2. Store raw data into hdfs location

cloudera@quickstart:~

```
[cloudera@quickstart ~]$ cd /tmp/hive_class_local_tmp
[cloudera@quickstart hive_class_local_tmp]$ ls
CustomersCSV.csv  dept_csv.txt  SalesOrderData.csv  sales_order_data_csv.txt
[cloudera@quickstart hive_class_local_tmp]$ ls
CustomersCSV.csv  dept_csv.txt  sales_order_data_csv.txt
[cloudera@quickstart hive_class_local_tmp]$ cd
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 15 items
-rw-r--r--   1 cloudera supergroup          0 2022-08-29 04:12 /.autofsck
-rw-r--r--   1 cloudera supergroup          0 2022-08-29 04:12 /.autorelabel
---xr--r--   1 cloudera supergroup         59 2022-08-29 04:51 /LocalFile
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - cloudera supergroup          0 2022-08-29 04:12 /bin
drwxr-xr-x   - cloudera supergroup          0 2022-08-29 04:12 /boot
drwxr-xr-x   - cloudera supergroup          0 2022-08-29 04:12 /dev
drwxr-xr-x   - hbase    supergroup          0 2023-01-01 01:51 /hbase
drwxr-xr-x   - cloudera supergroup          0 2022-09-10 09:54 /home
-rw-r--r--   1 cloudera supergroup     360233 2023-01-01 02:01 /sales_order_data_csv.txt
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
-rw-r--r--   1 cloudera supergroup         26 2022-12-27 02:41 /test.txt
drwxrwxrwt   - hdfs     supergroup          0 2022-12-27 23:44 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use hive_class_b1;
OK
Time taken: 0.489 seconds
hive> create table sales_order_data
    > (
    > ORDERNUMBER INT,
    > QUANTITYORDERED INT,
    > PRICEEACH FLOAT,
    > ORDERLINENUMBER INT,
    > SALES FLOAT,
    > STATUS VARCHAR(100),
    > QTR_ID INT,
    > MONTH_ID INT,
    > YEAR_ID INT,
    > PRODUCTLINE VARCHAR(100),
    > MSRP INT,
    > PRODUCTCODE VARCHAR(100),
    > PHONE VARCHAR(100),
    > CITY VARCHAR(100),
    > STATE VARCHAR(100),
    > POSTALCODE VARCHAR(100),
    > COUNTRY VARCHAR(100),
    > TERRITORY VARCHAR(100),
    > CONTACTLASTNAME VARCHAR(100),
    > CONTACTFIRSTNAME VARCHAR(100),
    > DEALSIZE VARCHAR(100) )
    > row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
    > tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.642 seconds
hive> load data inpath '/sales_order_data_csv.txt' into table sales_order_data
    > ;
Loading data to table hive_class_b1.sales_order_data
Table hive_class_b1.sales_order_data stats: [numFiles=1, totalSize=360233]
OK
Time taken: 1.116 seconds
hive>
```

