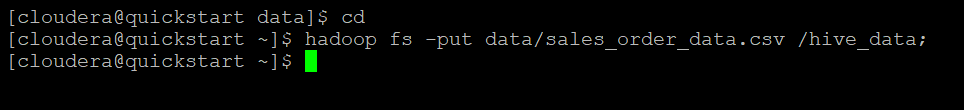
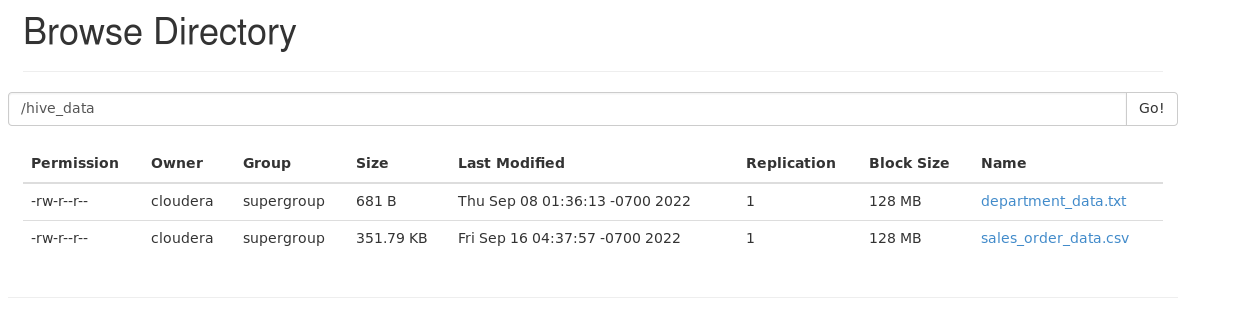
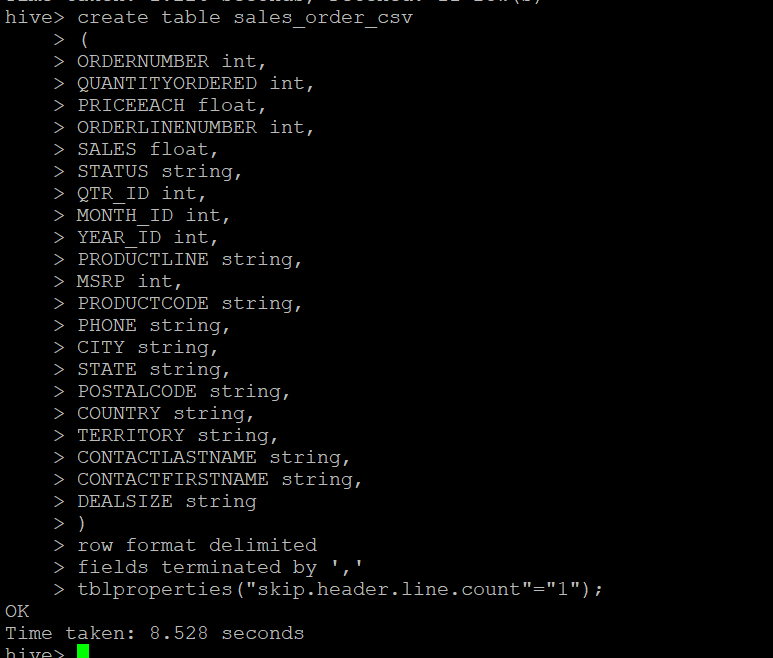
1. **Store raw data into hdfs location:**

