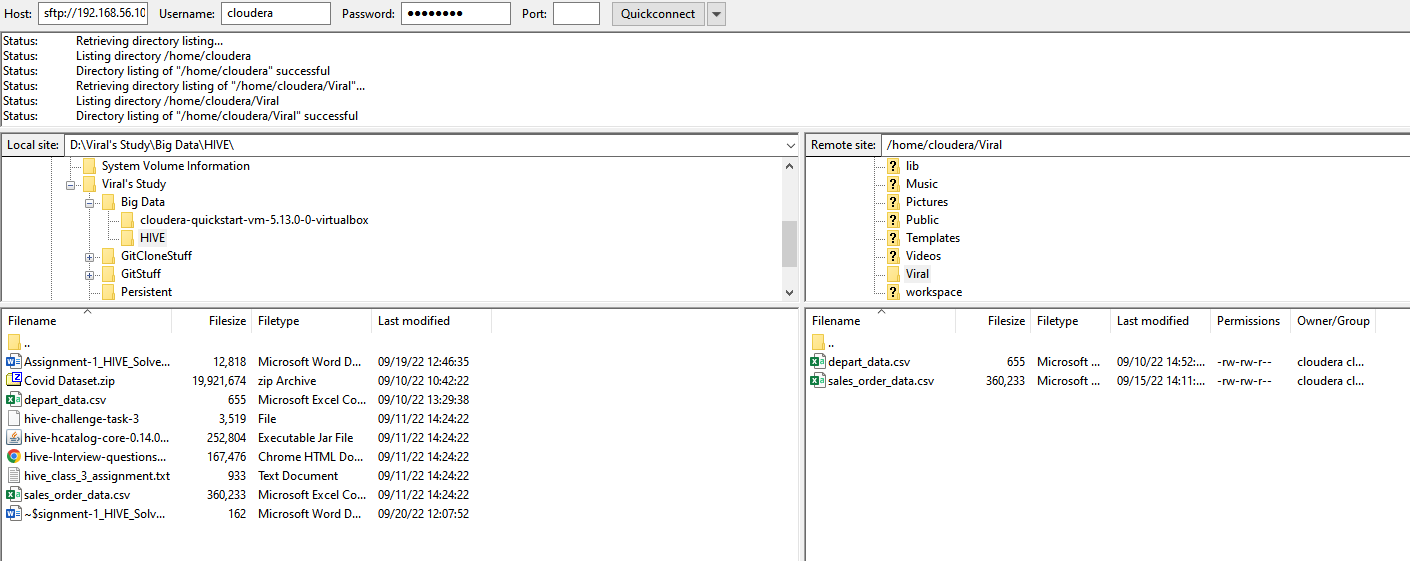
1. **Download vehicle sales data ->** [**https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv**](https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv)

**Answer** – Downloaded csv file and copied from local windows machine to Cloudera Local File System (LFS) path ‘/home/cloudera/Viral’ using FileZilla



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

**Answer** – CSV data from Local Fie System is copied to HDFS path ‘/Viral/’ using below syntax

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

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**

**Answer** – Created Internal table "sales\_order\_csv" in database “hive\_class\_b1” using below commands inside hive:

USE hive\_class\_b1;

create table sales\_order\_ csv

(

ORDERNUMBER int,

QUANTITYORDERED int,

PRICEEACH float,

ORDERLINENUMBER int,

SALES float,

STATUS string,

QTR\_ID int,

MONTH\_ID int,

YEAR\_ID int,

PRODUCTLINE string,

MSRP int,

PRODUCTCODE string,

PHONE string,

CITY string,

STATE string,

POSTALCODE string,

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

row format delimited

fields terminated by ','

tblproperties("skip.header.line.count"="1");

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

**Answer** – Loaded data from csv table present in HDFS location “/Viral/” using below syntax

Load data inpath “/Viral/sales\_order\_data.csv” INTO TABLE sales\_order\_csv;

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

**Answer** – Created Internal table "sales\_order\_orc" in database “hive\_class\_b1” using below commands inside hive:

USE hive\_class\_b1;

create table sales\_order\_orc

(

ORDERNUMBER int,

QUANTITYORDERED int,

PRICEEACH float,

ORDERLINENUMBER int,

SALES float,

STATUS string,

QTR\_ID int,

MONTH\_ID int,

YEAR\_ID int,

PRODUCTLINE string,

MSRP int,

PRODUCTCODE string,

PHONE string,

CITY string,

STATE string,

POSTALCODE string,

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

stored as orc;