4. Load data from hdfs path into "sales_order_csv"

```
hive> load data inpath '/sales_order_data_csv.txt' into table sales_order_data
    > ;
Loading data to table hive_class_b1.sales_order_data
Table hive_class_b1.sales_order_data stats: [numFiles=1, totalSize=360233]
OK
Time taken: 1.116 seconds
hive> set hive.cli.print.header = true;
hive> select * from sales_order_data limit 10;
OK
sales_order_data.ordernumber    sales_order_data.quantityordered        sales_order_data.priceeach      sales_order_data.orderlinenumber        sales_order_da
ta.sales        sales_order_data.status sales_order_data.qtr_id sales_order_data.month_id       sales_order_data.year_id        sales_order_data.productline s
ales_order_data.msrp    sales_order_data.productcode    sales_order_data.phone  sales_order_data.city   sales_order_data.state  sales_order_data.postalcode  s
ales_order_data.country sales_order_data.territory      sales_order_data.contactlastname        sales_order_data.contactfirstname       sales_order_data.deals
ize
10107   30      95.7    2       2871    Shipped 1       2       2003    Motorcycles     95      S10_1678        2125557818      NYC     NY      10022   USA  N
A       Yu      Kwai    Small
10121   34      81.35   5       2765.9  Shipped 2       5       2003    Motorcycles     95      S10_1678        26.47.1555      Reims           51100   France
EMEA    Henriot Paul    Small
10134   41      94.74   2       3884.34 Shipped 3       7       2003    Motorcycles     95      S10_1678        +33 1 46 62 7555        Paris           75508F
rance   EMEA    Da Cunha        Daniel  Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95      S10_1678        6265557265      Pasadena        CA      90003U
SA      NA      Young   Julie   Medium
10159   49      100     14      5205.27 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505551386      San Francisco   CA           U
SA      NA      Brown   Julie   Medium
10168   36      96.66   1       3479.76 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505556809      Burlingame      CA      94217U
SA      NA      Hirano  Juri    Medium
10180   29      86.13   9       2497.77 Shipped 4       11      2003    Motorcycles     95      S10_1678        20.16.1555      Lille           59000   France
EMEA    Rance   Martine Small
10188   48      100     1       5512.32 Shipped 4       11      2003    Motorcycles     95      S10_1678        +47 2267 3215   Bergen          N 5804  Norway
EMEA    Oeztan  Veysel  Medium
10201   22      98.57   2       2168.54 Shipped 4       12      2003    Motorcycles     95      S10_1678        6505555787      San Francisco   CA           U
SA      NA      Murphy  Julie   Small
10211   41      100     14      4708.44 Shipped 1       1       2004    Motorcycles     95      S10_1678        (1) 47.55.6555  Paris           75016   France
EMEA    Perrier Dominique       Medium
Time taken: 0.891 seconds, Fetched: 10 row(s)
hive>
```

5. Create an internal hive table which will store data in ORC format "sales_order_orc"

```
hive> create table sales_order_orc
    > (
    > ORDERNUMBER INT,
    > QUANTITYORDERED INT,
    > PRICEEACH FLOAT,
    > ORDERLINENUMBER INT,
    > SALES FLOAT,
    > STATUS VARCHAR(100),
    > QTR_ID INT,
    > MONTH_ID INT,
    > YEAR_ID INT,
    > PRODUCTLINE VARCHAR(100),
    > MSRP INT,
    > PRODUCTCODE VARCHAR(100),
    > PHONE VARCHAR(100),
    > CITY VARCHAR(100),
    > STATE VARCHAR(100),
    > POSTALCODE VARCHAR(100),
    > COUNTRY VARCHAR(100),
    > TERRITORY VARCHAR(100),
    > CONTACTLASTNAME VARCHAR(100),
    > CONTACTFIRSTNAME VARCHAR(100),
    > DEALSIZE VARCHAR(100) )
    > STORED AS ORC;
OK
Time taken: 0.208 seconds
```

