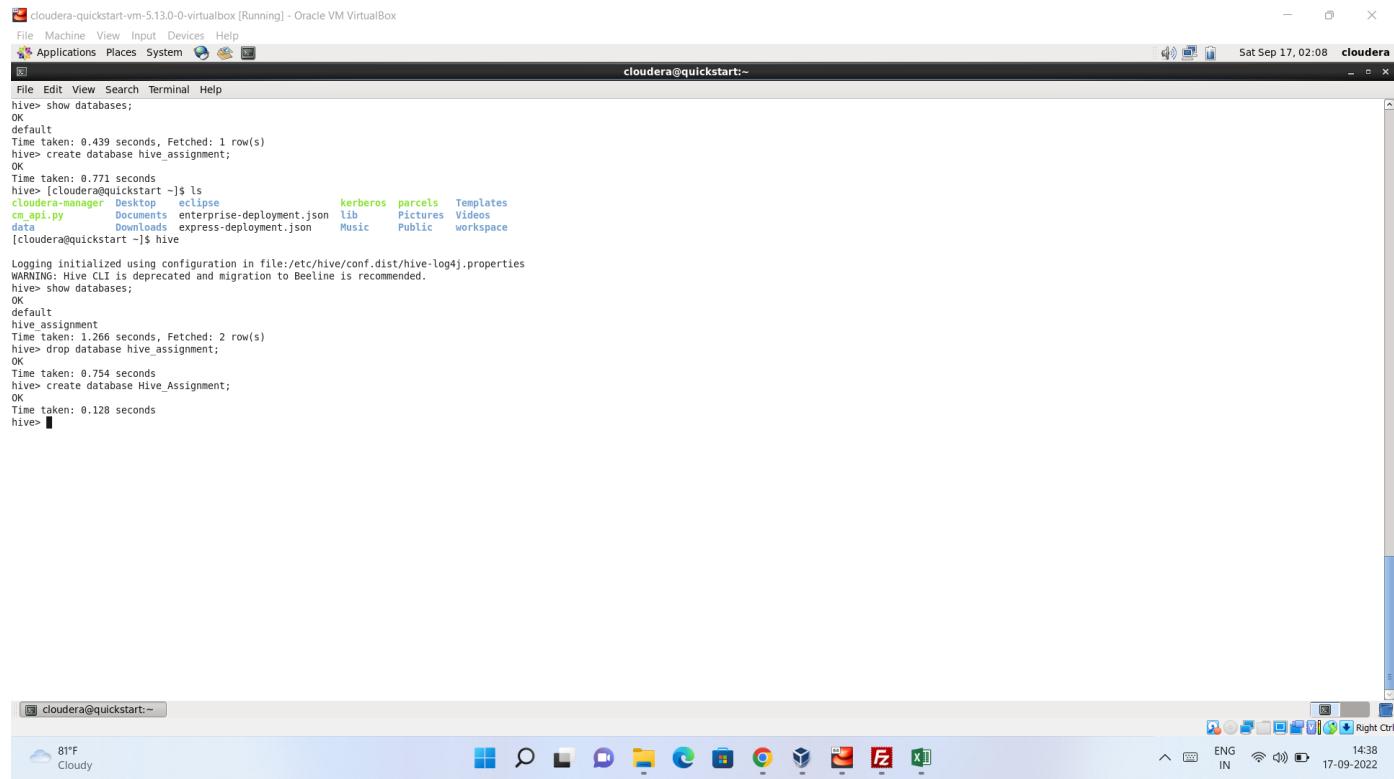