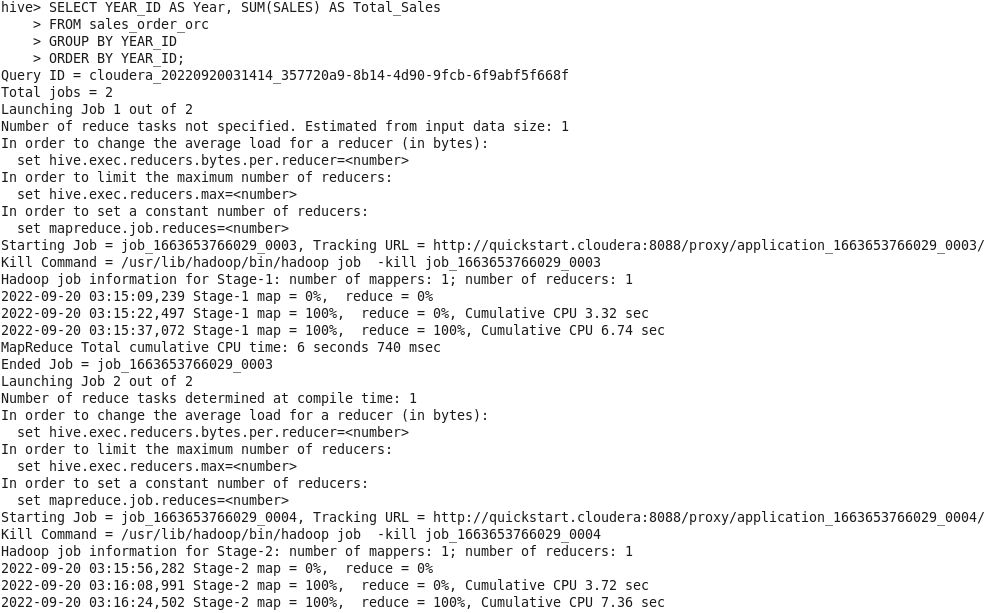
1. **Load data from "sales\_order\_csv" into "sales\_order\_orc"**

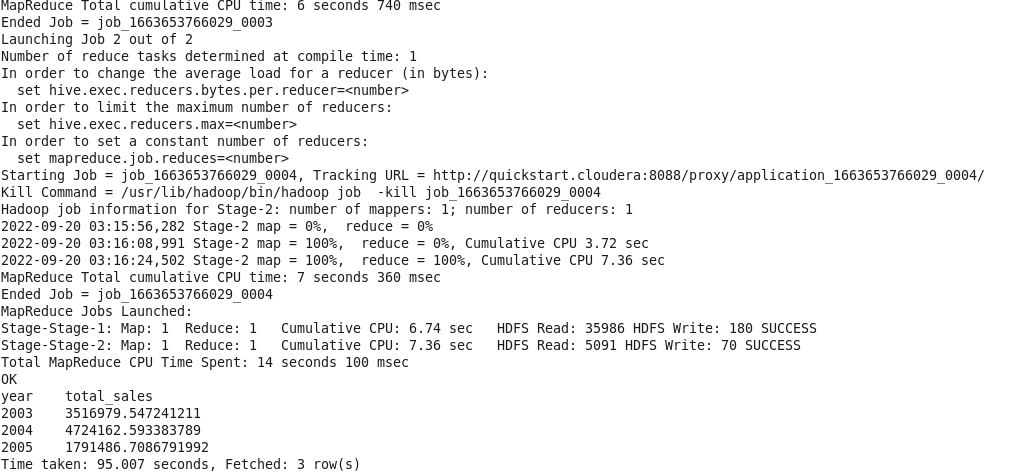
**Answer** – Loaded data from "sales\_order\_csv" into "sales\_order\_orc" using below syntax:

FROM sales\_order\_csv INSERT OVERWRITE TABLE sales\_order\_orc SELECT \*;

