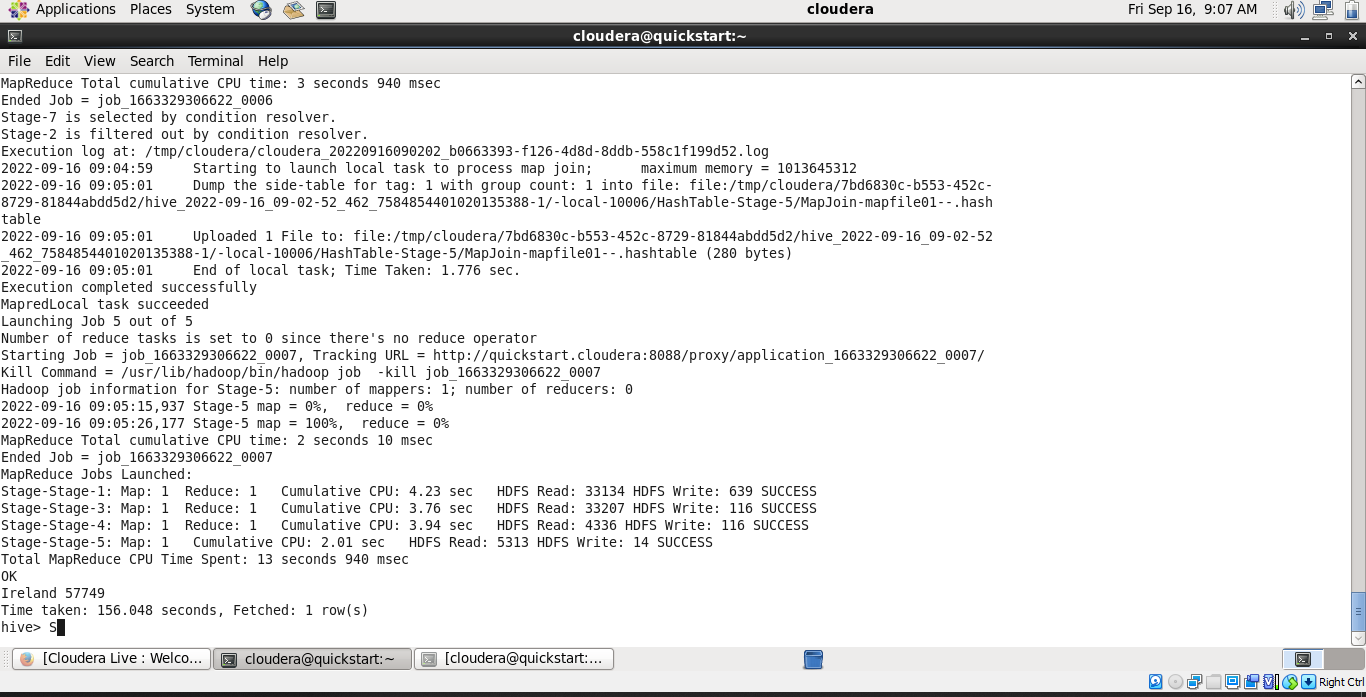
step 2: find country having min(sales)