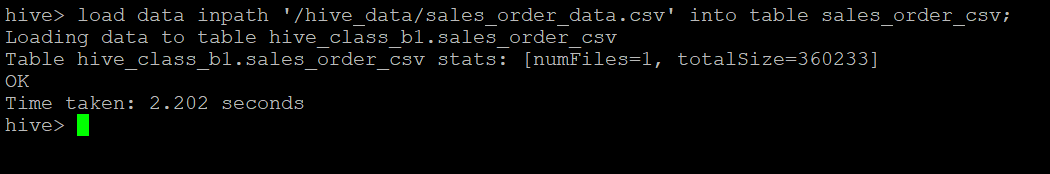




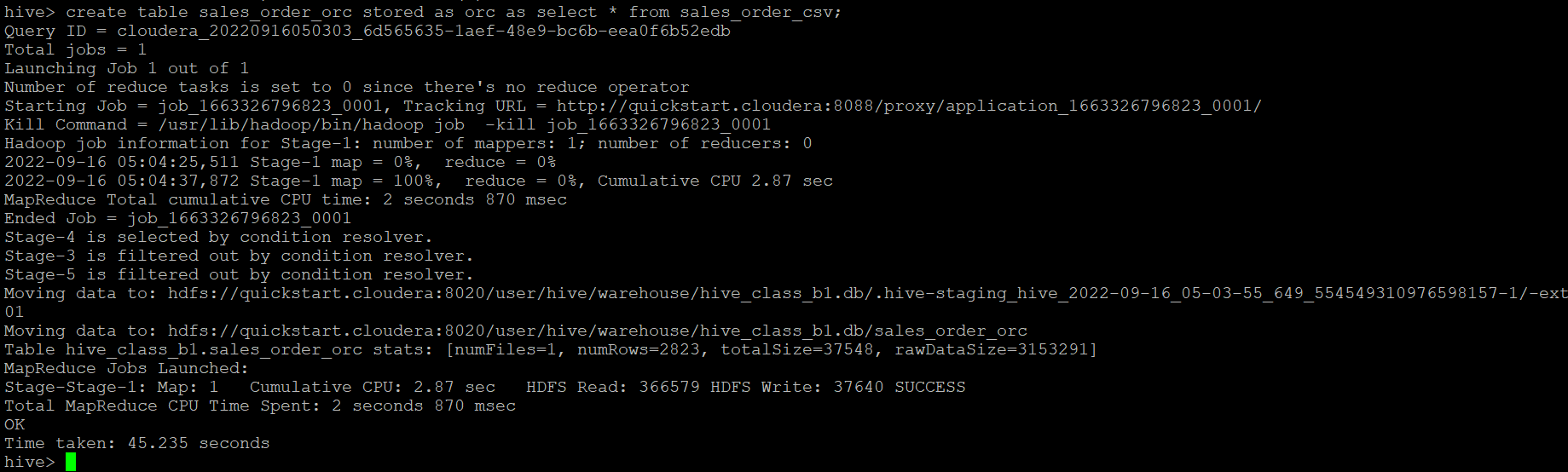
1. **Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table.**