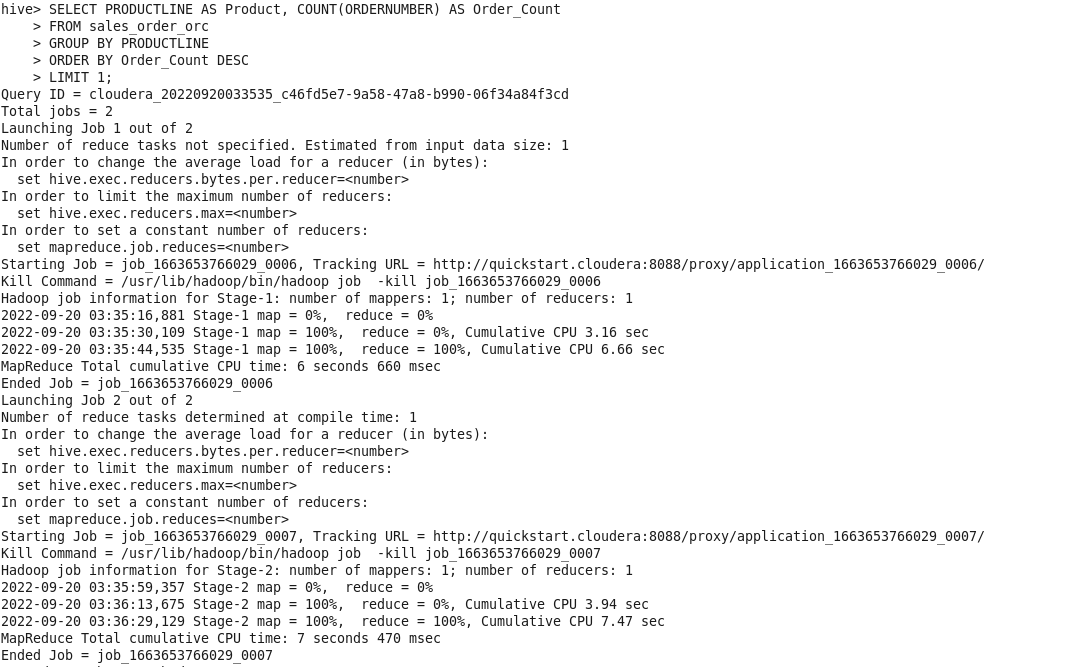
**Perform below mentioned queries on "sales\_order\_orc" table:**

