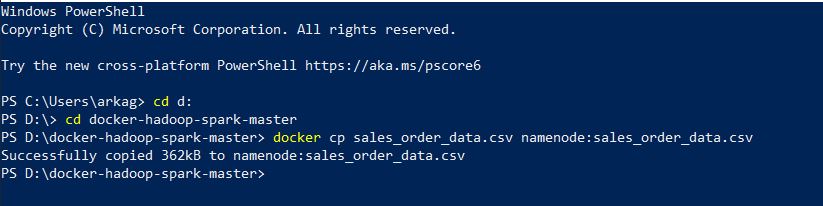
**Hive Assignment1**

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

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

**Step1: Moving data to namenode: docker cp sales\_order\_data.csv namenode:sales\_order\_data.csv**

****

**Step2: Go to the bash shell on the namenode: docker exec -it namenode bash**

**Step3: Creating HDFS directory /data/HiveAssignment1: hdfs dfs -mkdir -p /data/HiveAssignment1**

**Step4: Copy sales\_order\_data.csv from namenode to HDFS: hdfs dfs -put sales\_order\_data.csv /data/HiveAssignment1/sales\_order\_data.csv**

**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**

**Step1: Create database: create database HiveAssignment1;**

**Step2: Create table sales\_order\_data:**

**Use HiveAssignment1;**

**CREATE TABLE sales\_order\_data (**

**ORDERNUMBER int,**

**QUANTITYORDERED int,**

**PRICEEACH decimal,**

**ORDERLINENUMBER int,**

**SALES decimal,**