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



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

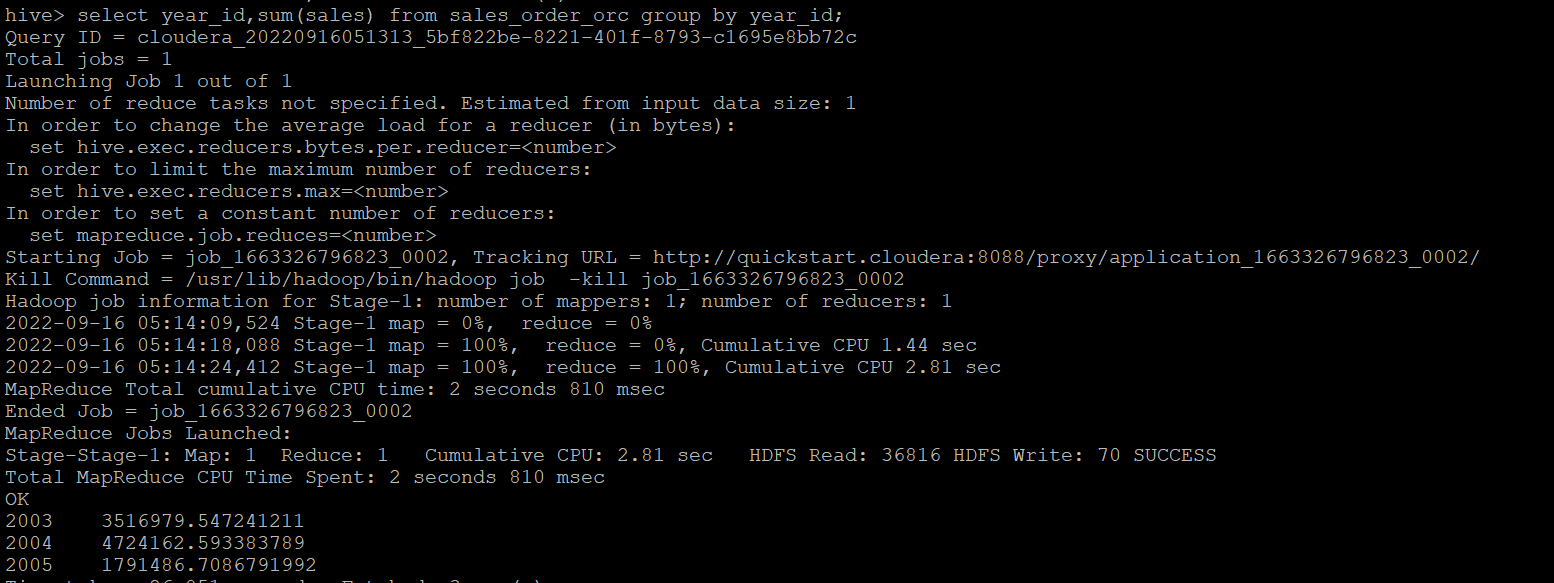


5. **Load data from "sales\_order\_csv" into "sales\_order\_orc"**

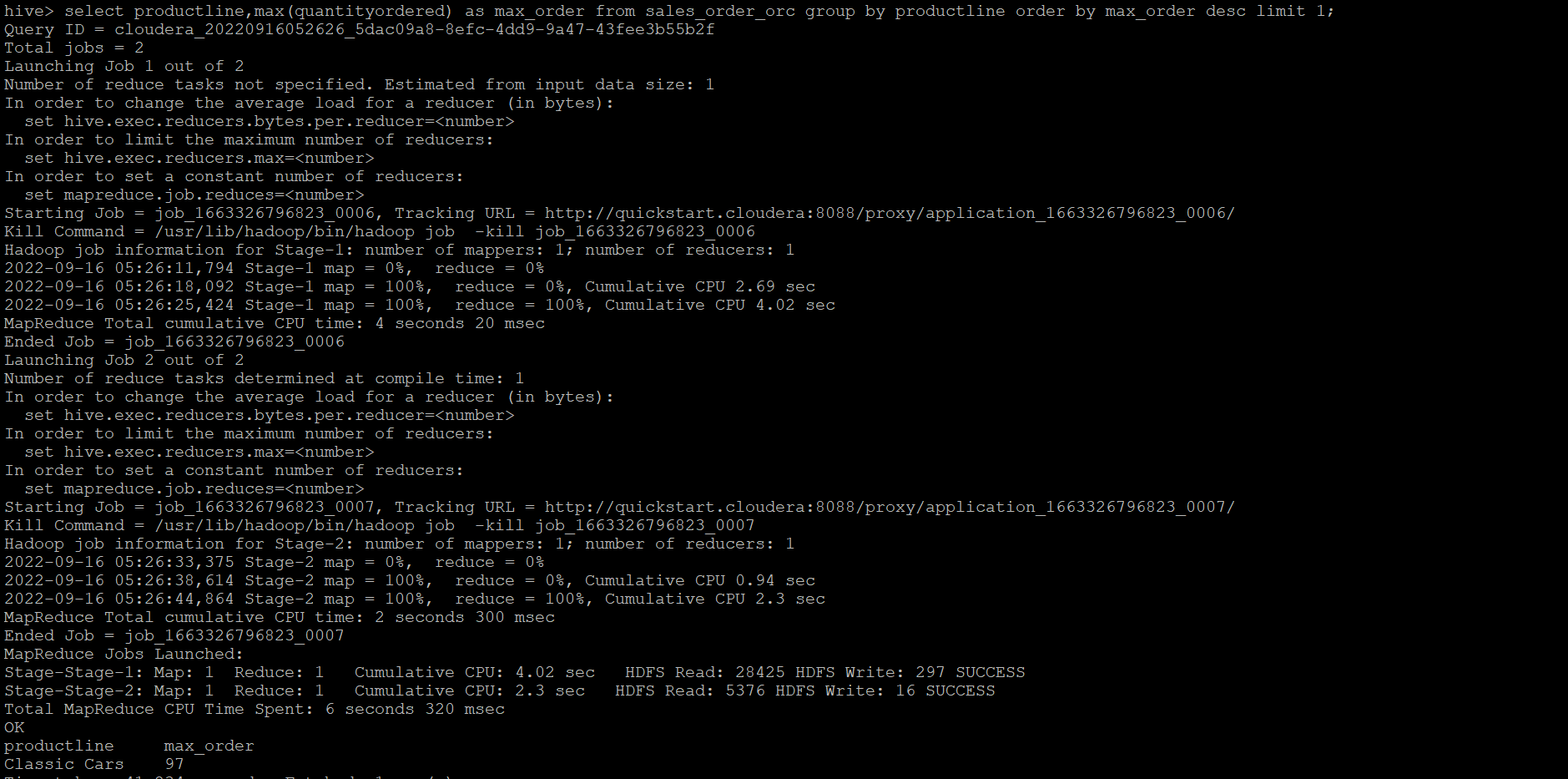
* **from sales\_order\_csv insert overwrite table sales\_order\_orc select \*;**