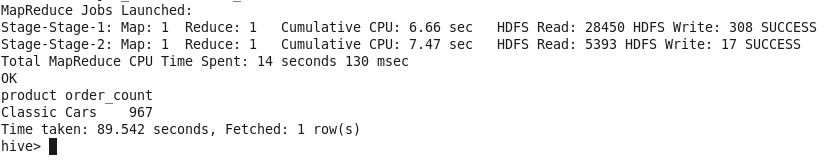
1. **Calculate total sales per year**



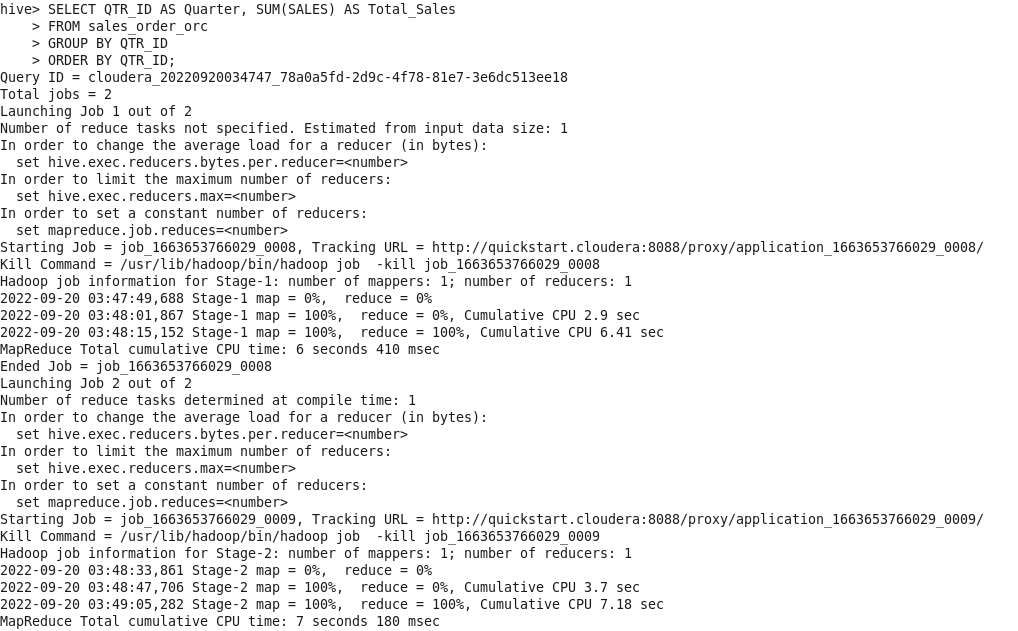


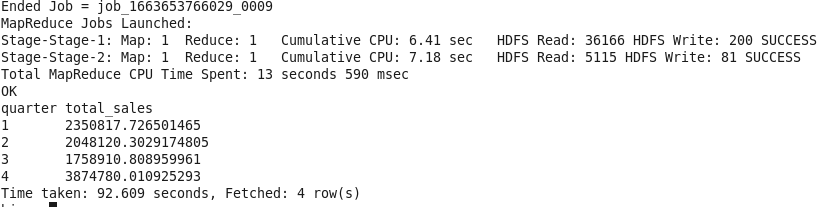
1. **Find a product for which maximum orders were placed**



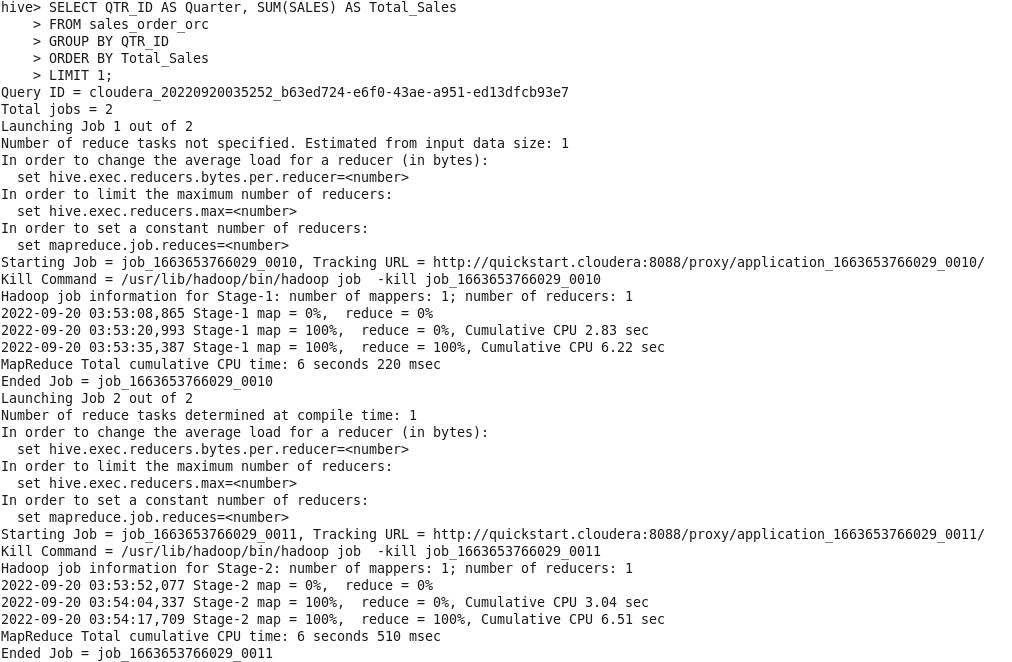


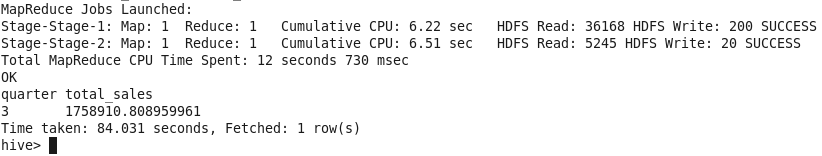
1. **Calculate the total sales for each quarter**