6. Load data from "sales_order_csv" into "sales_order_orc"

```
hive> INSERT OVERWRITE TABLE sales_order_orc SELECT * FROM sales_order_data;
FAILED: SemanticException [Error 10001]: Line 1:53 Table not found 'sales_order_data'
hive> INSERT OVERWRITE TABLE sales_order_orc SELECT * FROM sales_order_csv;
Query ID = cloudera_20230101022121_756e1219-de83-4a0f-9bca-324a0cef297d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1672566700277_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-01-01 02:22:11,579 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:22:23,646 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.44 sec
MapReduce Total cumulative CPU time: 4 seconds 440 msec
Ended Job = job_1672566700277_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/hive_class_b1.db/sales_order_orc/.hive-staging_hive_2023-01-01_02-21-49_908_43039553705373
96778-1/-ext-10000
Loading data to table hive_class_b1.sales_order_orc
Table hive_class_b1.sales_order_orc stats: [numFiles=1, numRows=2823, totalSize=37553, rawDataSize=3153291]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 4.44 sec   HDFS Read: 368291 HDFS Write: 37645 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 440 msec
OK
_col0   _col1   _col2   _col3   _col4   _col5   _col6   _col7   _col8   _col9   _col10  _col11  _col12  _col13  _col14  _col15  _col16  _col17  _col18  _col19
_col20
Time taken: 37.01 seconds
hive>
```

```
hive> SELECT * FROM sales_order_orc limit 10;
OK
sales_order_orc.ordernumber     sales_order_orc.quantityordered sales_order_orc.priceeach       sales_order_orc.orderlinenumber sales_order_orc.sales   sales_
order_orc.status        sales_order_orc.qtr_id  sales_order_orc.month_id        sales_order_orc.year_id sales_order_orc.productline     sales_order_orc.msrp s
ales_order_orc.productcode      sales_order_orc.phone   sales_order_orc.city    sales_order_orc.state   sales_order_orc.postalcode      sales_order_orc.countr
y       sales_order_orc.territory       sales_order_orc.contactlastname sales_order_orc.contactfirstname        sales_order_orc.dealsize
10107   30      95.7    2       2871.0  Shipped 1       2       2003    Motorcycles     95      S10_1678        2125557818      NYC     NY      10022   USA  N
A       Yu      Kwai    Small
10121   34      81.35   5       2765.9  Shipped 2       5       2003    Motorcycles     95      S10_1678        26.47.1555      Reims           51100   France
EMEA    Henriot Paul    Small
10134   41      94.74   2       3884.34 Shipped 3       7       2003    Motorcycles     95      S10_1678        +33 1 46 62 7555        Paris           75508F
rance   EMEA    Da Cunha        Daniel  Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95      S10_1678        6265557265      Pasadena        CA      90003U
SA      NA      Young   Julie   Medium
10159   49      100.0   14      5205.27 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505551386      San Francisco   CA          U
SA      NA      Brown   Julie   Medium
10168   36      96.66   1       3479.76 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505556809      Burlingame      CA      94217U
SA      NA      Hirano  Juri    Medium
10180   29      86.13   9       2497.77 Shipped 4       11      2003    Motorcycles     95      S10_1678        20.16.1555      Lille           59000   France
EMEA    Rance   Martine Small
10188   48      100.0   1       5512.32 Shipped 4       11      2003    Motorcycles     95      S10_1678        +47 2267 3215   Bergen          N 5804  Norway
EMEA    Oeztan  Veysel  Medium
10201   22      98.57   2       2168.54 Shipped 4       12      2003    Motorcycles     95      S10_1678        6505555787      San Francisco   CA          U
SA      NA      Murphy  Julie   Small
10211   41      100.0   14      4708.44 Shipped 1       1       2004    Motorcycles     95      S10_1678        (1) 47.55.6555  Paris           75016   France
EMEA    Perrier Dominique       Medium
Time taken: 0.096 seconds, Fetched: 10 row(s)
hive>
```

Perform below mentioned queries on "sales_order_orc" table :

a. Calculate total sales per year

```
hive> SELECT year_id,sum(sales) FROM sales_order_orc GROUP BY (year_id);
Query ID = cloudera_20230101023030_0e2559c2-5f60-4580-882a-513d7712c6eb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 02:30:26,503 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:30:38,588 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.99 sec
2023-01-01 02:30:47,034 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.01 sec
MapReduce Total cumulative CPU time: 10 seconds 10 msec
Ended Job = job_1672566700277_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.01 sec   HDFS Read: 37312 HDFS Write: 70 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 10 msec
OK
year_id _c1
2003    3516979.547241211
2004    4724162.593383789
2005    1791486.7086791992
Time taken: 33.647 seconds, Fetched: 3 row(s)
hive>
```