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

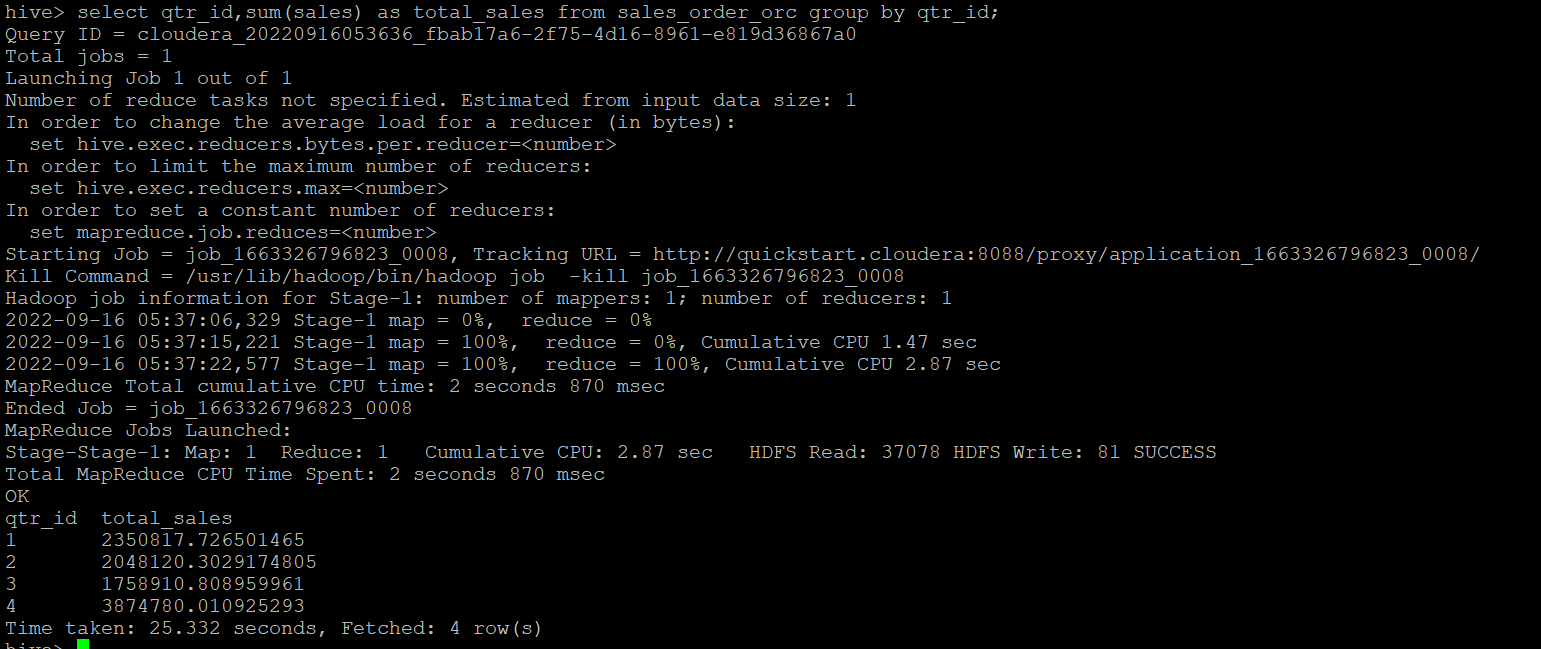
a. Calculate total