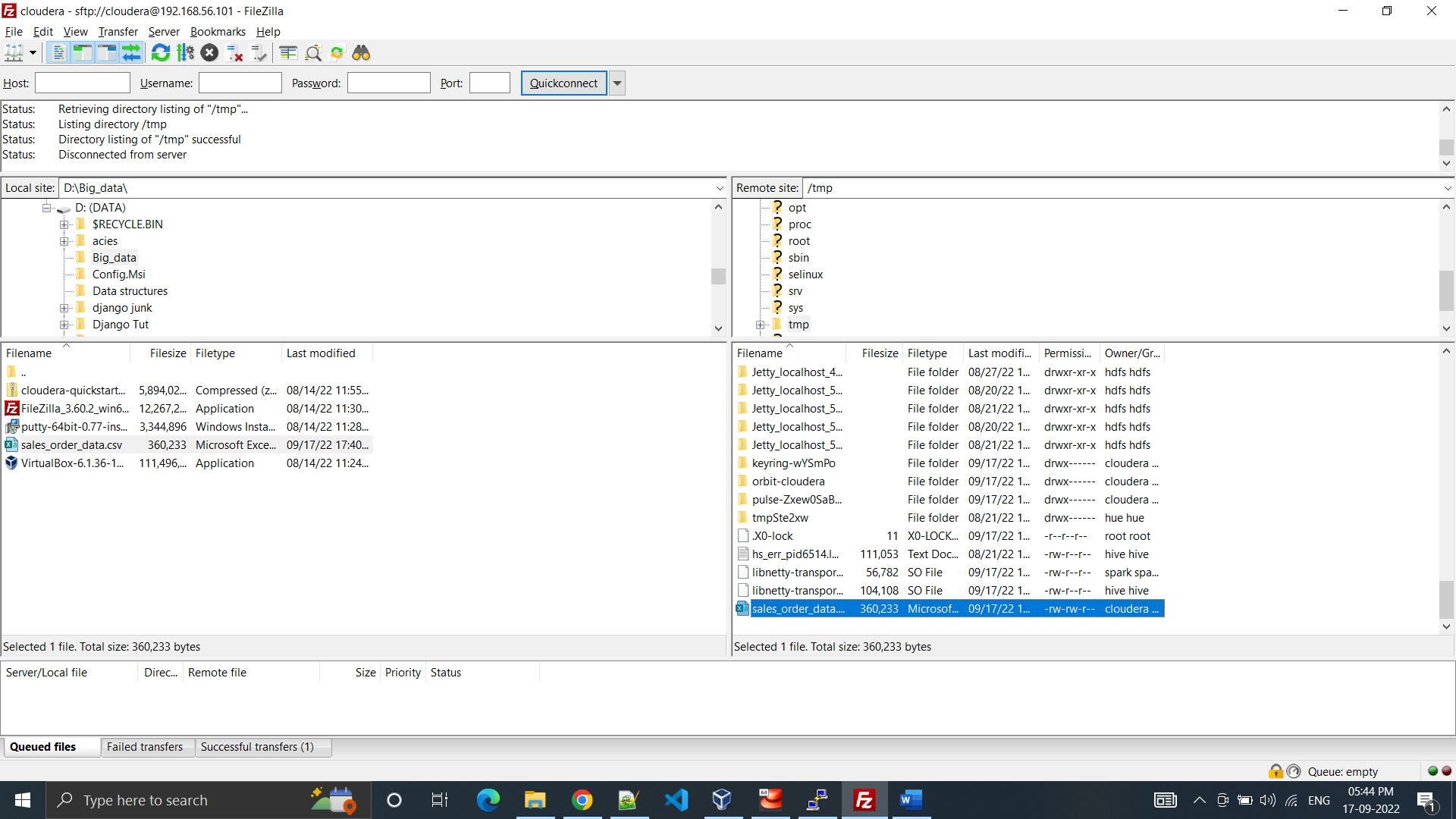
**Assignment – Hive**

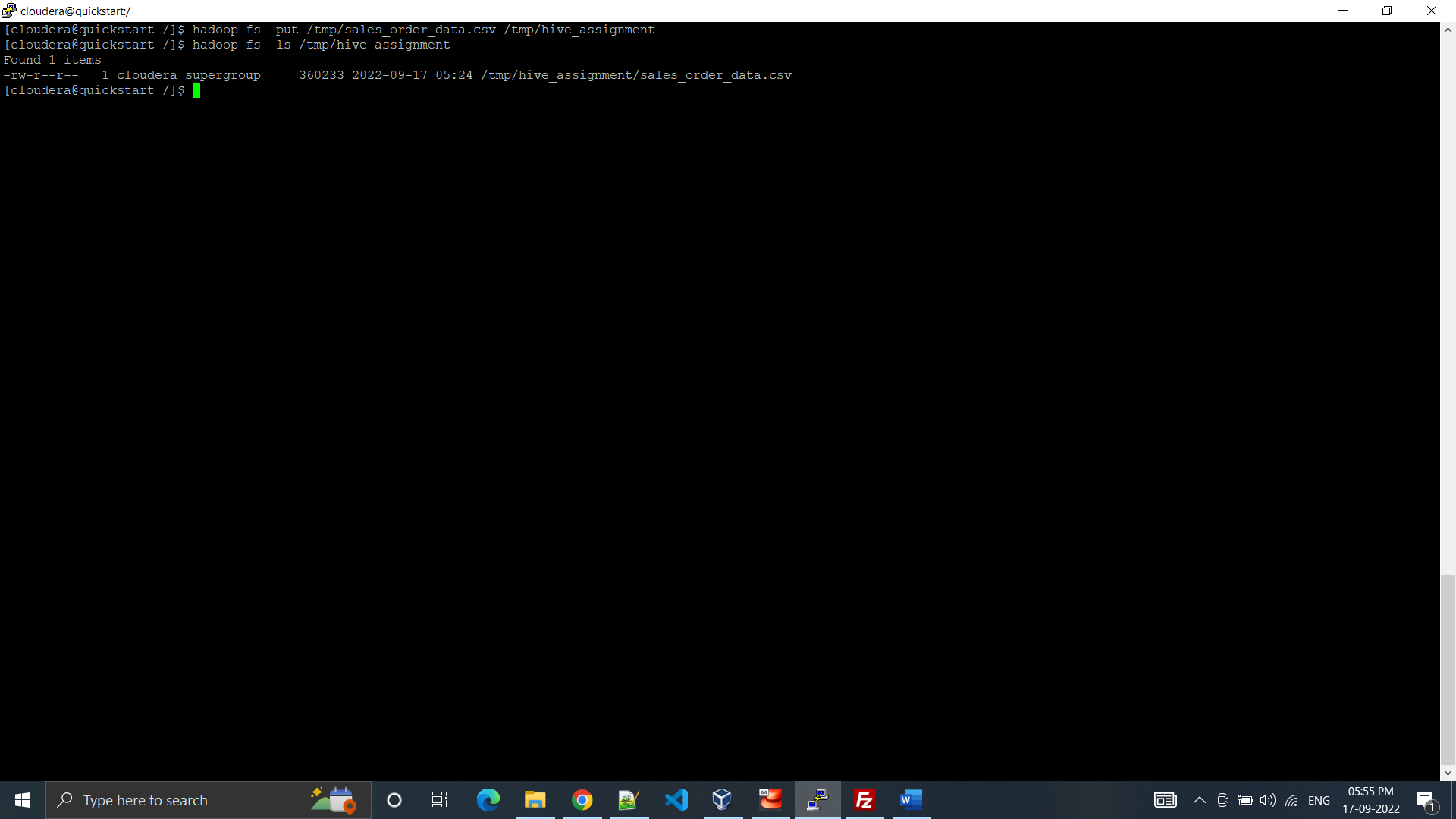
1. Download vehicle sales data -> <https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv>