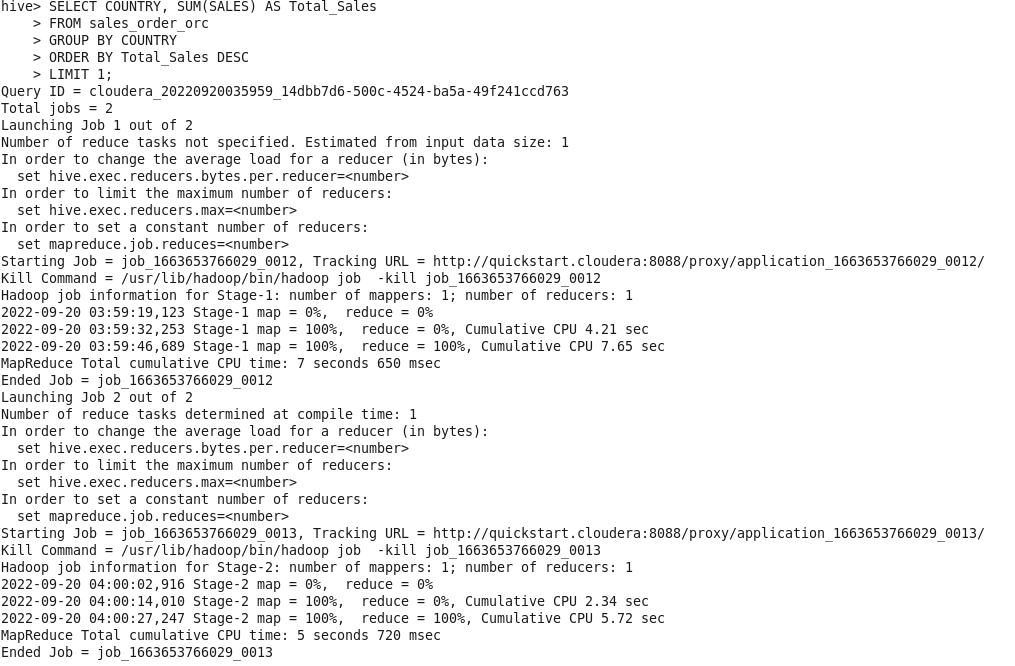
1. **In which quarter sales was minimum**

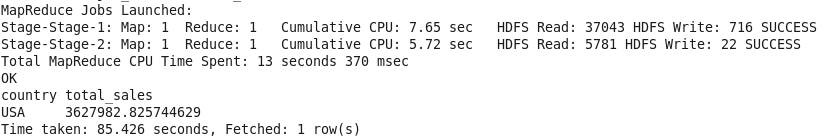




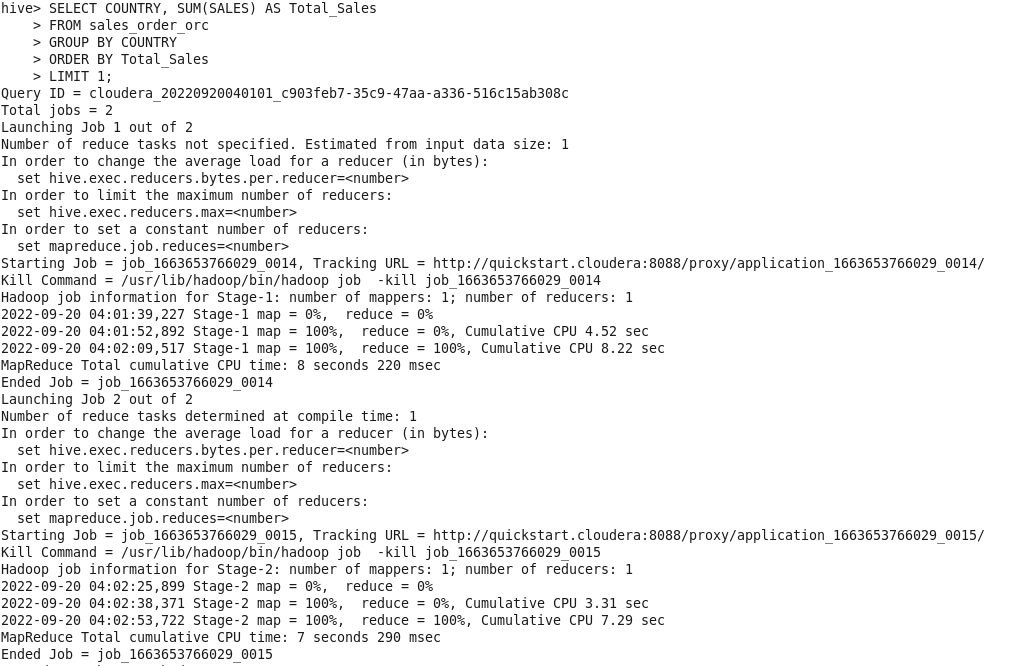
1. **In which country sales was maximum and in which country sales was minimum**