sales per year



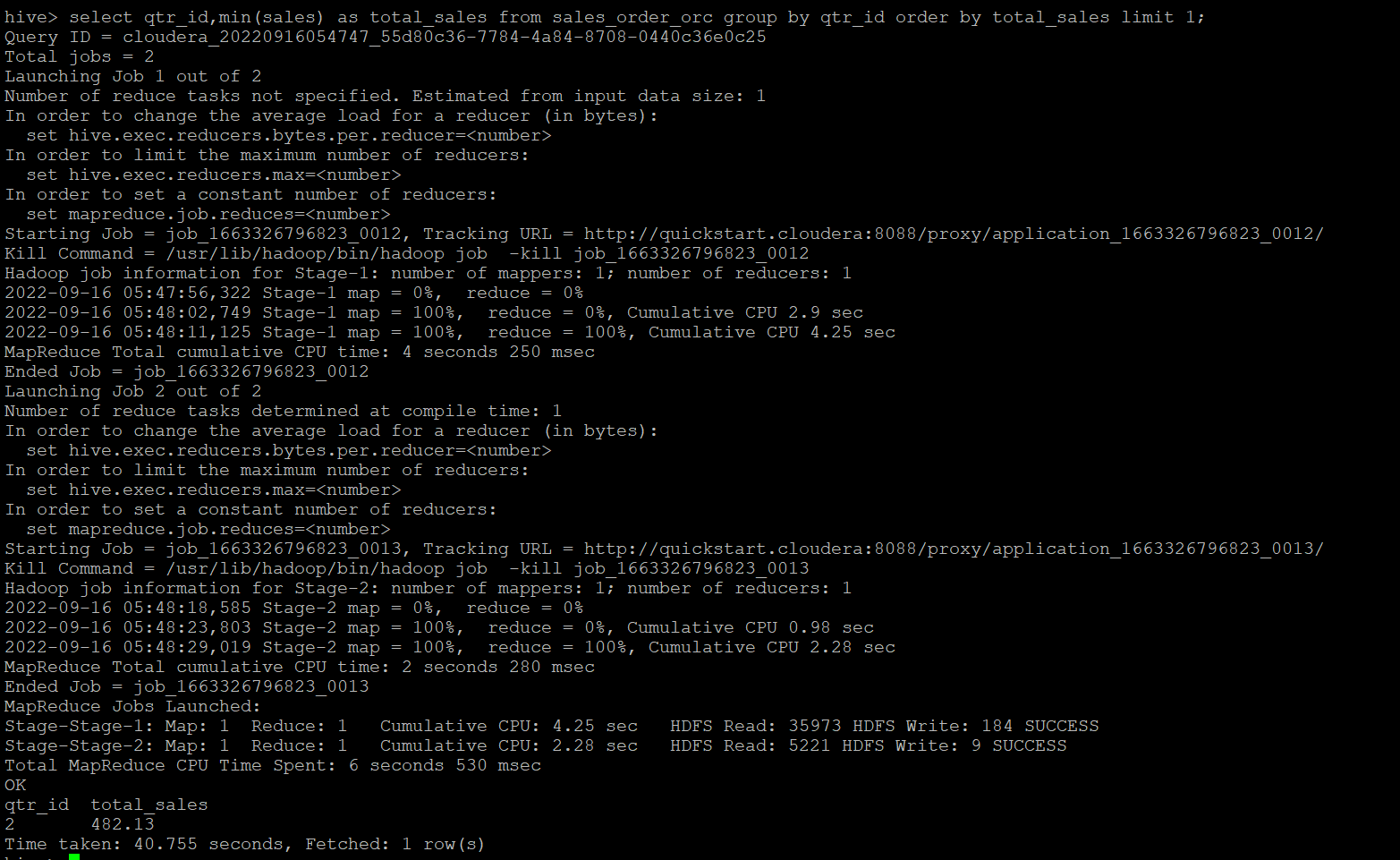
b.Find a product for which maximum orders were placed