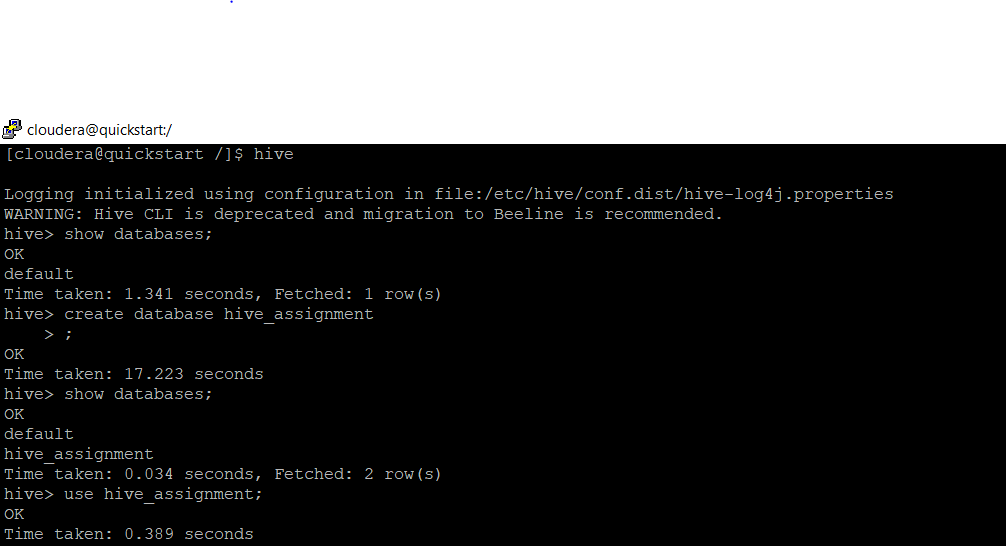
Text

Description automatically generated

2. Store raw data into hdfs location



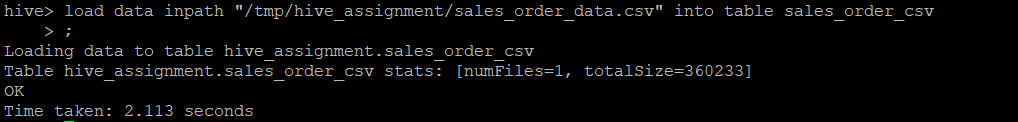
3. Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table