b. Find a product for which maximum orders were placed

cloudera@quickstart:~

```
hive> SELECT productline,SUM(quantityordered) AS Max_Orders FROM sales_order_orc GROUP BY (productline) ORDER BY Max_Orders DESC;
Query ID = cloudera_20230101024141_b7349aa5-d9c6-4b30-b689-87a714a3a8df
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 02:41:12,023 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:41:18,523 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.18 sec
2023-01-01 02:41:26,929 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.07 sec
MapReduce Total cumulative CPU time: 4 seconds 70 msec
Ended Job = job_1672566700277_0004
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-01 02:41:35,793 Stage-2 map = 0%,  reduce = 0%
2023-01-01 02:41:43,383 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.31 sec
2023-01-01 02:41:50,793 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1672566700277_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.07 sec   HDFS Read: 28919 HDFS Write: 311 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.29 sec   HDFS Read: 5445 HDFS Write: 115 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 360 msec
OK
productline     max_orders
Classic Cars    33992
Vintage Cars    21069
Motorcycles     11663
Trucks and Buses        10777
```

```
productline      max_orders
Classic Cars     33992
Vintage Cars     21069
Motorcycles      11663
Trucks and Buses          10777
Planes  10727
Ships   8127
Trains  2712
Time taken: 49.727 seconds, Fetched: 7 row(s)
hive>
```

C. Calculate the total sales for each quarter

```
hive> SELECT year_id,SUM(sales) AS Total_sales_Q1 FROM sales_order_orc WHERE month_id in (1,2,3) GROUP BY year_id;
Query ID = cloudera_20230101025555_b6837dcc-f30d-4e10-b509-fe29db6a6343
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 02:55:28,199 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:55:37,642 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.39 sec
2023-01-01 02:55:46,044 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.39 sec
MapReduce Total cumulative CPU time: 4 seconds 390 msec
Ended Job = job_1672566700277_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.39 sec   HDFS Read: 38854 HDFS Write: 70 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 390 msec
OK
year_id total_sales_q1
2003    445094.6897583008
2004    833730.6786499023
2005    1071992.3580932617
Time taken: 26.433 seconds, Fetched: 3 row(s)
hive>
```

d. In which quarter sales was minimum

```
hive> SELECT year_id,SUM(sales) AS Total_sales_Q2 FROM sales_order_orc WHERE month_id in (4,5,6) GROUP BY year_id;
Query ID = cloudera_20230101025757_ccb705c3-37f4-4b5e-a4a1-dfbf201afd85
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 02:57:48,386 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:57:57,013 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.13 sec
2023-01-01 02:58:05,414 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.15 sec
MapReduce Total cumulative CPU time: 5 seconds 150 msec
Ended Job = job_1672566700277_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.15 sec   HDFS Read: 38861 HDFS Write: 69 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 150 msec
OK
year_id total_sales_q2
2003    562365.2218017578
2004    766260.7305297852
2005    719494.3505859375
Time taken: 25.808 seconds, Fetched: 3 row(s)
hive>
```

```
hive> SELECT year_id,SUM(sales) AS Total_sales_Q3 FROM sales_order_orc WHERE month_id in (7,8,9) GROUP BY year_id;
Query ID = cloudera_20230101025858_63c22df2-0395-4152-868d-76f82cefbfd6
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 02:59:03,902 Stage-1 map = 0%,  reduce = 0%
2023-01-01 02:59:12,454 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.12 sec
2023-01-01 02:59:20,863 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.14 sec
MapReduce Total cumulative CPU time: 5 seconds 140 msec
Ended Job = job_1672566700277_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.14 sec   HDFS Read: 38861 HDFS Write: 47 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 140 msec
OK
year_id total_sales_q3
2003    649514.5415039062
2004    1109396.2674560547
Time taken: 25.434 seconds, Fetched: 2 row(s)
hive>
```

```
hive> SELECT year_id,SUM(sales) AS Total_sales_Q4 FROM sales_order_orc WHERE month_id in (10,11,12) GROUP BY year_id;
Query ID = cloudera_20230101030000_64c8df10-d6ed-43ef-bea2-387b2d7302d5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 03:00:19,375 Stage-1 map = 0%,  reduce = 0%
2023-01-01 03:00:26,745 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.79 sec
2023-01-01 03:00:35,133 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.78 sec
MapReduce Total cumulative CPU time: 4 seconds 780 msec
Ended Job = job_1672566700277_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.78 sec   HDFS Read: 38861 HDFS Write: 47 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 780 msec
OK
year_id total_sales_q4
2003    1860005.094177246
2004    2014774.9167480469
Time taken: 24.483 seconds, Fetched: 2 row(s)
hive>
```