**STATUS string,**

**QTR\_ID int,**

**MONTH\_ID int,**

**YEAR\_ID int,**

**PRODUCTLINE string,**

**MSRP decimal,**

**PRODUCTCODE string,**

**PHONE string,**

**CITY string,**

**STATE string,**

**POSTALCODE string,**

**COUNTRY string,**

**TERRITORY string,**

**CONTACTLASTNAME string,**

**CONTACTFIRSTNAME string,**

**DEALSIZE string**

**) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'**

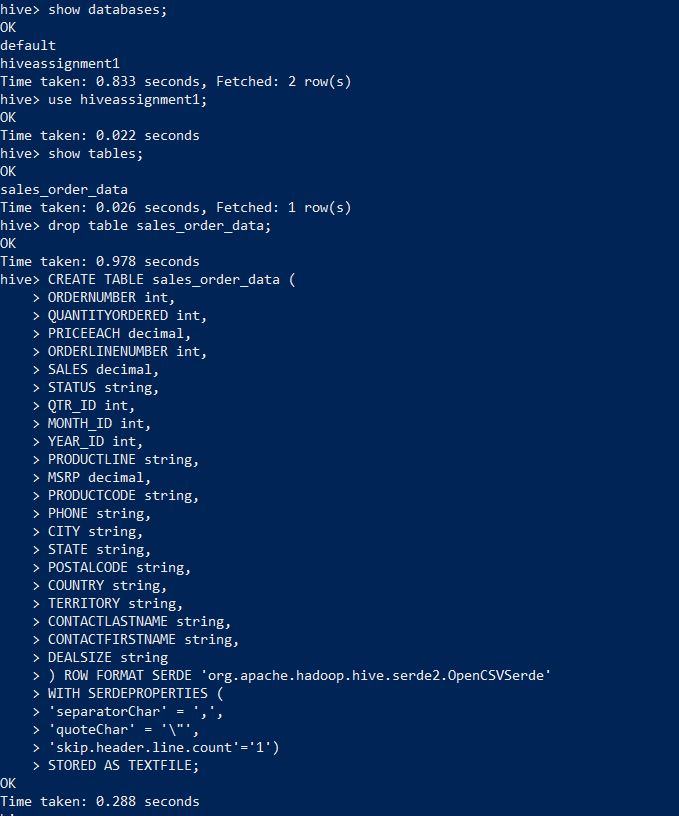
**WITH SERDEPROPERTIES (**

**'separatorChar' = ',',**

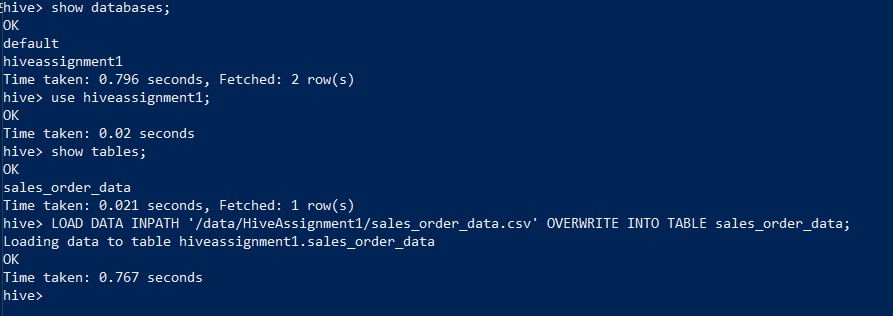
**'quoteChar' = '\"',**

**'skip.header.line.count'='1')**

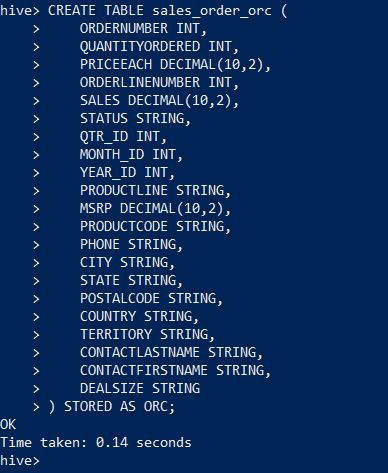
**STORED AS TEXTFILE;**

****

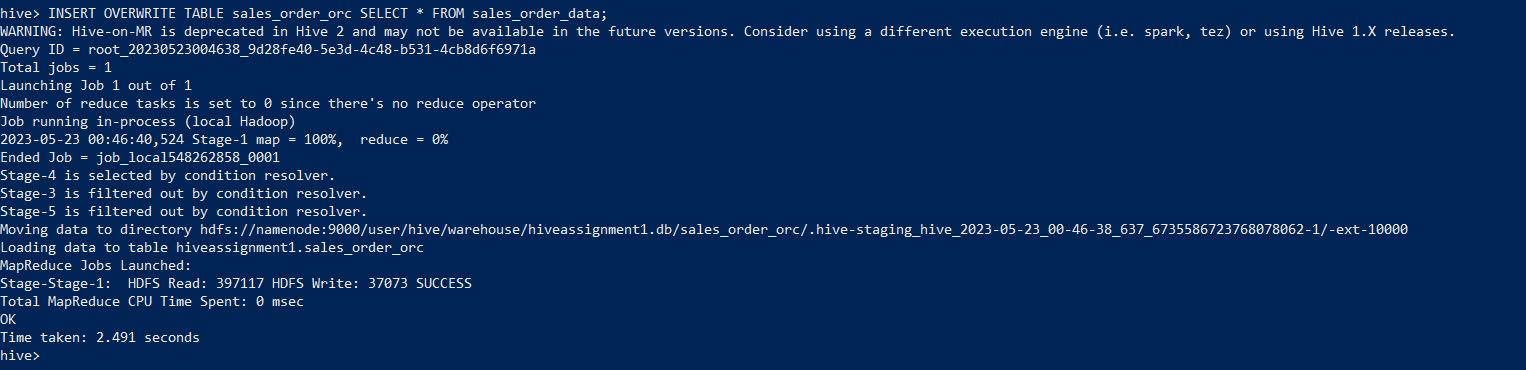
**4. Load data from hdfs path into "sales\_order\_csv"**

****

**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”: LOAD DATA INPATH '/data/HiveAssignment1/sales\_order\_data.csv' OVERWRITE INTO TABLE sales\_order\_data;**