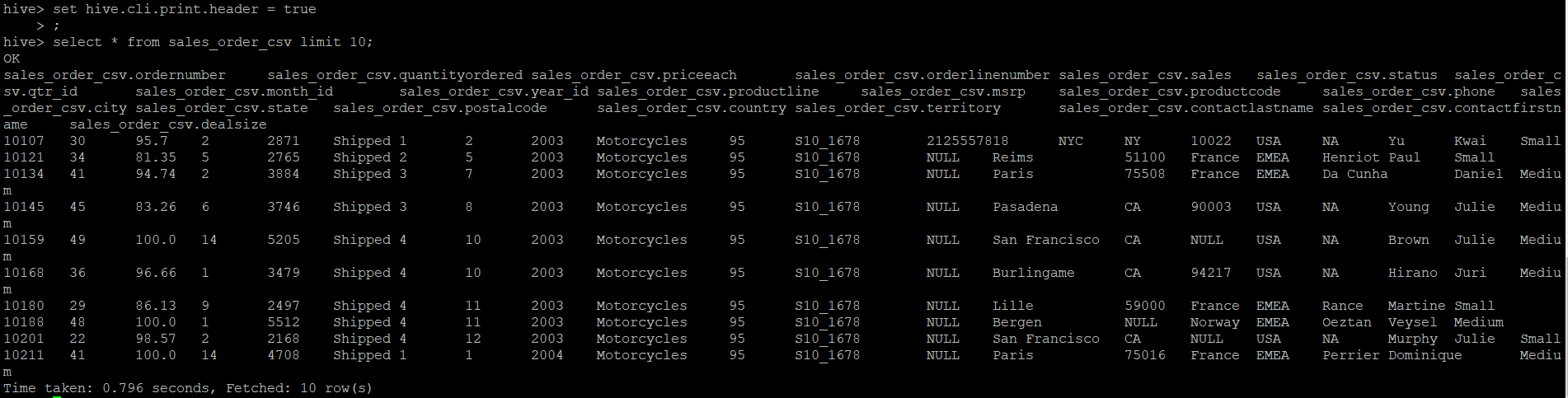


Text

Description automatically generated

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



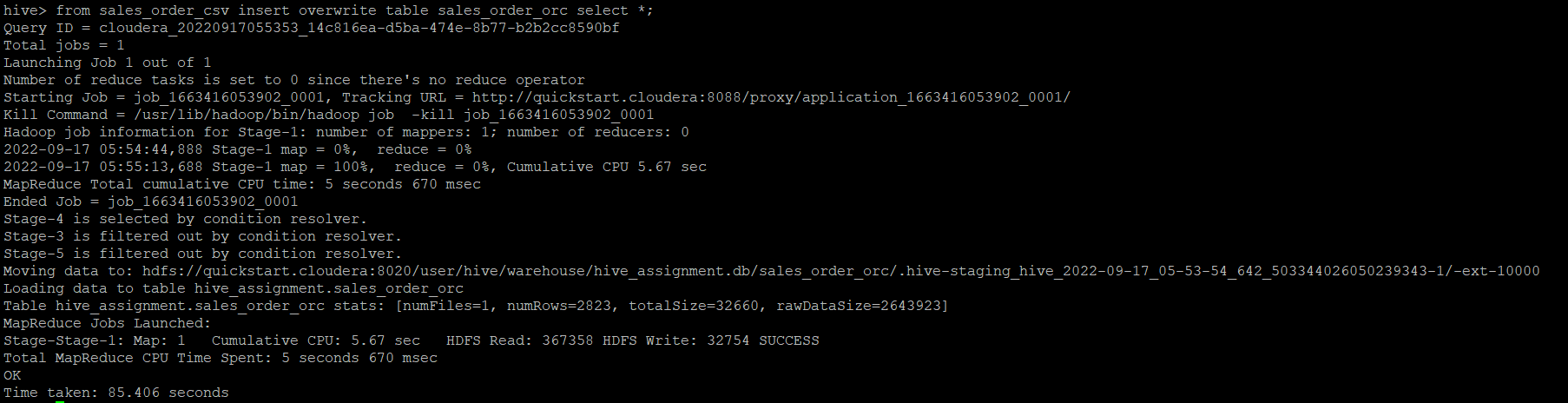


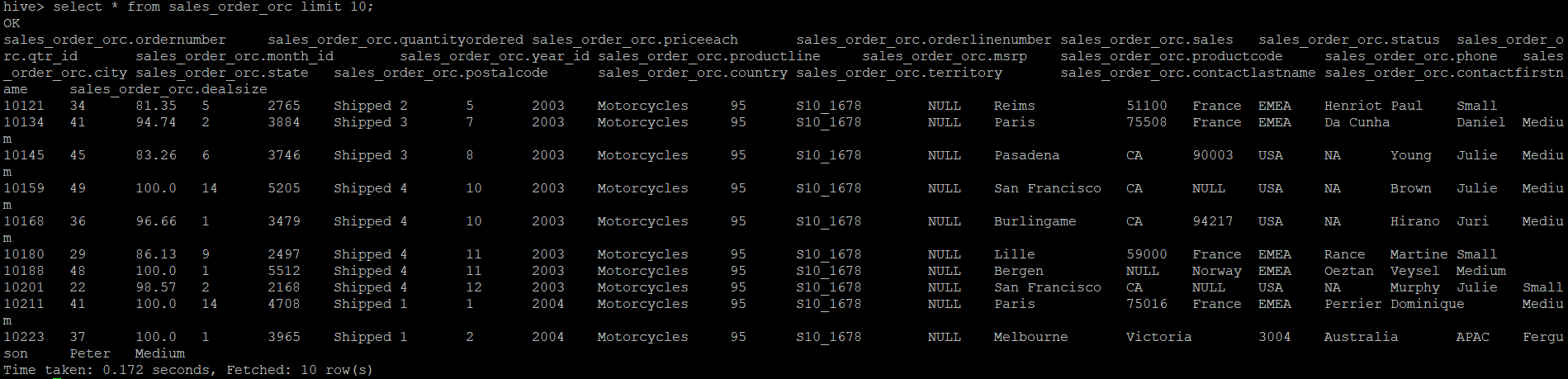
5. Create an internal hive table which will store data in ORC format "sales\_order\_orc".