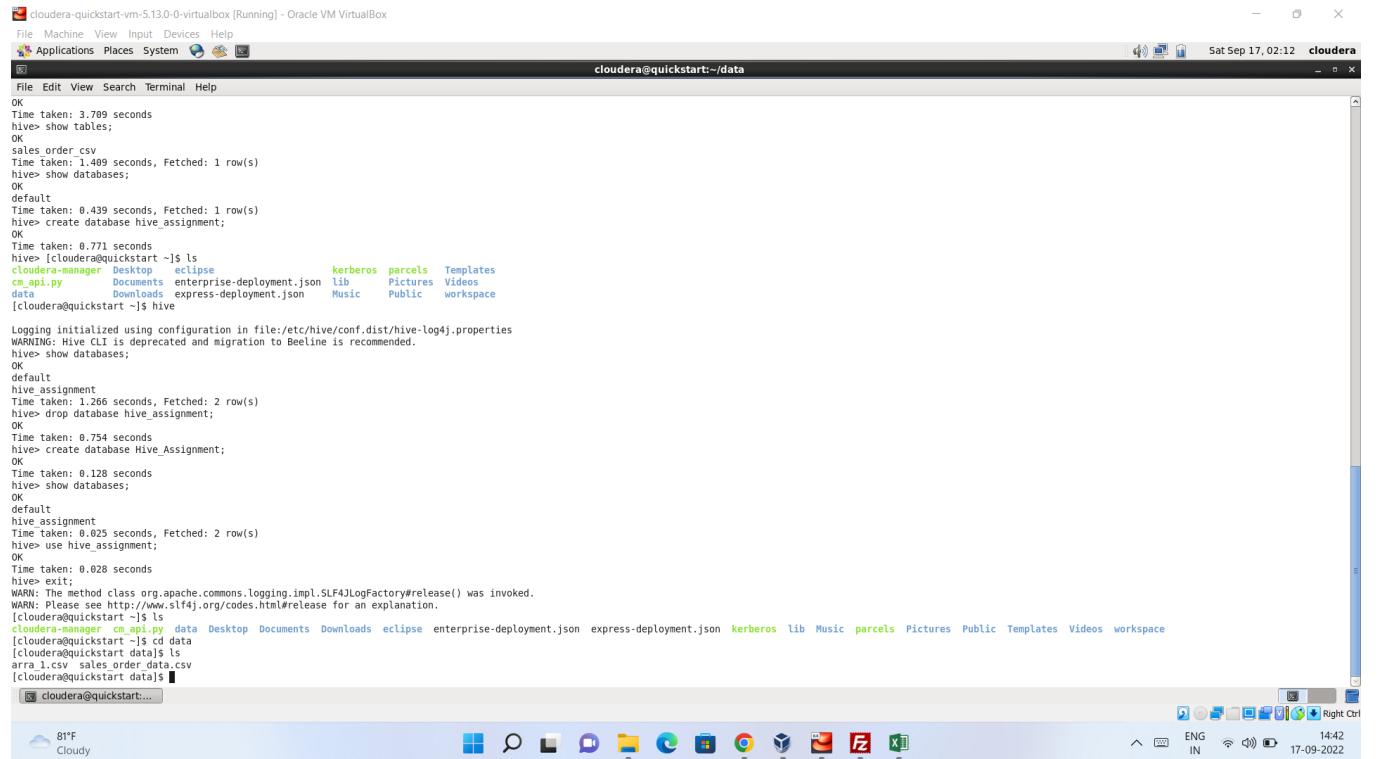