****

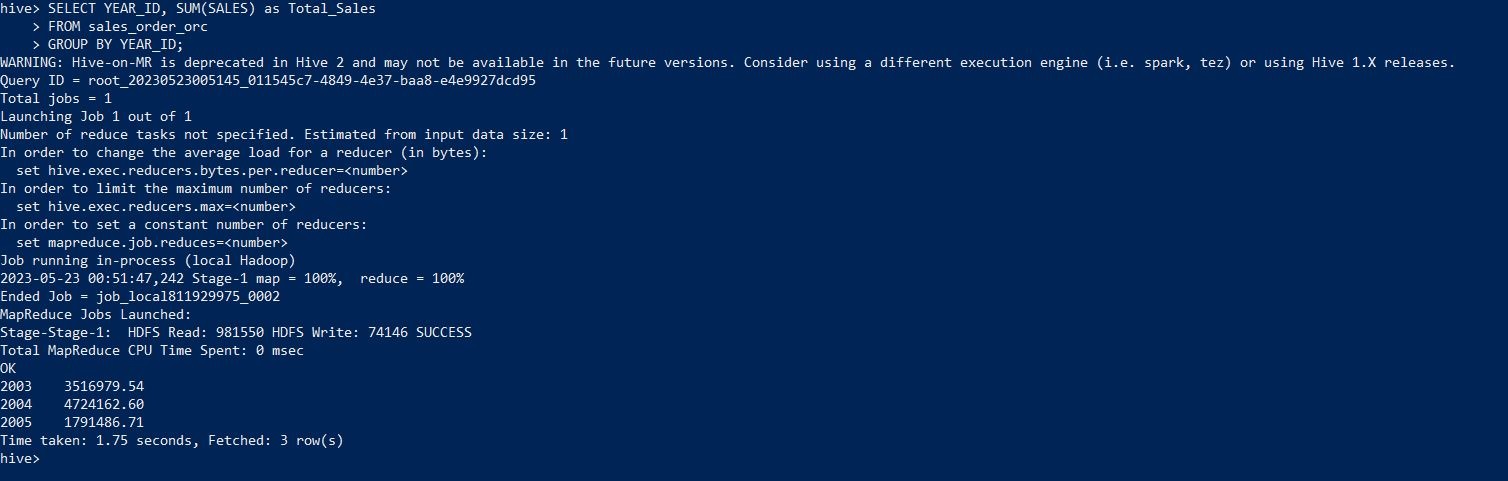
**Perform below menioned queries on "sales\_order\_orc" table:**

1. **Calculatye total sales per year**

**SELECT YEAR\_ID, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY YEAR\_ID;**

****

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

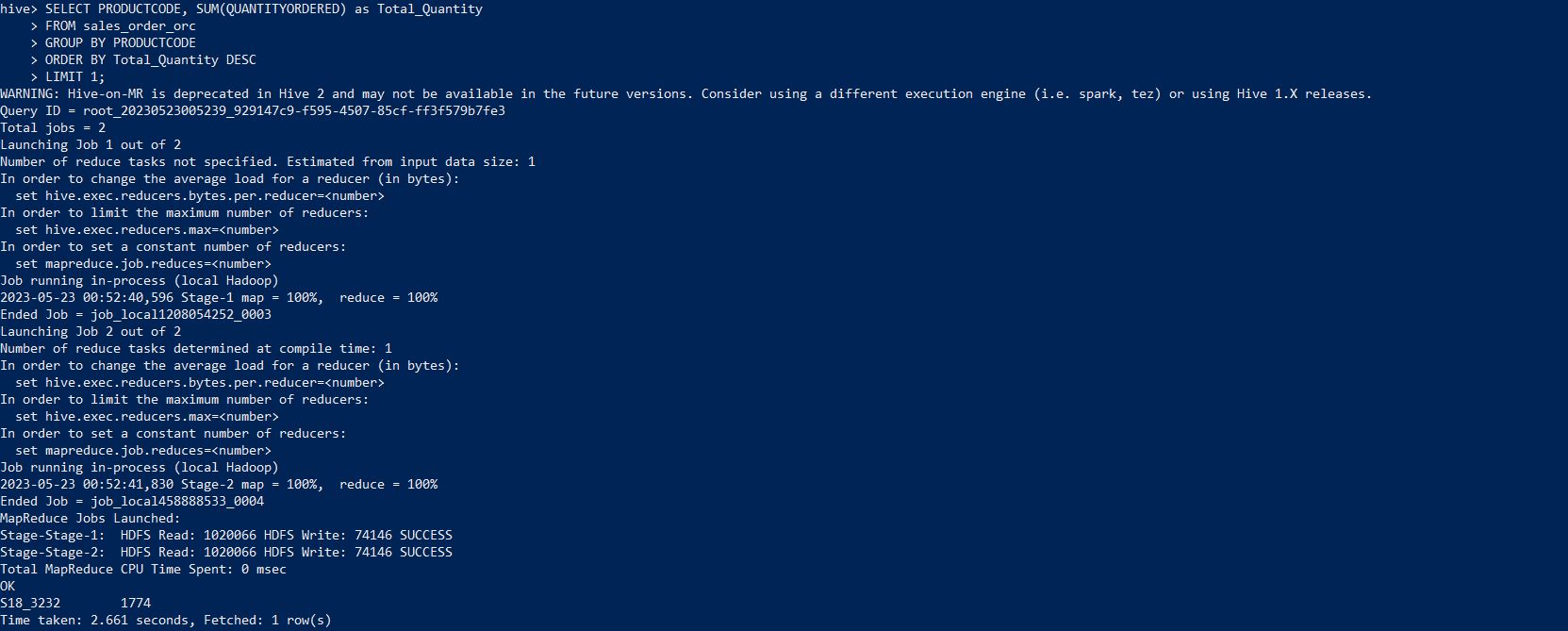
**SELECT PRODUCTCODE, SUM(QUANTITYORDERED) as Total\_Quantity**