Text

Description automatically generated

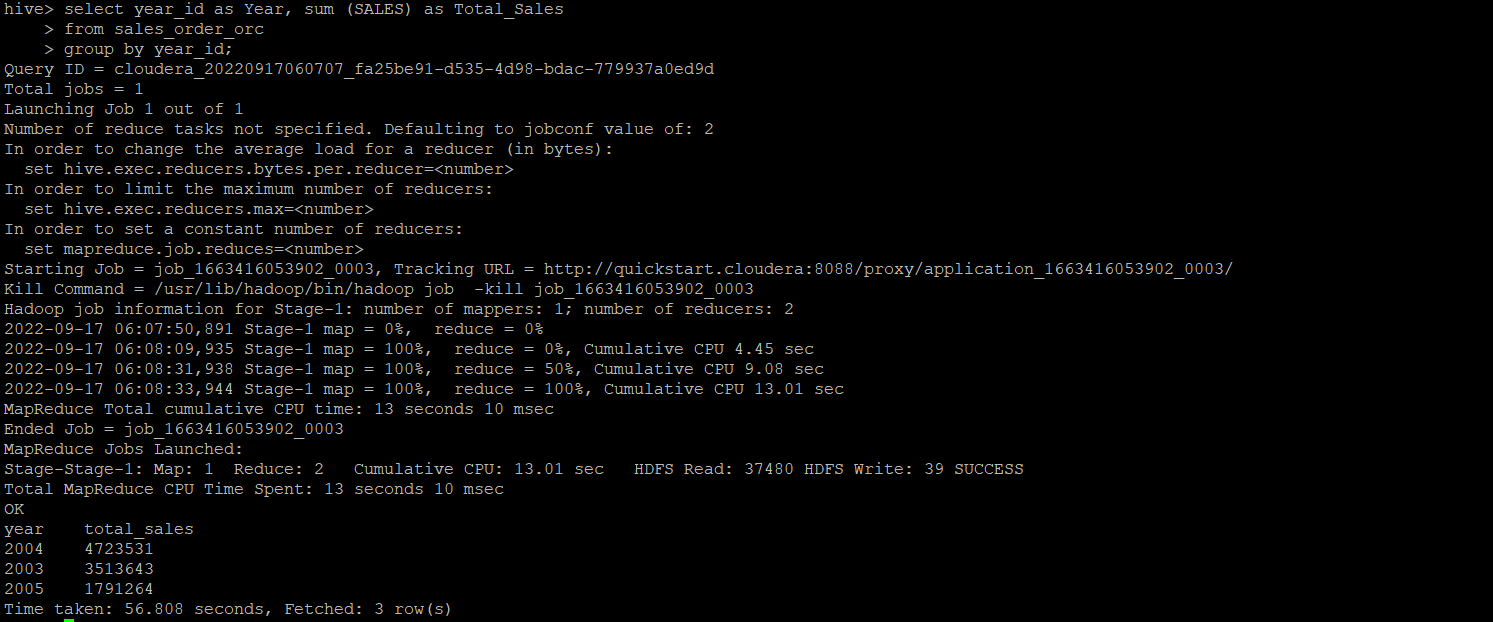
6. Load data from "sales\_order\_csv" into "sales\_order\_orc"