Created the database



```
cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~ Sat Sep 17, 02:08 cloudera
File Edit View Search Terminal Help
hive> show databases;
OK
default
Time taken: 0.439 seconds, Fetched: 1 row(s)
hive> create database hive_assignment;
OK
Time taken: 0.771 seconds
hive> [cloudera@quickstart ~]$ ls
cloudera-manager eclipse kerberos parcels Templates
cm_api.py Documents enterprise-deployment.json lib Pictures Videos
data Downloads express-deployment.json Music Public workspace
[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> show databases;
OK
hive_assignment
Time taken: 1.266 seconds, Fetched: 2 row(s)
hive> drop database hive_assignment;
OK
Time taken: 0.754 seconds
hive> create database Hive_Assignment;
OK
Time taken: 0.128 seconds
hive> [cloudera@quickstart ~]$
```

Loaded the data into local file system



```
cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~/data Sat Sep 17, 02:12 cloudera
File Edit View Search Terminal Help
OK
Time taken: 3.709 seconds
hive> show tables;
OK
sales_order.csv
Time Taken: 1.409 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
Time taken: 0.439 seconds, Fetched: 1 row(s)
hive> create database hive_assignment;
OK
Time taken: 0.771 seconds
hive> [cloudera@quickstart ~]$ ls
cloudera-manager Desktop eclipse kerberos parcels Templates
cm_api.py Documents enterprise-deployment.json lib Pictures Videos
data Downloads express-deployment.json Music Public workspace
[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> show databases;
OK
hive_assignment
Time taken: 1.266 seconds, Fetched: 2 row(s)
hive> drop database hive_assignment;
OK
Time taken: 0.754 seconds
hive> create database Hive_Assignment;
OK
Time taken: 0.128 seconds
hive> show databases;
OK
default
hive_assignment
Time taken: 0.025 seconds, Fetched: 2 row(s)
hive> use hive_assignment;
OK
Time taken: 0.026 seconds
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ ls
cloudera-manager cm_api.py data Desktop Documents Downloads eclipse enterprise-deployment.json express-deployment.json kerberos lib Music parcels Pictures Public Templates Videos workspace
[cloudera@quickstart ~]$ cd data
[cloudera@quickstart data]$ ls
arra_1.csv sales_order_data.csv
[cloudera@quickstart data]$ [cloudera@quickstart data]$
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~/data
File Edit View Search Terminal Help
hive> show tables;
OK
sales_order.csv
Time taken: 1.409 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
Time taken: 0.439 seconds, Fetched: 1 row(s)
hive> create database hive_assignment;
OK
Time taken: 0.771 seconds
hive> [cloudera@quickstart ~]$ ls
cloudera-manager Desktop eclipse kerberos parcels Templates
cm_api.py Documents enterprise-deployment.json lib Pictures Videos
data Downloads express-deployment.json Music Public workspace
[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> show databases;
OK
default
hive_assignment
Time taken: 1.266 seconds, Fetched: 2 row(s)
hive> drop database hive_assignment;
OK
Time taken: 0.754 seconds
hive> create database Hive_Assignment;
OK
Time taken: 0.128 seconds
hive> show databases;
OK
default
hive_assignment
Time taken: 0.025 seconds, Fetched: 2 row(s)
hive> use hive_assignment;
OK
Time taken: 0.028 seconds
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ ls
cloudera-manager cm_api.py data Desktop Documents Downloads eclipse enterprise-deployment.json express-deployment.json kerberos lib Music parcels Pictures Public Templates Videos workspace
[cloudera@quickstart ~]$ cd data
[cloudera@quickstart data]$ ls
arquivos_gerais sales_order.csv
[cloudera@quickstart data]$ hdfs dfs -mkdir /data
mkdir: '/data': File exists
[cloudera@quickstart data]$ [cloudera@quickstart ~]$
```

Created the data directory in hdfs location.