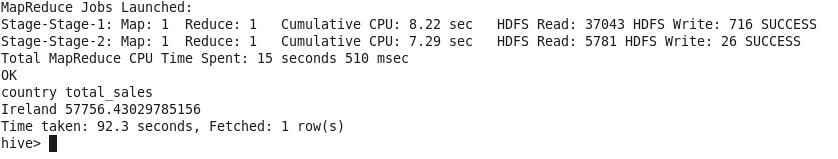
**Country with maximum sales**



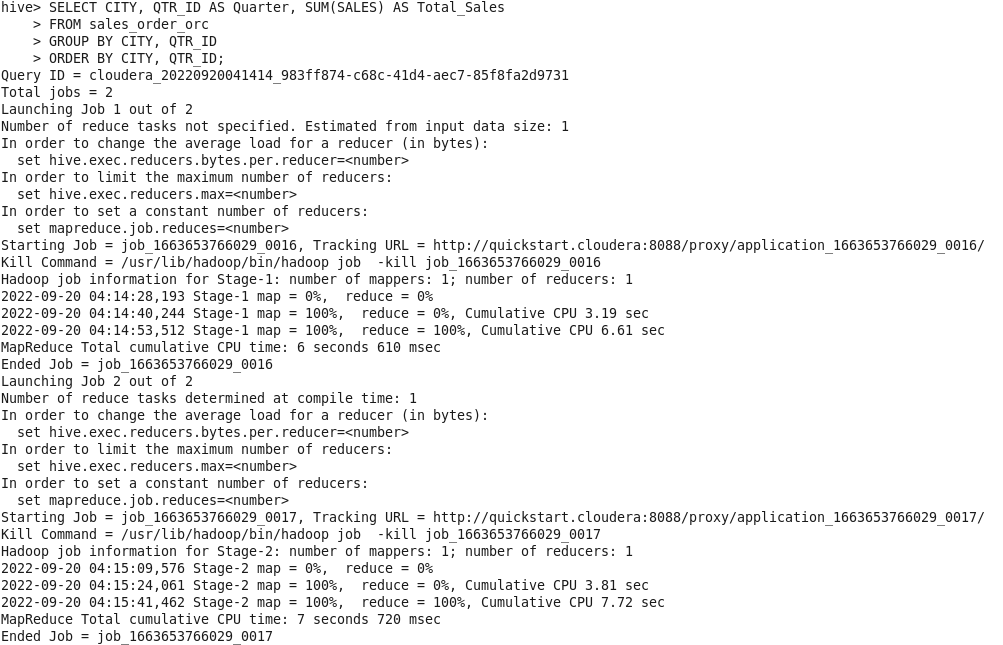


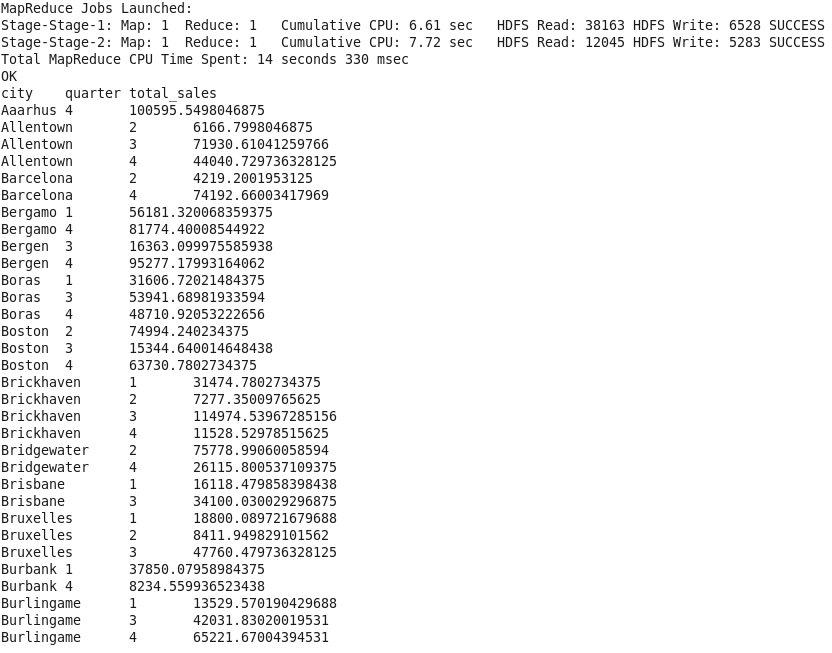
**Country with MINIMUM sales**





1. **Calculate quarterly sales for each city**





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