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

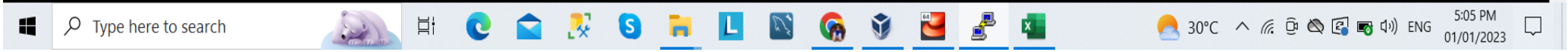
cloudera@quickstart:~

```
hive> SELECT country,MAX(sales) AS Max_sales,MIN(sales) AS Min_sales FROM sales_order_orc GROUP BY (country);
Query ID = cloudera_20230101033030_fad3a324-145a-46d5-9b3f-47424e3d743f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 03:30:26,148 Stage-1 map = 0%,  reduce = 0%
2023-01-01 03:30:41,396 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.7 sec
2023-01-01 03:30:55,625 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.17 sec
MapReduce Total cumulative CPU time: 11 seconds 170 msec
Ended Job = job_1672566700277_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.17 sec   HDFS Read: 38830 HDFS Write: 428 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 170 msec
OK
country max_sales       min_sales
Australia       9774.03 652.35
Austria 9240.0  640.05
Belgium 6804.63 881.4
Canada  9064.89 1119.93
Denmark 10468.9 1146.5
Finland 10606.2 891.03
France  11739.7 482.13
Germany 8940.96 948.99
Ireland 8258.0  1056.4
Italy   9160.36 577.6
Japan   10758.0 553.95
Norway  8844.12 1129.04
Philippines     7483.98 1173.15
Singapore       10993.5 785.64
Spain   12001.0 683.8
Sweden  7209.11 1467.48
Switzerland     6761.6  1205.04
UK      11886.6 710.2
USA     14082.8 541.14
Time taken: 45.904 seconds, Fetched: 19 row(s)
hive>
```

f. Calculate quarterly sales for each city

cloudera@quickstart:~

```
hive> SELECT year_id,SUM(sales) AS Total_sales_Q1,city FROM sales_order_orc WHERE month_id in (1,2,3) GROUP BY year_id,city ;
Query ID = cloudera_20230101033333_2a3af61d-8418-4c53-b3ac-526c19d7b91a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 03:34:07,516 Stage-1 map = 0%,  reduce = 0%
2023-01-01 03:34:24,751 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.48 sec
2023-01-01 03:34:39,018 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.9 sec
MapReduce Total cumulative CPU time: 11 seconds 900 msec
Ended Job = job_1672566700277_0017
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.9 sec   HDFS Read: 41155 HDFS Write: 1809 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 900 msec
OK
year_id total_sales_q1  city
2003    56181.320068359375      Bergamo
2003    11432.33984375  Frankfurt
2003    58871.110107421875      Kobenhavn
2003    9748.999755859375       Lule
2003    44621.96008300781       Madrid
2003    55245.02014160156       Makati City
2003    51017.919860839844      Manchester
2003    32647.809814453125      NYC
2003    12133.25        Nashua
2003    27398.820434570312      Philadelphia
2003    18695.579833984375      San Francisco
2003    12398.56005859375       San Rafael
2003    54701.999755859375      Stavern
2004    16118.479858398438      Brisbane
2004    18800.089721679688      Bruxelles
2004    37850.07958984375       Burbank
2004    21782.699951171875      Cambridge
2004    26906.68017578125       Cowes
2004    38784.470458984375      Dublin
2004    37266.48937988281       Frankfurt
2004    50432.549560546875      Gensve
```