**FROM sales\_order\_orc**

**GROUP BY PRODUCTCODE**

**ORDER BY Total\_Quantity DESC**

**LIMIT 1;**

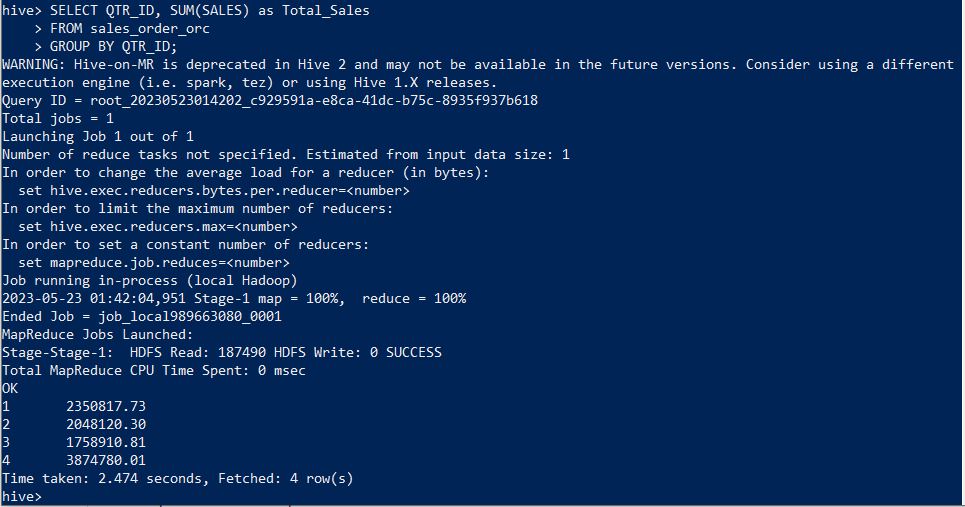
****

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

**SELECT QTR\_ID, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY QTR\_ID;**

****

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

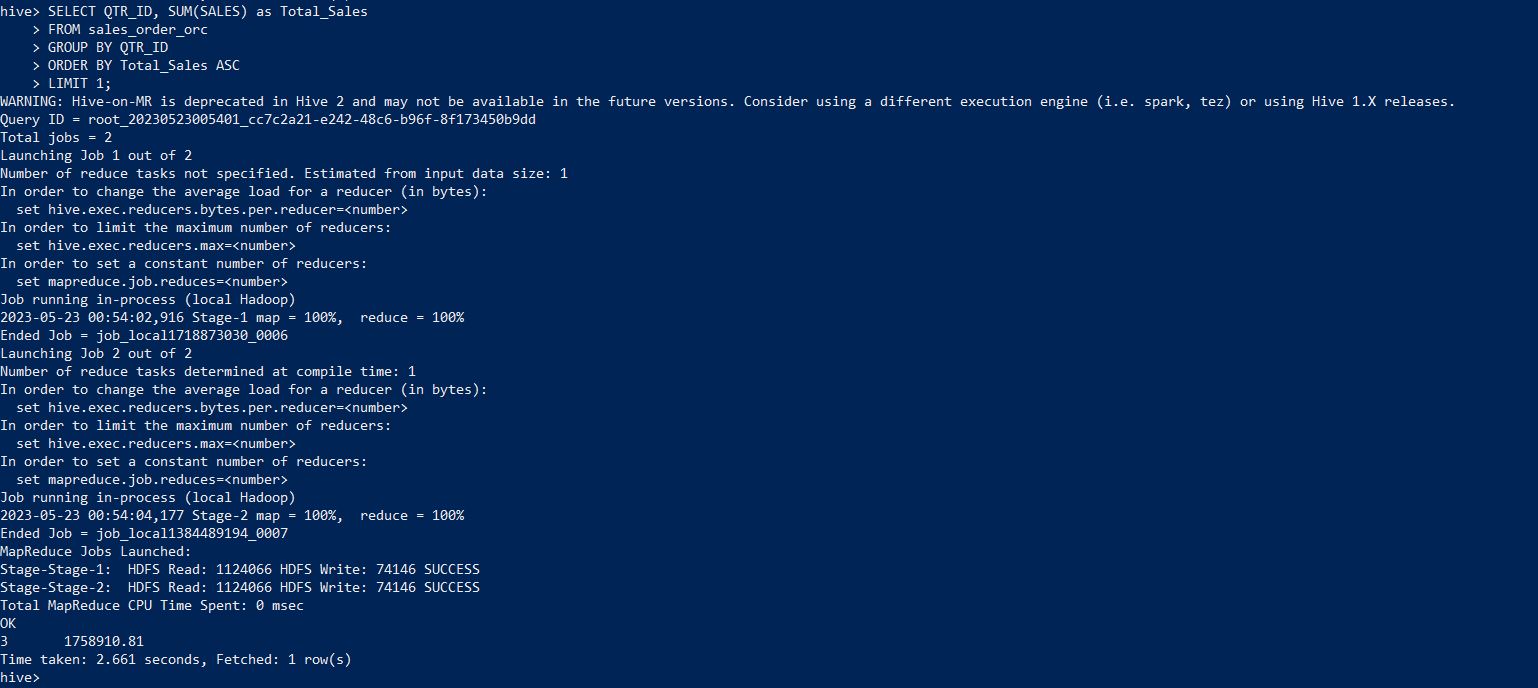
**SELECT QTR\_ID, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY QTR\_ID**

