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:

Query:

hadoop fs -put /home/cloudera/Downloads/sales_order_data.csv /tmp/hiveChallenge1

```
[cloudera@quickstart Downloads]$ hadoop fs -put /home/cloudera/Downloads/sales_order_data.csv /tmp/hiveChallenge1
[cloudera@quickstart Downloads]$ hadoop fs -ls /tmp/hiveChallenge1/
Found 1 items
-rw-r--r--   1 cloudera supergroup     360233 2022-11-05 10:07 /tmp/hiveChallenge1/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:

Query:

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 varchar(30),

PHONE varchar(30),

CITY string,

STATE string,

POSTALCODE varchar(30),

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

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

```
hive> 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 varchar(30),
    > PHONE varchar(30),
    > CITY string,
    > STATE string,
    > POSTALCODE varchar(30),
    > COUNTRY string,
    > TERRITORY string,
    > CONTACTLASTNAME string,
    > CONTACTFIRSTNAME string,
    > DEALSIZE string
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > tblproperties("skip.header.line.count"="1")
    > ;
OK
Time taken: 0.223 seconds
```

4) Load data from hdfs path into "sales_order_csv"

Query:

load data inpath '/tmp/hiveChallenge1/' into table sales_order_csv;

```
hive> load data inpath '/tmp/hiveChallenge1/' into table sales_order_csv;
Loading data to table hive_challenge_1.sales_order_csv
Table hive_challenge_1.sales_order_csv stats: [numFiles=1, totalSize=360233]
OK
Time taken: 0.513 seconds
```

Verifying row count and data:

select count(*) from sales_order_csv;

```
hive> select count(*) from sales_order_csv;
Query ID = cloudera_20221105102222_0a971e18-c78b-454a-8d28-19f66b51e89a
Total jobs = 1
Launching Job 1 out of 1
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_1667665836887_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 10:22:17,620 Stage-1 map = 0%,  reduce = 0%
2022-11-05 10:22:30,853 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.4 sec
2022-11-05 10:22:46,249 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.74 sec
MapReduce Total cumulative CPU time: 4 seconds 740 msec
Ended Job = job_1667665836887_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.74 sec   HDFS Read: 370561 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 740 msec
OK
2823
Time taken: 46.452 seconds, Fetched: 1 row(s)
```

select * from sales_order_csv limit 5;

```
hive> select * from sales_order_csv limit 5;
OK
10107   30      95.7    2       2871.0  Shipped 1       2       2003    Motorcycles     95      S10_1678        2125557818      N
YC      NY      10022   USA     NA      Yu      Kwai    Small
10121   34      81.35   5       2765.9  Shipped 2       5       2003    Motorcycles     95      S10_1678        26.47.1555      R
eims            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           75508   France  EMEA    Da Cunha        Daniel  Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95      S10_1678        6265557265      P
asadena CA      90003   USA     NA      Young   Julie   Medium
10159   49      100.0   14      5205.27 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505551386      S
an Francisco    CA              USA     NA      Brown   Julie   Medium
Time taken: 0.175 seconds, Fetched: 5 row(s)
```

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

Query:

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 varchar(30),
PHONE varchar(30),
CITY string,
STATE string,
POSTALCODE varchar(30),
COUNTRY string,
TERRITORY string,
CONTACTLASTNAME string,
CONTACTFIRSTNAME string,
DEALSIZE string
)
stored as ORC;
```

```
hive> 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 varchar(30),
    > PHONE varchar(30),
    > CITY string,
    > STATE string,
    > POSTALCODE varchar(30),
    > COUNTRY string,
    > TERRITORY string,
    > CONTACTLASTNAME string,
    > CONTACTFIRSTNAME string,
    > DEALSIZE string
    > )
    > stored as ORC;
OK
Time taken: 0.114 seconds
```

6) Load data from "sales_order_csv" into "sales_order_orc":

Query:

from sales_order_csv insert overwrite table sales_order_orc select *;

```
hive> from sales_order_csv insert overwrite table sales_order_orc select *;
Query ID = cloudera_20221105103131_5a4ff341-cdc4-4802-b8da-a3cc7da80767
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_1667665836887_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-11-05 10:31:40,397 Stage-1 map = 0%,  reduce = 0%
2022-11-05 10:31:53,146 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.32 sec
MapReduce Total cumulative CPU time: 3 seconds 320 msec
Ended Job = job_1667665836887_0003
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_challenge_1.db/sales_order_orc/.hive-staging_hive_2022-
11-05_10-31-23_840_108349783662121000-1/-ext-10000
Loading data to table hive_challenge_1.sales_order_orc
Table hive_challenge_1.sales_order_orc stats: [numFiles=1, numRows=2823, totalSize=37557, rawDataSize=3153291]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.32 sec   HDFS Read: 367661 HDFS Write: 37652 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 320 msec
OK
Time taken: 31.207 seconds
```

Verifying the data from ORC table:

select count(*) from sales_order_orc;

```
hive> select count(*) from sales_order_orc;
Query ID = cloudera_20221105103232_2bd1ba4f-315e-46ae-88be-24ff925cc381
Total jobs = 1
Launching Job 1 out of 1
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_1667665836887_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 10:32:30,212 Stage-1 map = 0%,  reduce = 0%
2022-11-05 10:32:43,018 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.99 sec
2022-11-05 10:32:59,230 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1667665836887_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.29 sec   HDFS Read: 27074 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
2823
Time taken: 44.431 seconds, Fetched: 1 row(s)
```

select * from sales_order_orc limit 5;

```
hive> select * from sales_order_orc limit 5;
OK
10107   30      95.7    2       2871.0  Shipped 1       2       2003    Motorcycles     95      S10_1678        2125557818      N
YC      NY      10022   USA     NA      Yu      Kwai    Small
10121   34      81.35   5       2765.9  Shipped 2       5       2003    Motorcycles     95      S10_1678        26.47.1555      R
eims            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           75508   France  EMEA    Da Cunha        Daniel  Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95      S10_1678        6265557265      P
asadena CA      90003   USA     NA      Young   Julie   Medium
10159   49      100.0   14      5205.27 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505551386      S
an Francisco    CA              USA     NA      Brown   Julie   Medium
Time taken: 0.135 seconds, Fetched: 5 row(s)
```

Calculate total sales per year:

Query:

select YEAR_ID as Year,sum(SALES) as Yearly_Sales_Sum from sales_order_orc group by YEAR_ID;

```
hive> select YEAR_ID as Year,sum(SALES) as Yearly_Sales_Sum from sales_order_orc group by YEAR_ID;
Query ID = cloudera_20221105104545_7fd0ff1b-31c7-4960-9b07-e6d7186aa302
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_1667665836887_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 10:45:55,432 Stage-1 map = 0%,  reduce = 0%
2022-11-05 10:46:07,345 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.07 sec
2022-11-05 10:46:22,089 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.47 sec
MapReduce Total cumulative CPU time: 4 seconds 470 msec
Ended Job = job_1667665836887_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.47 sec   HDFS Read: 37288 HDFS Write: 70 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 470 msec
OK
year    yearly_sales_sum
2003    3516979.547241211
2004    4724162.593383789
2005    1791486.7086791992
Time taken: 42.872 seconds, Fetched: 3 row(s)
```

Find a product for which maximum orders were placed:

Query:

select PRODUCTCODE as Product,sum(QUANTITYORDERED) as MaxSales from sales_order_orc group by PRODUCTCODE order by MaxSales desc limit 1;

```
hive> select PRODUCTCODE as Product,sum(QUANTITYORDERED) as MaxSales from sales_order_orc group by PRODUCTCODE order by MaxSal
 desc limit 1;
Query ID = cloudera_20221105113030_57a1c618-e9d5-45c9-9251-7ca53d58450b
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_1667665836887_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0014/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 11:30:20,044 Stage-1 map = 0%,  reduce = 0%
2022-11-05 11:30:35,684 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.64 sec
2022-11-05 11:30:51,515 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.09 sec
MapReduce Total cumulative CPU time: 5 seconds 90 msec
Ended Job = job_1667665836887_0014
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_1667665836887_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0015
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-11-05 11:31:09,542 Stage-2 map = 0%,  reduce = 0%
2022-11-05 11:31:22,326 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.81 sec
2022-11-05 11:31:38,226 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.22 sec
MapReduce Total cumulative CPU time: 4 seconds 220 msec
Ended Job = job_1667665836887_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.09 sec   HDFS Read: 29215 HDFS Write: 3289 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.22 sec   HDFS Read: 8526 HDFS Write: 14 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 310 msec
OK
product maxsales
S18_3232        1774
Time taken: 94.054 seconds, Fetched: 1 row(s)
```

Calculate the total sales for each quarter:

Query:

select QTR_ID as Quarter,sum(SALES) as TotalSales from sales_order_orc group by QTR_ID;

```
hive> select QTR_ID as Quarter,sum(SALES) as TotalSales from sales_order_orc group by QTR_ID;
Query ID = cloudera_20221105113535_3617c6e1-dfd9-4005-b943-438d9e48b5fb
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_1667665836887_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 11:36:09,949 Stage-1 map = 0%,  reduce = 0%
2022-11-05 11:36:25,551 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.23 sec
2022-11-05 11:36:43,476 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.69 sec
MapReduce Total cumulative CPU time: 4 seconds 690 msec
Ended Job = job_1667665836887_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.69 sec   HDFS Read: 37474 HDFS Write: 81 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 690 msec
OK
quarter totalsales
1       2350817.726501465
2       2048120.3029174805
3       1758910.808959961
4       3874780.010925293
Time taken: 54.656 seconds, Fetched: 4 row(s)
```

In which quarter sales was minimum:

Query:

select QTR_ID as Quarter,sum(SALES) as TotalSales from sales_order_orc group by QTR_ID order by TotalSales asc limit 1;

```
hive> select QTR_ID as Quarter,sum(SALES) as TotalSales from sales_order_orc group by QTR_ID order by TotalSales asc limit 1;
Query ID = cloudera_20221105113939_4c953b67-c89d-40e1-935f-28d8cedbc3e6
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_1667665836887_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 11:39:32,460 Stage-1 map = 0%,  reduce = 0%
2022-11-05 11:39:44,127 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.98 sec
2022-11-05 11:40:00,298 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1667665836887_0017
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_1667665836887_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0018
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-11-05 11:40:17,806 Stage-2 map = 0%,  reduce = 0%
2022-11-05 11:40:32,887 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.19 sec
2022-11-05 11:40:49,030 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.69 sec
MapReduce Total cumulative CPU time: 4 seconds 690 msec
Ended Job = job_1667665836887_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.29 sec   HDFS Read: 36546 HDFS Write: 200 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.69 sec   HDFS Read: 5243 HDFS Write: 20 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 980 msec
OK
quarter totalsales
3       1758910.808959961
Time taken: 92.078 seconds, Fetched: 1 row(s)
```

In which country sales was maximum and in which country sales was minimum:

Query:

SELECT A.country, A.sales FROM(SELECT country,Sales,RANK() OVER (ORDER BY sales) rnk_min,RANK() OVER (ORDER BY Sales DESC) rnk_max FROM sales_order_orc)A WHERE rnk_min = 1 OR rnk_max = 1 ORDER BY Sales;

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.41 sec   HDFS Read: 39057 HDFS Write: 85350 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 6.92 sec   HDFS Read: 93973 HDFS Write: 149 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 4.1 sec   HDFS Read: 4936 HDFS Write: 26 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 430 msec
OK
a.country       a.sales
France  482.13
USA     14082.8
Time taken: 142.862 seconds, Fetched: 2 row(s)
```

Calculate quartelry sales for each city:

Query:

select city,QTR_ID,sum(SALES) as TotalSales from sales_order_orc group by city,QTR_ID;

```
hive> select city,QTR_ID,sum(SALES) as TotalSales from sales_order_orc group by city,QTR_ID;
Query ID = cloudera_20221105120101_cea36887-ffc2-4cf8-b6c3-ff74abfb6201
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_1667665836887_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0023
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 12:01:19,915 Stage-1 map = 0%,  reduce = 0%
2022-11-05 12:01:32,718 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.16 sec
2022-11-05 12:01:48,688 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.24 sec
MapReduce Total cumulative CPU time: 5 seconds 240 msec
Ended Job = job_1667665836887_0023
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.24 sec   HDFS Read: 39480 HDFS Write: 5283 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 240 msec
OK
city    qtr_id  totalsales
Aaarhus 4       100595.5498046875
Allentown       2       6166.7998046875
Allentown       3       71930.61041259766
Allentown       4       44040.729736328125
Barcelona       2       4219.2001953125
Barcelona       4       74192.66003417969
Bergamo 1       56181.320068359375
Bergamo 4       81774.40008544922
Bergen  3       16363.099975585938
Bergen  4       95277.17993164062
Boras   1       31606.72021484375
```

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

Query:

select MONTH_ID,YEAR_ID,QUANTITYORDERED from (select MONTH_ID,YEAR_ID,QUANTITYORDERED ,dense_rank() over(partition by YEAR_ID order by QUANTITYORDERED desc) as rnk from sales_order_orc)a where a.rnk =1;

```
hive> select MONTH_ID,YEAR_ID,QUANTITYORDERED  from (select MONTH_ID,YEAR_ID,QUANTITYORDERED ,dense_rank() over(partition by YEAR_ID order by QUANTITYORDERED desc) as r
nk from sales_order_orc)a where a.rnk =1;
Query ID = cloudera_20221105120606_587bcbdc-4ee2-44d9-9f5e-a65b401adf48
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_1667665836887_0025, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667665836887_0025/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667665836887_0025
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 12:06:33,042 Stage-1 map = 0%,  reduce = 0%
2022-11-05 12:06:44,754 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.26 sec
2022-11-05 12:07:02,701 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.1 sec
MapReduce Total cumulative CPU time: 6 seconds 100 msec
Ended Job = job_1667665836887_0025
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.1 sec   HDFS Read: 32843 HDFS Write: 392 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 100 msec
OK
month_id        year_id quantityordered
11      2003    50
3       2003    50
11      2003    50
5       2003    50
9       2003    50
2       2003    50
11      2003    50
10      2003    50
2       2003    50
2       2003    50
5       2003    50
11      2003    50
11      2003    50
10      2003    50
6       2003    50
9       2003    50
5       2003    50
10      2003    50
```