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~/data
File Edit View Search Terminal Help
hive> show tables;
OK
sales_order.csv
Time taken: 0.439 seconds, Fetched: 1 row(s)
hive> create database hive_assignment;
OK
Time taken: 0.771 seconds
hive> [cloudera@quickstart ~]$ ls
cloudera-manager Desktop eclipse kerberos parcels Templates
cm_api.py Documents enterprise-deployment.json lib Pictures Videos
data Downloads express-deployment.json Music Public workspace
[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> show databases;
OK
default
hive_assignment
Time taken: 1.266 seconds, Fetched: 2 row(s)
hive> drop database hive_assignment;
OK
Time taken: 0.754 seconds
hive> create database Hive_Assignment;
OK
Time taken: 0.128 seconds
hive> show databases;
OK
default
hive_assignment
Time taken: 0.025 seconds, Fetched: 2 row(s)
hive> use hive_assignment;
OK
Time taken: 0.028 seconds
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ ls
cloudera-manager cm_api.py data Desktop Documents Downloads eclipse enterprise-deployment.json express-deployment.json kerberos lib Music parcels Pictures Public Templates Videos workspace
[cloudera@quickstart ~]$ cd data
[cloudera@quickstart data]$ ls
arquivos_gerais sales_order.csv
[cloudera@quickstart data]$ hdfs dfs -mkdir /data
mkdir: '/data': File exists
[cloudera@quickstart data]$ hdfs dfs -put /data/sales_order_data.csv /data/
put: /data/sales_order_data.csv: No such file or directory
[cloudera@quickstart data]$ hdfs dfs -put /home/cloudera/data/sales_order_data.csv /data/
[cloudera@quickstart data]$ hdfs dfs -ls /data/
Found 1 items
-rw-r--r-- 1 cloudera supergroup 360233 2022-09-17 02:14 /data/sales_order_data.csv
[cloudera@quickstart data]$ [cloudera@quickstart ~]$
```

Loaded the data from the local file system to a directory data in hdfs location.

```
[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_assignment;
OK
Time taken: 0.884 seconds
hive> create table sales.order_csv
    > (ordernumber int,
     > quantityordered int,
     > priceeach float,
     > orderline 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,
     > postalcoding 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: 1.156 seconds
hive> [cloudera@quickstart ~]$
```

Created the sales_order_csv table.

```
[cloudera@quickstart ~]$ hive
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:-
File Edit View Search Terminal Help
[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_assignment;
OK
Time taken: 2.766 seconds
hive> show tables;
OK
sales.order_csv
Time taken: 1.085 seconds, Fetched: 1 row(s)
hive> load data local inpath '/data/sales.order.data.csv' into table sales.order_csv;
FAILED: SemanticException Line 1:23 Invalid path ''/data/sales.order.data.csv'': No files matching path file:/data/sales.order.data.csv
hive> load data local inpath 'hdfs://data/sales.order.data.csv' into table sales.order_csv;
FAILED: SemanticException [Error 10028]: Line 1:23 Path is not legal 'hdfs://data/sales.order.data.csv'': Source file system should be "file" if "local" is specified
hive> load data local inpath 'hdfs://data/sales.order.data.csv' into table sales.order_csv;
FAILED: SemanticException [Error 10028]: Line 1:23 Path is not legal 'hdfs://data/sales.order.data.csv'': Source file system should be "file" if "local" is specified
hive> load data local inpath '/data/sales.order.data.csv' into table sales.order_csv;
FAILED: SemanticException Line 1:23 Invalid path ''/data/sales.order.data.csv'': No files matching path file:/data/sales.order.data.csv
hive> load data inpath '/data/sales.order.data.csv' into table sales.order_csv;
Loading data to table hive_assignment.sales.order_csv...
Table hive_assignment.sales.order_csv stats: [numFiles=1, totalSize=360233]
OK
Time taken: 6.006 seconds
hive> select * from sales.order_csv limit 10;
OK
10107 38 95.7 2 2871.0 Shipped 1 2 2003 Motorcycles 95 $10.1678 2125597818 NYC NY 10022 USA NA Yu Kwai Small
10121 34 81.35 5 2765.9 Shipped 2 5 2003 Motorcycles 95 $10.1678 26.47.155 Reims 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 Daniel Medium
10145 45 83.26 6 3746.7 Shipped 3 8 2003 Motorcycles 95 $10.1678 6265557265 Pasadena CA 90083 USA EMEA 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
10168 36 96.66 1 3479.76 Shipped 4 10 2003 Motorcycles 95 $10.1678 6505556809 Burlingame CA 94217 USA NA Hirano Juri Medium
10180 29 86.13 9 2497.77 Shipped 4 11 2003 Motorcycles 95 $10.1678 20.16.1555 Lille 59000 France EMEA Rance Martine Small
10188 48 100.0 1 5512.32 Shipped 4 11 2003 Motorcycles 95 $10.1678 +47 2267 3215 Bergen N 5804 Norway EMEA Oeztan Veysel Medium
10201 22 98.57 2 2168.54 Shipped 4 12 2003 Motorcycles 95 $10.1678 6505555787 San Francisco CA USA NA Murphy Julie Small
10211 41 100.0 14 4708.44 Shipped 1 1 2004 Motorcycles 95 $10.1678 (1) 47.55.6555 Paris 75016 France EMEA Perrier Dominique Medium
Time taken: 1.253 seconds, Fetched: 10 row(s)
hive> [cloudera@quickstart ~]$
```

Queried the sales_order_csv table to verify whether loading was successful.