**ORDER BY Total\_Sales ASC**

**LIMIT 1;**

****

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

**-- Maximum sales**

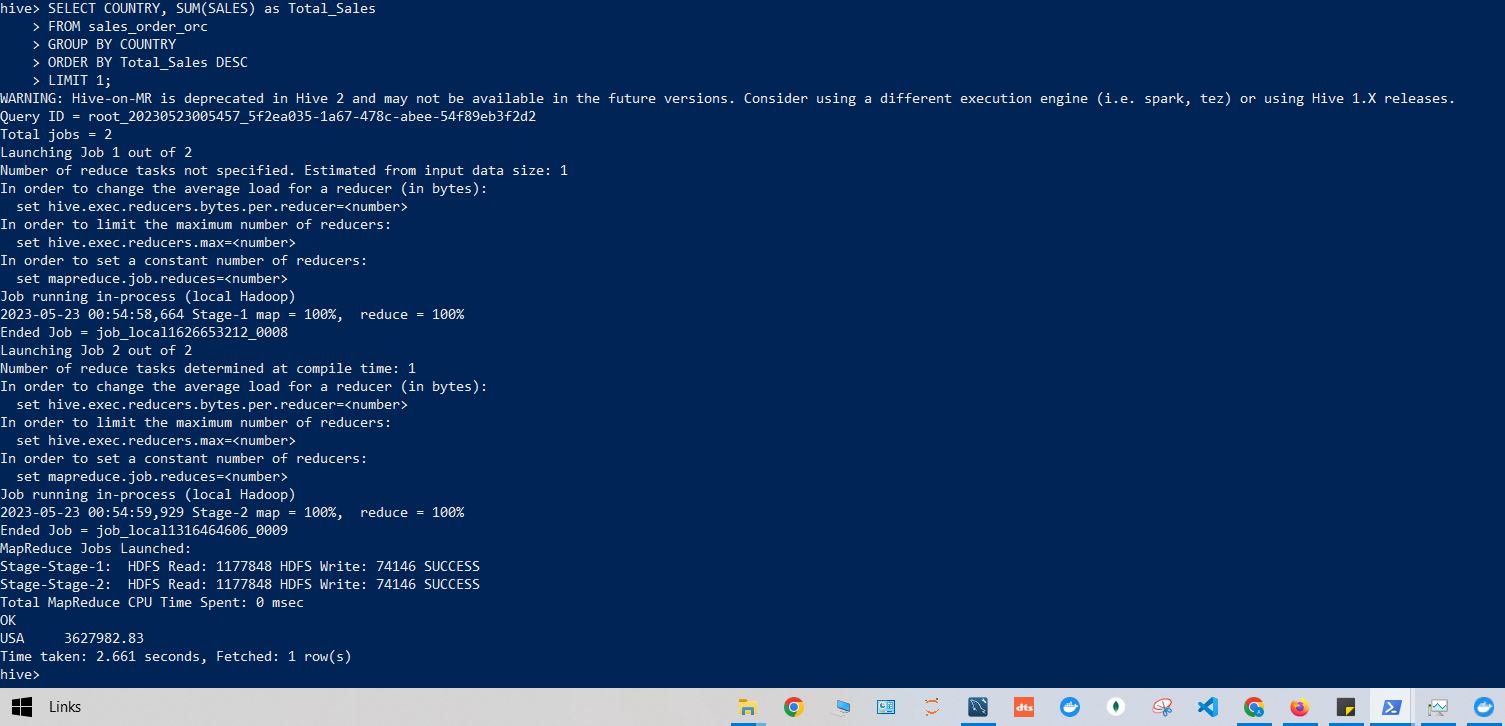
**SELECT COUNTRY, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY COUNTRY**