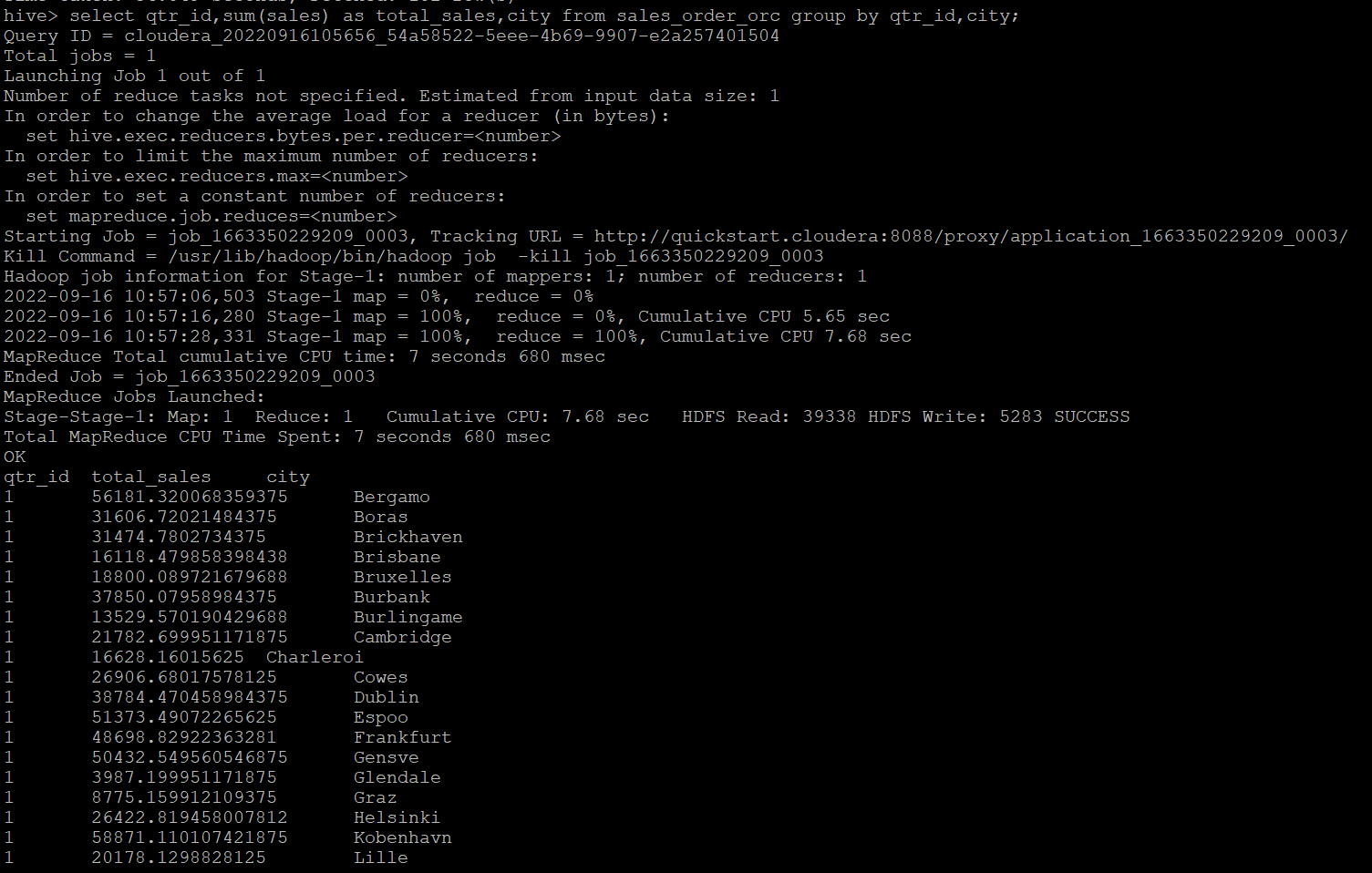
c. Calculate the total sales for each quarter



d. In which quarter sales was minimum



f. Calculate quartelry sales for each city



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