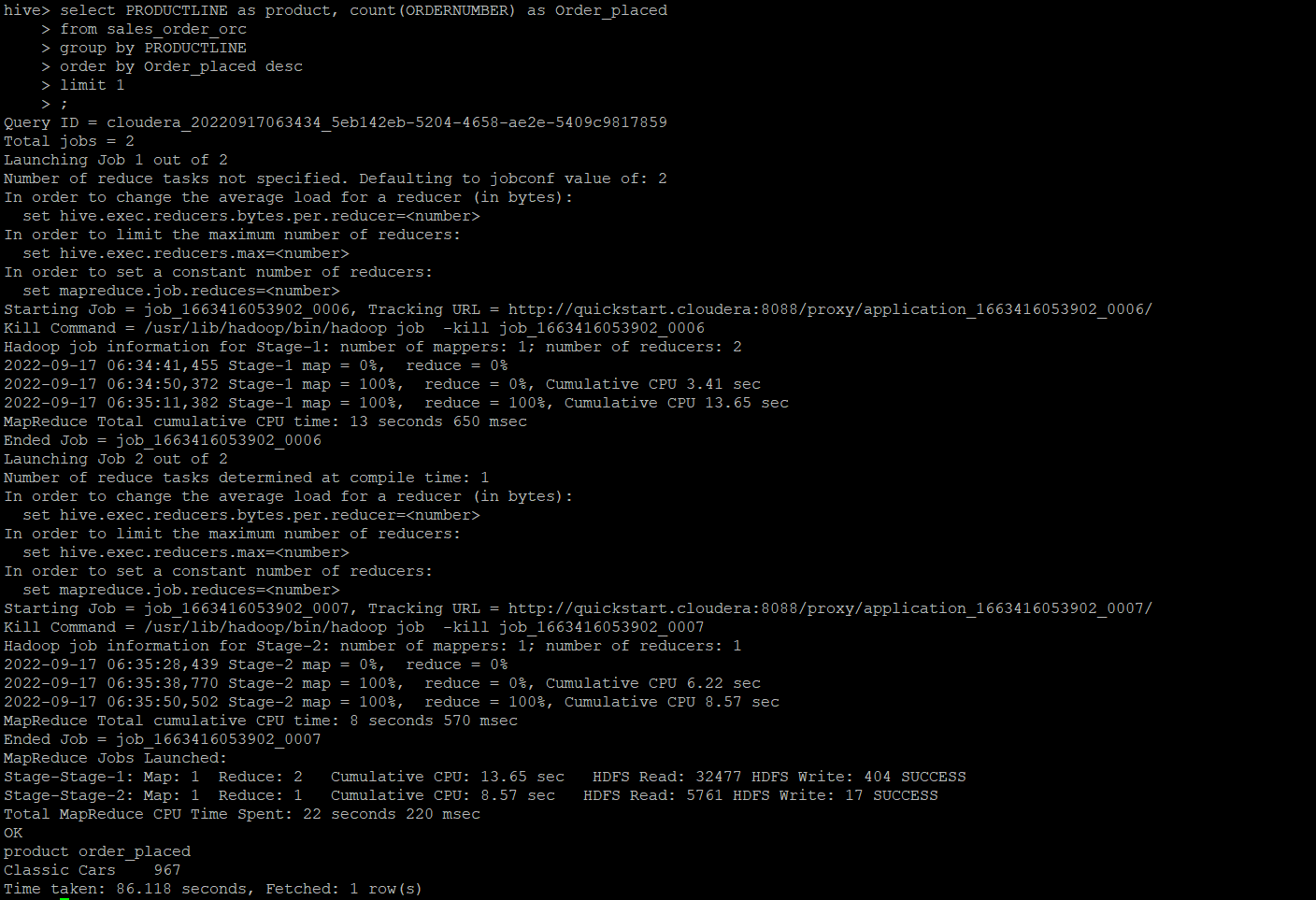


Perform below mentioned queries on "sales\_order\_orc" table :