WAY 1: using group by and order by in same query



WAY 2: using the harder way

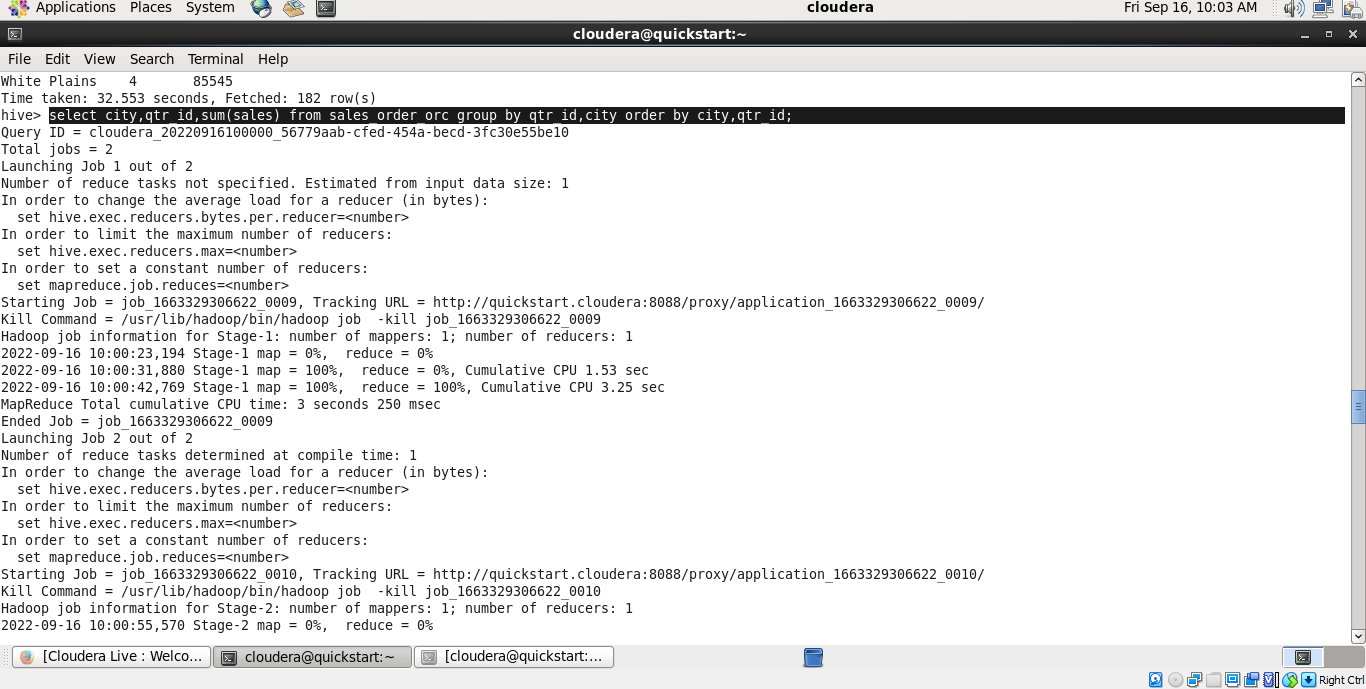