**ORDER BY Total\_Sales DESC**

**LIMIT 1;**

****

**-- Minimum sales**

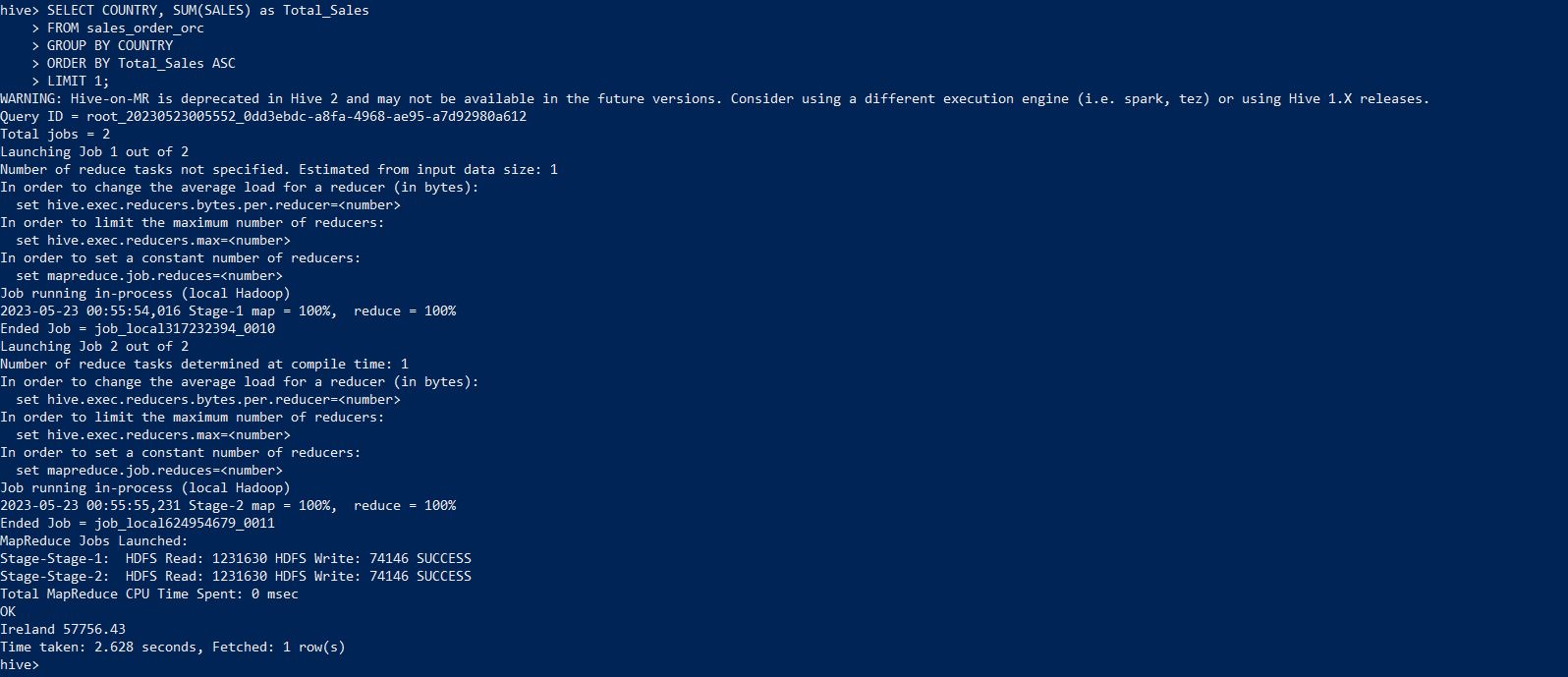
**SELECT COUNTRY, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY COUNTRY**