1. Calculate total sales per year



b. Find a product for which maximum orders were placed



c. Calculate the total sales for each quarter

Text

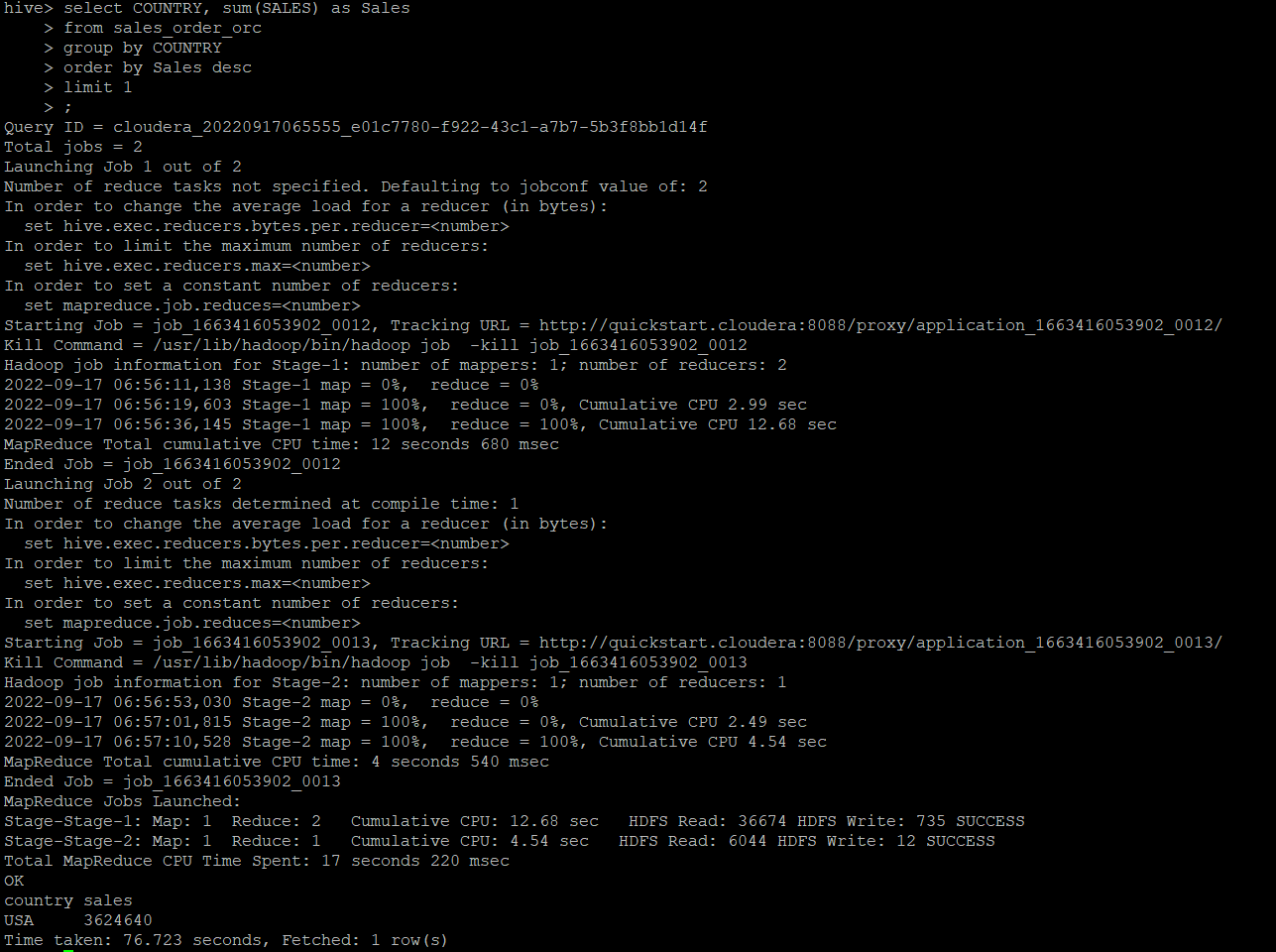
Description automatically generated

d. In which quarter sales was minimum

Text

Description automatically generated

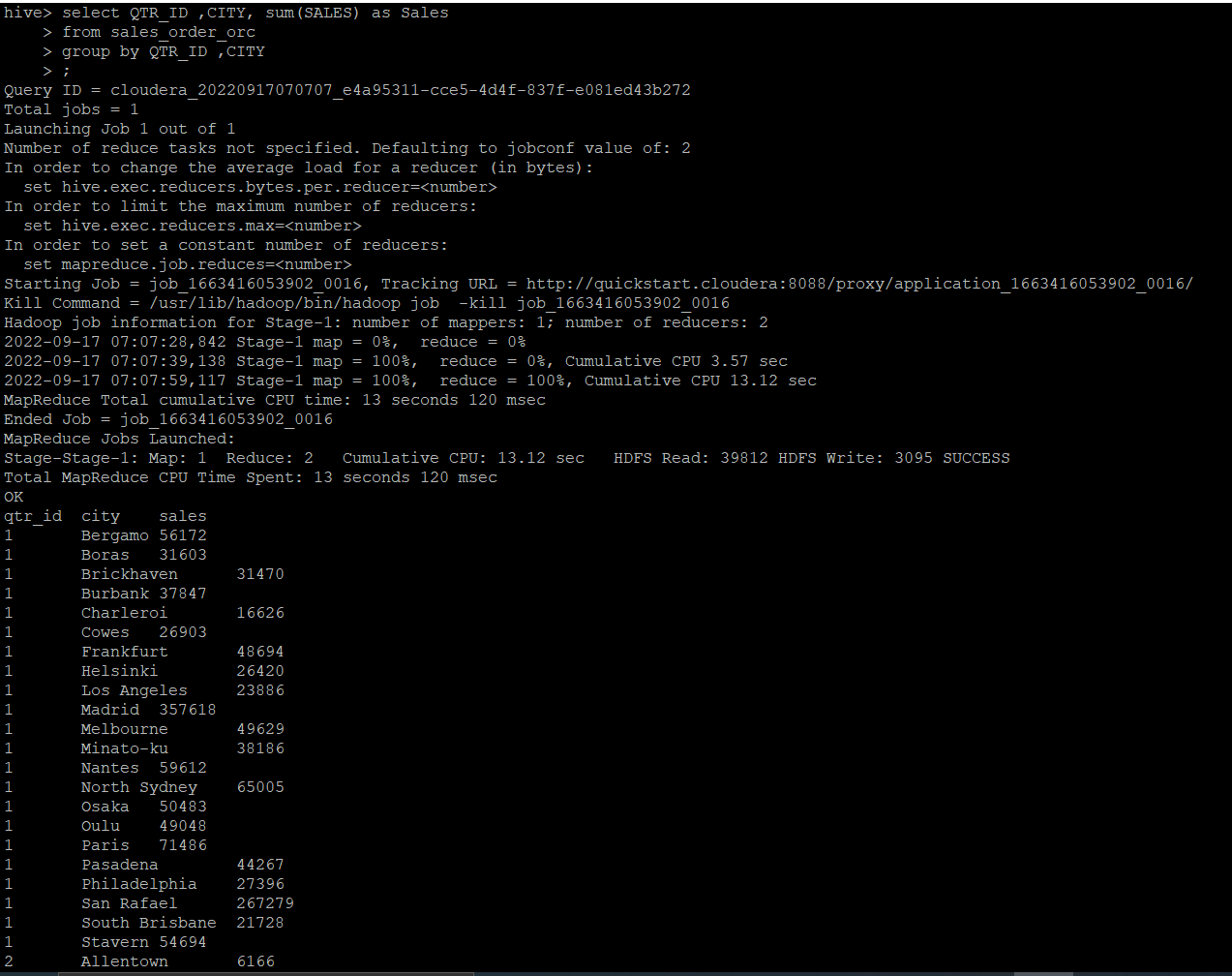
e. In which country sales was maximum and in which country sales was minimum