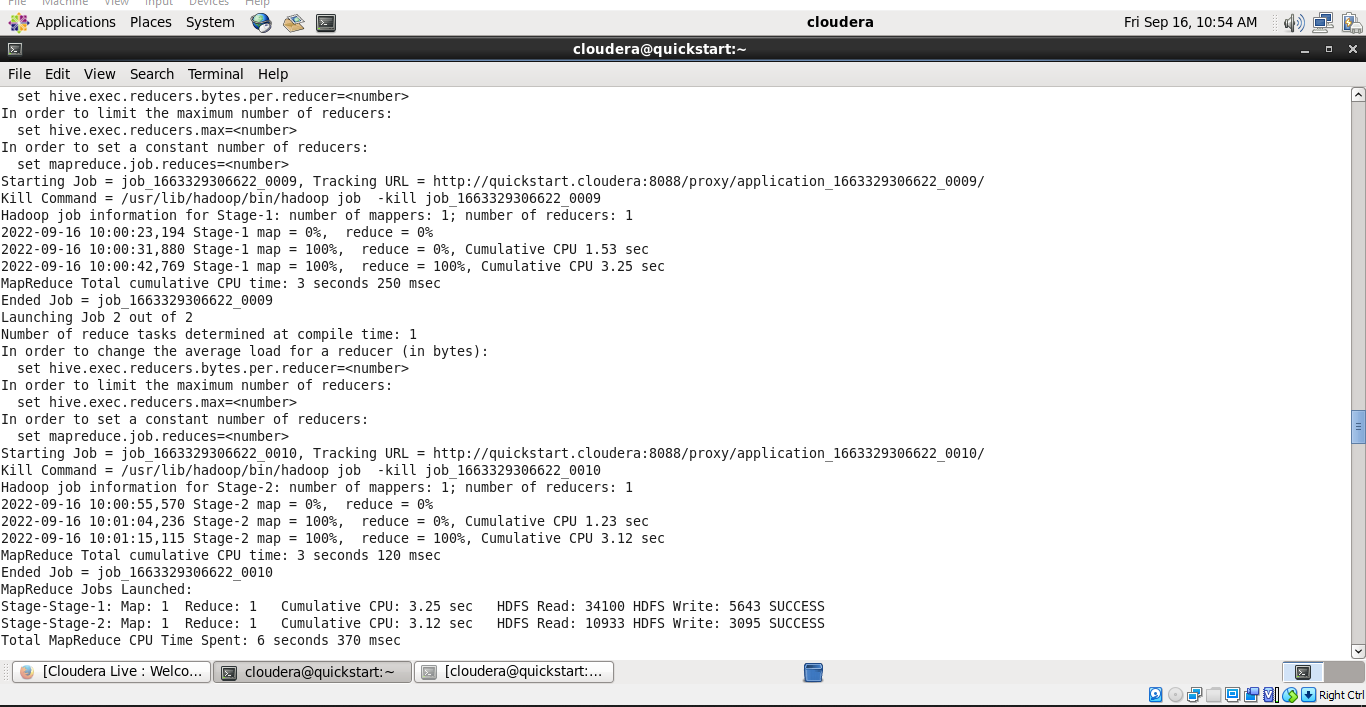


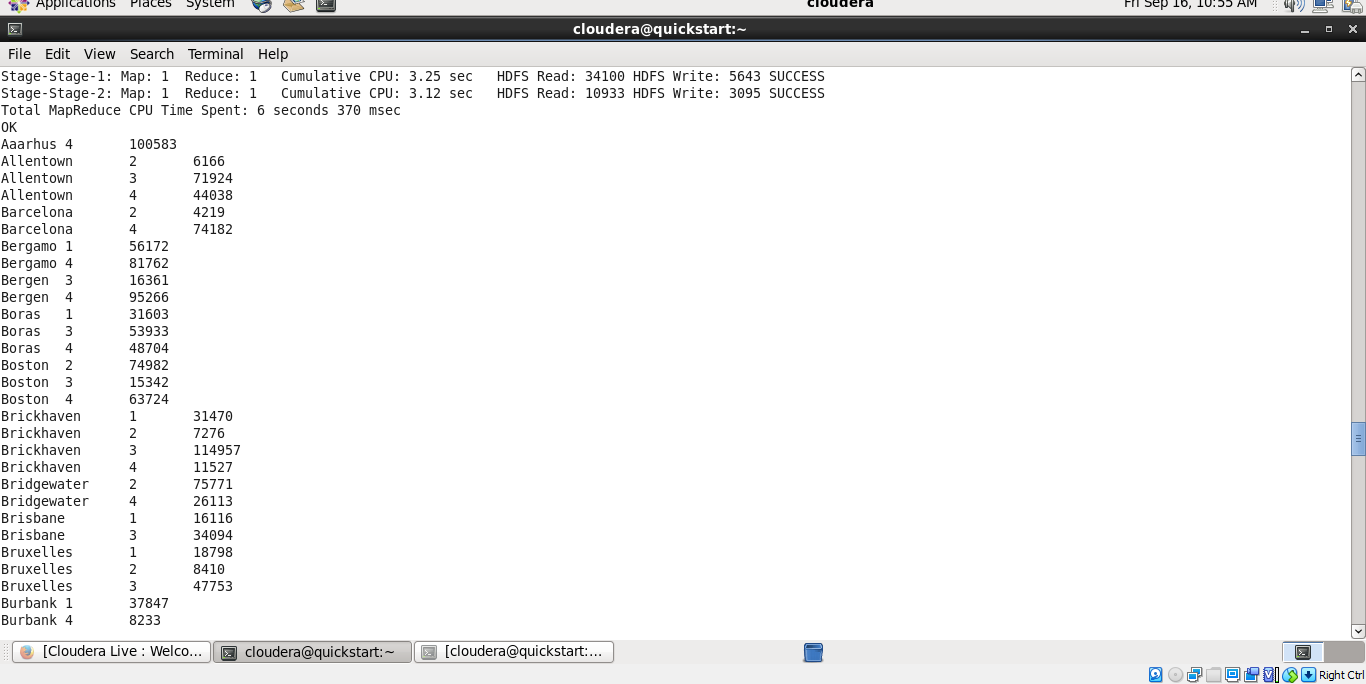
f. Calculate quartelry sales for each city