**ORDER BY Total\_Sales ASC**

**LIMIT 1;**

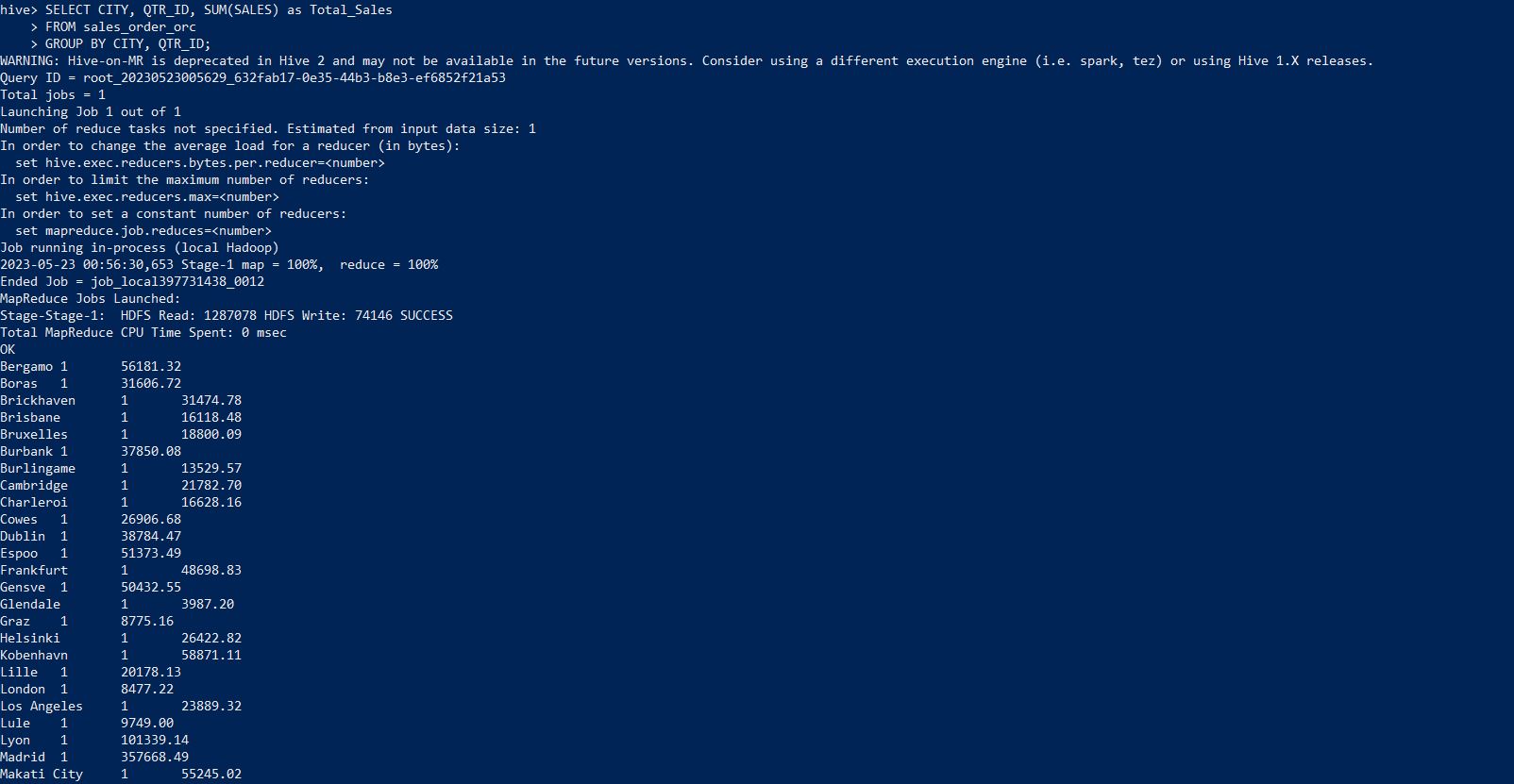
****

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

**SELECT CITY, QTR\_ID, SUM(SALES) as Total\_Sales**

**FROM sales\_order\_orc**

**GROUP BY CITY, QTR\_ID;**

****

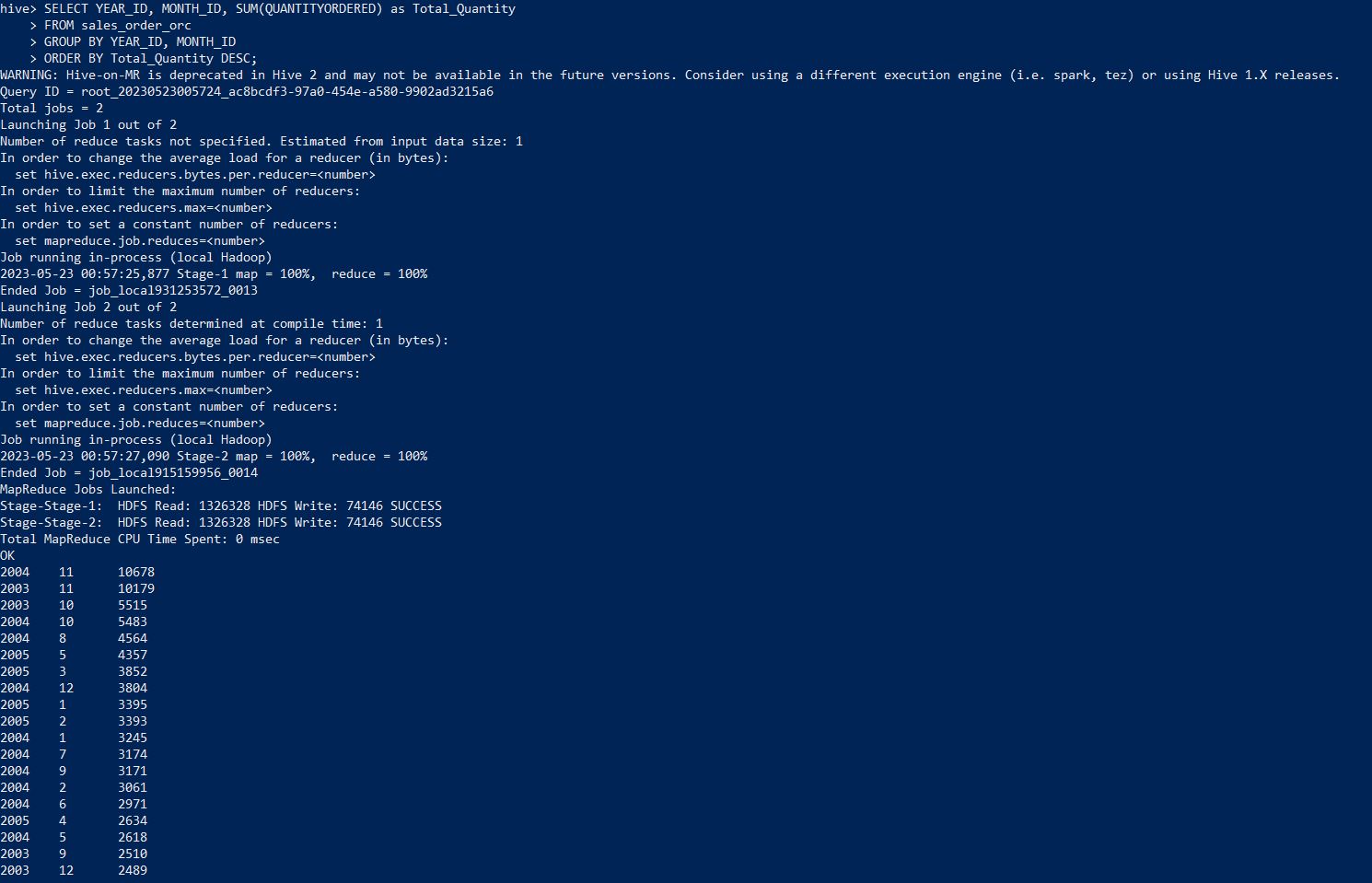
1. **Find a month for each year in which maximum number of quantities were sold**

**SELECT YEAR\_ID, MONTH\_ID, SUM(QUANTITYORDERED) as Total\_Quantity**

**FROM sales\_order\_orc**

**GROUP BY YEAR\_ID, MONTH\_ID**

**ORDER BY Total\_Quantity DESC;**

****