Text

Description automatically generated

f. Calculate quarterly sales for each city



Text

Description automatically generated

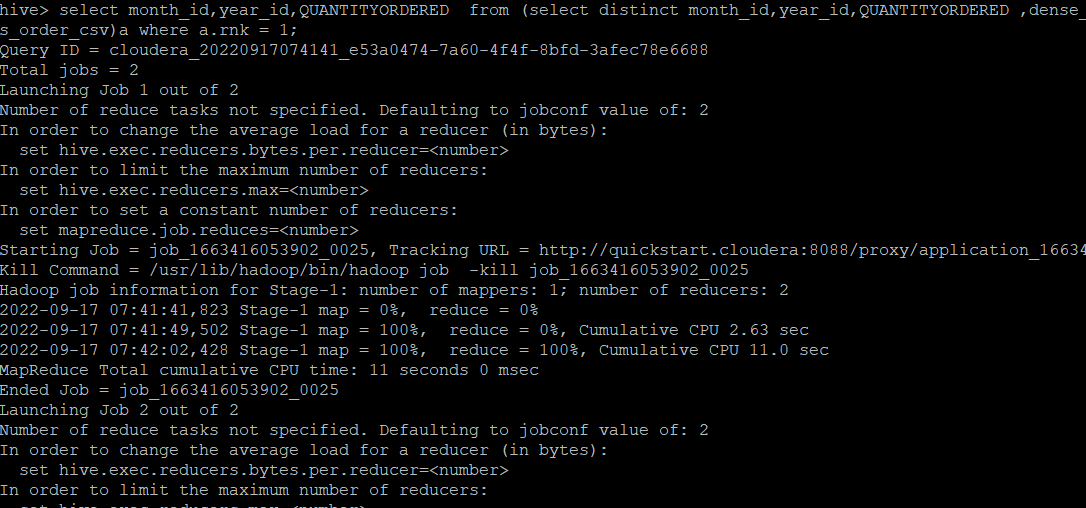
Graphical user interface

Description automatically generated

Text

Description automatically generated with low confidence

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



Text

Description automatically generated