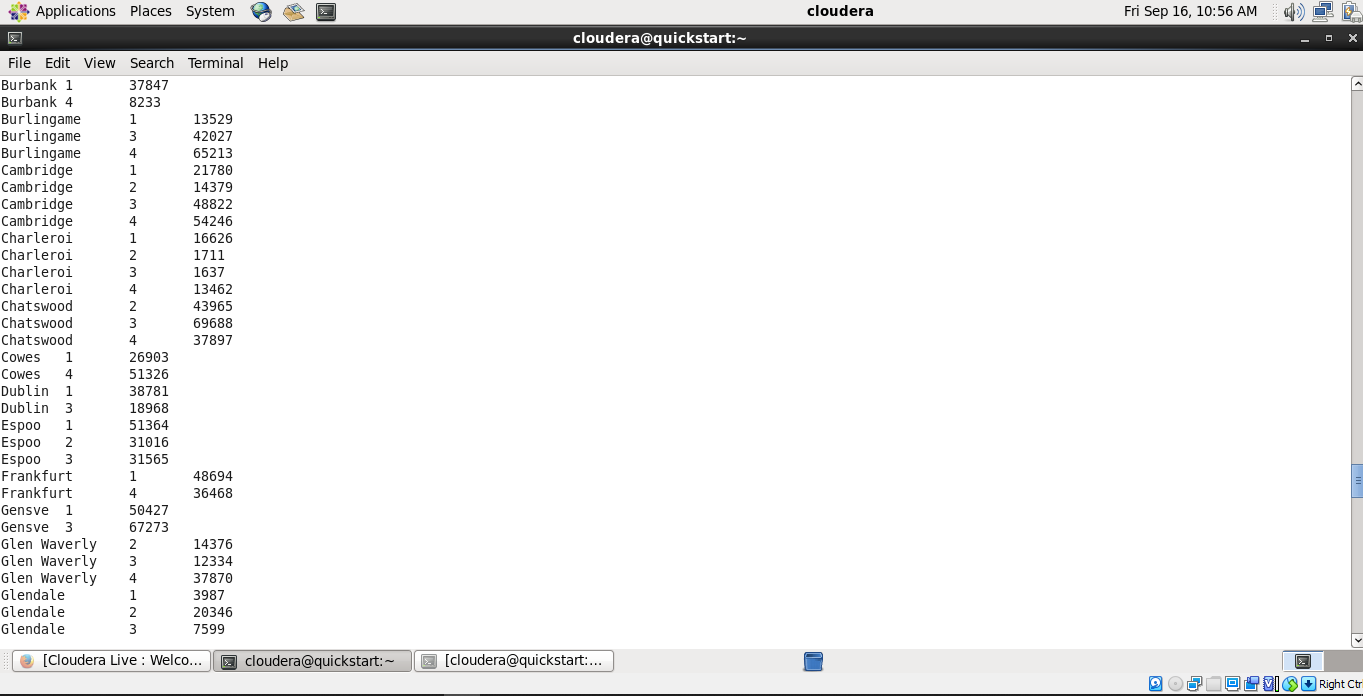
step 1: find quarterly sales

step 2: group by city

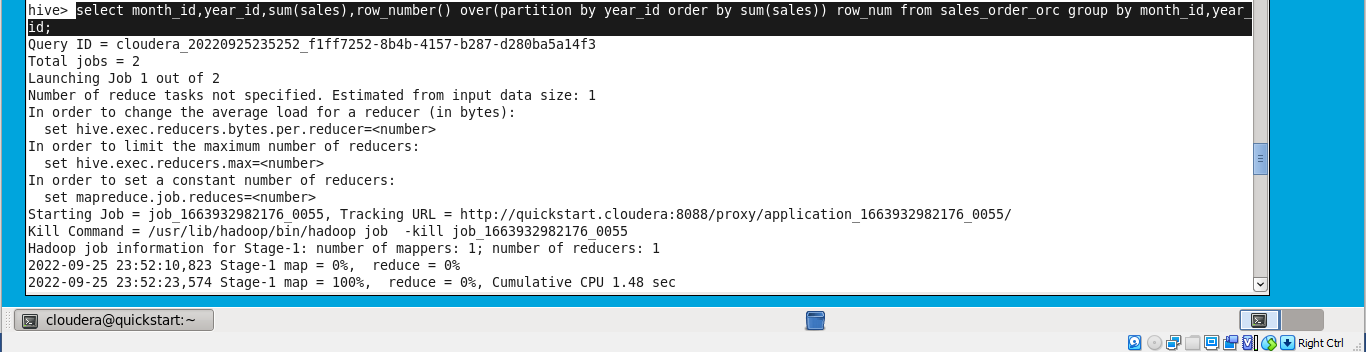


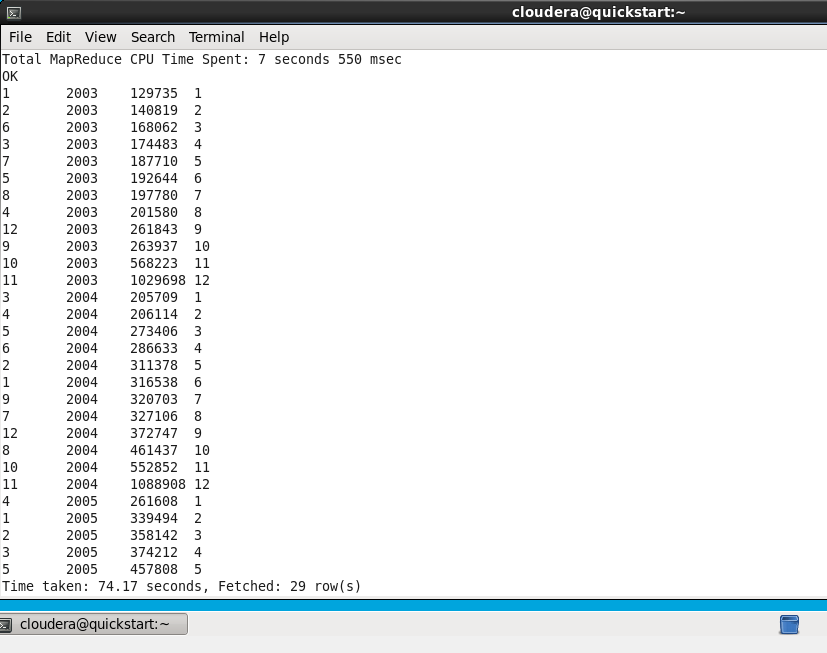






h. Find a month for each year in which maximum sales happened





The Final Filtering on this intermediate output:

