

## 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).

## 2. Store raw data into hdfs location

```
cloudera@quickstart:~/Hive_class/
[cloudera@quickstart ~]$ cd Hive_class/
[cloudera@quickstart Hive_class]$ ls
array_data.csv  csv_file.csv  department_data.csv  full_grouped.csv  map_data.csv  sales_order_data.csv
[cloudera@quickstart Hive_class]$ hadoop fs -ls /
Found 6 items
drwxrwxrwx  - hdfs  supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x  - hbase supergroup          0 2022-09-15 06:04 /hbase
drwxr-xr-x  - solr  solr               0 2017-10-23 09:18 /solr
drwxrwxrwt  - hdfs  supergroup          0 2022-09-15 06:41 /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 Hive_class]$ hadoop fs -ls /tmp/
Found 5 items
drwxr-xr-x  - cloudera supergroup      0 2022-09-15 06:27 /tmp/Hive_class_data
drwxrwxrwt  - mapred   mapred           0 2017-10-23 09:15 /tmp/hadoop-yarn
drwx-wx-wx  - hive    supergroup        0 2022-09-02 23:08 /tmp/hive
drwxrwxrwt  - mapred   hadoop          0 2017-10-23 09:17 /tmp/logs
-rw-r--r--  1 cloudera supergroup      17 2022-08-31 06:30 /tmp/test.txt
[cloudera@quickstart Hive_class]$ hadoop fs -copyFromLocal /home/cloudera/Hive_class/sales_order_data.csv /tmp/assignment
[cloudera@quickstart Hive_class]$ hadoop fs -ls /tmp/
Found 6 items
drwxr-xr-x  - cloudera supergroup      0 2022-09-15 06:27 /tmp/Hive_class_data
-rw-r--r--  1 cloudera supergroup  360233 2022-09-15 06:43 /tmp/assignment
drwxrwxrwt  - mapred   mapred           0 2017-10-23 09:15 /tmp/hadoop-yarn
drwx-wx-wx  - hive    supergroup        0 2022-09-02 23:08 /tmp/hive
drwxrwxrwt  - mapred   hadoop          0 2017-10-23 09:17 /tmp/logs
-rw-r--r--  1 cloudera supergroup      17 2022-08-31 06:30 /tmp/test.txt
[cloudera@quickstart Hive_class]$
```

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

```
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 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")
  > ;
OK
Time taken: 0.111 seconds
hive> [REDACTED]
```

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

```
hive> load data inpath 'hdfs://tmp/assignment/' into table sales_order_csv;
Loading data to table hive_class_b1.sales_order_csv
Table hive_class_b1.sales_order_csv stats: (numFiles=1, totalSize=360233)
OK
Time taken: 0.721 seconds
hive> select * from sales_order_csv limit 5;
OK
10107 30    95.7   2     2871.0 Shipped 1    2     2003   Motorcycles 95    $10_1678   2125557818   NYC    NY    10022   USA    NA    Yu    Kwai    Small
10121 34    81.35  5     2765.9 Shipped 2    5     2003   Motorcycles 95    $10_1678   26.47.1555   Reims   Paris  51100   France  EMEA  Henriet Paul  Small
10134 41    94.74  2     3884.34 Shipped 3    7     2003   Motorcycles 95    $10_1678   +33 1 46 62 7555   Paris  75508   France  EMEA  Da Cunha  D
aniel  Medium
10145 45    83.26  6     3746.7 Shipped 3    8     2003   Motorcycles 95    $10_1678   6265557265   Pasadena  CA    90003   USA    NA    Young  Julie
Medium
10159 49    100.0   14    5205.27 Shipped 4    10    2003   Motorcycles 95    $10_1678   6505551386   San Francisco  CA    USA    NA    Brown  Julie
Medium
Time taken: 0.719 seconds, Fetched: 5 row(s)
hive> [REDACTED]
```

**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 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;
```

OK

Time taken: 0.263 seconds

hive> █

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

```
hive> 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_20220915065353_ed91803d-aicc-4d07-bdfb-3df0b7cd1f5f
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_1663247034259_0001, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1663247034259_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663247034259_0001
Hadoop job information for Stage-1: number of mappers: 17 number of reducers: 0
2022-09-15 06:54:11,257 Stage-1 map = 0%, reduce = 0%
2022-09-15 06:54:19,232 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.61 sec
MapReduce Total cumulative CPU time: 2 seconds 610 msec
Ended Job = job_1663247034259_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_2022-09-15_06-53-58_817_3645248176405301154-1/-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=37548, rawDataSize=3153291]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Cumulative CPU: 2.61 sec  HDFS Read: 367327 HDFS Write: 37640 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 610 msec
OK
Time taken: 22.038 seconds
hive> █
```

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

**a. Calculate total sales per year.**

```
hive> select year_id, sum(sales) as total_sales from sales_order_orc group by year_id;
```

```
hive> select year_id, sum(sales) from sales_order_orc group by year_id;
Query ID = cloudera_20220915065757_0df9eff-01ff-4839-8224-38fbe14dbd15
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_1663247034259_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663247034259_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663247034259_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 06:57:21,972 Stage-1 map = 0%,  reduce = 0%
2022-09-15 06:57:28,539 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.23 sec
2022-09-15 06:57:35,896 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.99 sec
MapReduce Total cumulative CPU time: 2 seconds 990 msec
Ended Job = job_1663247034259_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 2.99 sec  HDFS Read: 36816 HDFS Write: 70 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 990 msec
OK
2003    3516979.547241211
2004    4724162.593383789
2005    1791486.7086791992
Time taken: 22.882 seconds, Fetched: 3 row(s)
hive>
```

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

```
hive> select productline, sum(quantityordered) as total_quantity from sales_order_orc group by productline order by total_quantity desc limit 1;
```

```
hive> select productline, sum(quantityordered) as total_quantity from sales_order_orc group by productline order by total_quantity desc limit 1;
Query ID = cloudera_20220915070505_193b39c0-faffd-4258-9ba7-57c87c1c622b
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_1663247034259_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663247034259_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663247034259_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 07:05:48,748 Stage-1 map = 0%,  reduce = 0%
2022-09-15 07:05:55,089 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.18 sec
2022-09-15 07:06:01,467 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.61 sec
MapReduce Total cumulative CPU time: 2 seconds 610 msec
Ended Job = job_1663247034259_0005
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_1663247034259_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663247034259_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663247034259_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-15 07:06:11,379 Stage-2 map = 0%,  reduce = 0%
2022-09-15 07:06:17,652 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.99 sec
2022-09-15 07:06:24,112 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.34 sec
MapReduce Total cumulative CPU time: 2 seconds 340 msec
Ended Job = job_1663247034259_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 2.61 sec  HDFS Read: 28424 HDFS Write: 311 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 2.34 sec  HDFS Read: 5414 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 950 msec
OK
Classic Cars    33992
Time taken: 43.157 seconds, Fetched: 1 row(s)
hive>
```

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

```
hive> select year_id, qtr_id, sum(sales) as total_sales from sales_order_orc group by year_id,qtr_id;
```

```
hive> select year_id, qtr_id, sum(sales) as total_sales from sales_order_orc group by year_id,qtr_id;
Query ID = cloudera_20220915223030_78821dad-077f-4a64-b7de-d4570185937f
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_1663301977043_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 22:30:27,023 Stage-1 map = 0%,  reduce = 0%
2022-09-15 22:30:33,187 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.26 sec
2022-09-15 22:30:40,554 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.73 sec
MapReduce Total cumulative CPU time: 2 seconds 730 msec
Ended Job = job_1663301977043_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 2.73 sec  HDFS Read: 37729 HDFS Write: 253 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 730 msec
OK
year_id qtr_id  total_sales
2003    1        445094.6897583008
2003    2        562365.2218017578
2003    3        649514.5415039062
2003    4        1860005.094177246
2004    1        833730.6786499023
2004    2        766260.7305297852
2004    3        1109396.2674560547
2004    4        2014774.9167480469
2005    1        1071992.3580932617
2005    2        719494.3505859375
Time taken: 21.03 seconds, Fetched: 10 row(s)
hive> █
```

**d. In which quarter sales were minimum.**

```
hive> WITH cte AS
```

```
  > ( select year_id,min(sales) as min_sales from sales_order_orc group by year_id )
  > select distinct a.year_id,a.month_id,a.sales from sales_order_orc a join cte b on a.year_id =
b.year_id where a.sales = b.min_sales;
```

```
a.year_id      a.month_id      a.sales
2003    3        710.2
2004    8        577.6
2005    5        482.13
Time taken: 51.924 seconds, Fetched: 3 row(s)
hive> █
```

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

**Minimum sales:**

```
hive> select country, sum(sales) as total_sales from sales_order_orc group by country order by total_sales limit 1;
```

```
hive> select country, sum(sales) as total_sales from sales_order_orc group by country order by total_sales limit 1;
Query ID = cloudera_20220915214040_e8ad37fd-ceab-47bd-802a-b76a10c2dea3
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_1663301977043_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 21:40:15,354 Stage-1 map = 0%,  reduce = 0%
2022-09-15 21:40:20,605 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.36 sec
2022-09-15 21:40:28,042 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.74 sec
MapReduce Total cumulative CPU time: 2 seconds 740 msec
Ended Job = job_1663301977043_0002
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_1663301977043_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-15 21:40:37,865 Stage-2 map = 0%,  reduce = 0%
2022-09-15 21:40:44,150 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.0 sec
2022-09-15 21:40:49,443 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.45 sec
MapReduce Total cumulative CPU time: 2 seconds 450 msec
Ended Job = job_1663301977043_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 2.74 sec  HDFS Read: 36960 HDFS Write: 716 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 2.45 sec  HDFS Read: 5781 HDFS Write: 26 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 190 msec
OK
Ireland 57756.43029785156
Time taken: 44.999 seconds, Fetched: 1 row(s)
hive> 
```

## Maximum sales

```
hive> select country, sum(sales) as total_sales from sales_order_orc group by country order by total_sales desc limit 1;
```

```
hive> select country, sum(sales) as total_sales from sales_order_orc group by country order by total_sales desc limit 1;
Query ID = cloudera_20220915214242_579b95eb-409a-4f38-982f-b8714baee18a
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_1663301977043_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 21:42:20,850 Stage-1 map = 0%,  reduce = 0%
2022-09-15 21:42:26,182 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.29 sec
2022-09-15 21:42:33,533 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.61 sec
MapReduce Total cumulative CPU time: 2 seconds 610 msec
Ended Job = job_1663301977043_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_1663301977043_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-15 21:42:41,016 Stage-2 map = 0%,  reduce = 0%
2022-09-15 21:42:47,332 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.09 sec
2022-09-15 21:42:53,619 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.57 sec
MapReduce Total cumulative CPU time: 2 seconds 570 msec
Ended Job = job_1663301977043_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 2.61 sec  HDFS Read: 37043 HDFS Write: 716 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 2.57 sec  HDFS Read: 5781 HDFS Write: 22 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 180 msec
OK
USA      3627982.825744629
Time taken: 41.549 seconds, Fetched: 1 row(s)
hive>
```

## f. Calculate quarterly sales for each city.

```
hive> select city , QTR_ID , sum(sales) as total_sales from sales_order_orc group by city,QTR_ID order by city, QTR_ID;
```

```
hive> select city , QTR_ID , sum(sales) as total_sales from sales_order_orc group by city,QTR_ID order by city, QTR_ID;
Query ID = cloudera_20220915215151_f3fcac51-c5dc-4192-8b9c-f2113e1804ba
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_1663301977043_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-15 21:51:46,855 Stage-1 map = 0%, reduce = 0%
2022-09-15 21:51:53,127 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.27 sec
2022-09-15 21:51:58,374 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.63 sec
MapReduce Total cumulative CPU time: 2 seconds 630 msec
Ended Job = job_1663301977043_0009
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_1663301977043_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663301977043_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663301977043_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-15 21:52:08,937 Stage-2 map = 0%, reduce = 0%
2022-09-15 21:52:15,308 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.38 sec
2022-09-15 21:52:21,784 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.83 sec
MapReduce Total cumulative CPU time: 2 seconds 830 msec
Ended Job = job_1663301977043_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.63 sec HDFS Read: 38164 HDFS Write: 6528 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.83 sec HDFS Read: 12052 HDFS Write: 5203 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 460 msec
OK
+-----+
|city   |qtr_id|total_sales|
+-----+
|Aarhus|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|
```

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

hive> WITH cte AS

```
> ( select year_id,max(quantityordered) as max_quantityordered from sales_order_orc  
group by year_id )  
> select distinct a.year_id,a.month_id,a.quantityordered from sales_order_orc a join cte b  
on a.year_id = b.year_id where a.quantityordered = b.max_quantityordered;
```

a.year_id	a.month_id	a.quantityordered
2003	1	50
2003	2	50
2003	3	50
2003	4	50
2003	5	50
2003	6	50
2003	9	50
2003	10	50
2003	11	50
2004	11	55
2005	4	97

Time taken: 51.873 seconds, Fetched: 11 row(s)