cloudera@quickstart:~

```
year_id total_sales_q1  city
2003    56181.320068359375      Bergamo
2003    11432.33984375  Frankfurt
2003    58871.110107421875      Kobenhavn
2003    9748.999755859375       Lule
2003    44621.96008300781       Madrid
2003    55245.02014160156       Makati City
2003    51017.919860839844      Manchester
2003    32647.809814453125      NYC
2003    12133.25        Nashua
2003    27398.820434570312      Philadelphia
2003    18695.579833984375      San Francisco
2003    12398.56005859375       San Rafael
2003    54701.999755859375      Stavern
2004    16118.479858398438      Brisbane
2004    18800.089721679688      Bruxelles
2004    37850.07958984375       Burbank
2004    21782.699951171875      Cambridge
2004    26906.68017578125       Cowes
2004    38784.470458984375      Dublin
2004    37266.48937988281       Frankfurt
2004    50432.549560546875      Gensve
2004    20178.1298828125        Lille
2004    8477.219970703125       London
2004    23889.320068359375      Los Angeles
2004    101339.13977050781      Lyon
2004    105491.33990478516      Madrid
2004    49637.57067871094       Melbourne
2004    8722.1201171875 Newark
2004    50490.64013671875       Osaka
2004    51172.649658203125      Paris
2004    87489.23010253906       San Diego
2004    48922.76989746094       San Rafael
2004    24219.58984375  Singapore
2004    5759.419921875  Versailles
2005    31606.72021484375       Boras
2005    31474.7802734375        Brickhaven
2005    13529.570190429688      Burlingame
2005    16628.16015625  Charleroi
2005    51373.49072265625       Espoo
2005    3987.199951171875       Glendale
2005    8775.159912109375       Graz
2005    26422.819458007812      Helsinki
2005    207555.18994140625      Madrid
2005    2317.43994140625        Marseille
```

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

```
cloudera@quickstart:~
hive> SELECT month_id,year_id,SUM(quantityordered)AS Quantity_orders FROM sales_order_orc GROUP BY month_id,year_id ORDER BY Quantity_orders DESC;
Query ID = cloudera_20230101034242_e987db95-5591-402e-a23a-fa9ae6ef998e
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-01 03:43:06,974 Stage-1 map = 0%,  reduce = 0%
2023-01-01 03:43:20,579 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.17 sec
2023-01-01 03:43:35,959 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.24 sec
MapReduce Total cumulative CPU time: 9 seconds 240 msec
Ended Job = job_1672566700277_0019
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1672566700277_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-01 03:43:52,514 Stage-2 map = 0%,  reduce = 0%
2023-01-01 03:44:07,227 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.47 sec
2023-01-01 03:44:22,654 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 8.76 sec
MapReduce Total cumulative CPU time: 8 seconds 760 msec
Ended Job = job_1672566700277_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.24 sec   HDFS Read: 30003 HDFS Write: 792 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 8.76 sec   HDFS Read: 6082 HDFS Write: 356 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 0 msec
OK
month_id        year_id quantity_orders
11      2004    10678
11      2003    10179
10      2003    5515
10      2004    5483
```

cloudera@quickstart:~

```
Starting Job = job_1672566700277_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1672566700277_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1672566700277_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-01 03:43:52,514 Stage-2 map = 0%,  reduce = 0%
2023-01-01 03:44:07,227 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 4.47 sec
2023-01-01 03:44:22,654 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 8.76 sec
MapReduce Total cumulative CPU time: 8 seconds 760 msec
Ended Job = job_1672566700277_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.24 sec   HDFS Read: 30003 HDFS Write: 792 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 8.76 sec   HDFS Read: 6082 HDFS Write: 356 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 0 msec
OK
month_id        year_id quantity_orders
11      2004    10678
11      2003    10179
10      2003    5515
10      2004    5483
8       2004    4564
5       2005    4357
3       2005    3852
12      2004    3804
1       2005    3395
2       2005    3393
1       2004    3245
7       2004    3174
9       2004    3171
2       2004    3061
6       2004    2971
4       2005    2634
5       2004    2618
9       2003    2510
12      2003    2489
4       2004    2077
5       2003    2017
4       2003    1993
3       2004    1978
8       2003    1974
3       2003    1755
7       2003    1725
6       2003    1649
2       2003    1449
1       2003    1357
Time taken: 91.535 seconds, Fetched: 29 row(s)
hive>
```