```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sat Sep 17, 5:27 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
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> show databases;
OK
hive> use hive_assignment;
OK
Time taken: 0.06 seconds
hive> show tables;
OK
sales_order_csv
Time taken: 0.055 seconds, Fetched: 1 row(s)
hive> create table sales.order_orc
> ;
FAILED: SemanticException [Error 10043]: Either list of columns or a custom serializer should be specified
hive> create table sales.order_orc
> (
>   ordernumber int,
>   quantityordered int,
>   priceeach float,
>   orderlinenumber int,
>   sales float,
>   status string,
>   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: 0.212 seconds
hive> show tables;
OK
sales_order_csv
sales.order_orc
Time taken: 0.025 seconds, Fetched: 2 row(s)
hive> |

```

Created the sales_order_orc table.

```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sat Sep 17, 5:30 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
> msrp int,
> productcode string,
> phone string,
> city string,
> state string,
> postalcode string,
> territory string,
> contactlastname string,
> contactfirstname string,
> dealsize string
> ) stored as orc;
OK
Time taken: 0.212 seconds
hive> show tables;
OK
sales_order_csv
sales.order_orc
Time taken: 0.025 seconds, Fetched: 2 row(s)
hive> from sales_order_csv insert overwrite table sales.order_orc select *;
Query ID = cloudera_20220917052929_84e70226-7852-4b7d-99f8-09dbff6c8558
Total 1 jobs
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663407121153_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-09-17 05:29:30,464 Stage-1 map = 0%, reduce = 0%
2022-09-17 05:29:42,464 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.65 sec
MapReduce total cumulative CPU time: 2 seconds 650 msec
Ended Job = job_1663407121153_0002
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_assignment.db/sales.order_orc.hive-staging_hive_2022-09-17_05-29-19_579_4056859447157271750-1-ext-10000
Loading data to table hive_assignment.sales.order_orc
Table hive_assignment.sales.order_orc stats: [numFiles=1, numRows=2823, totalSize=37548, rawDataSize=3153291]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.65 sec HDFS Read: 367540 HDFS Write: 37642 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 650 msec
OK
Time taken: 26.183 seconds
hive> select * from sales.order_orc limit 3;
OK
10107 30 95.7 2 2871.8 Shipped 1 2 2003 Motorcycles 95 $10_1678 2125557818 NYC NY 10022 USA NA Yu Kwai Small
10121 34 81.35 9 2765.9 Shipped 2 5 2003 Motorcycles 95 $10_1678 26.47.1555 Reims Paris 51108 France EMEA Henriot 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 Daniel Medium
Time taken: 0.098 seconds, Fetched: 3 row(s)
hive> |

```

Loaded the data from the sales_order_csv to sales_order_orc and queried the sales_order_orc for verification.

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System Sat Sep 17, 5:35 AM cloudera

cloudera@quickstart:~

```

File Edit View Search Terminal Help
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663407121153_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 05:34:04,501 Stage-1 map = 0%, reduce = 0%
2022-09-17 05:34:10,887 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.38 sec
2022-09-17 05:34:19,210 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.98 sec
MapReduce Total cumulative CPU time: 2 seconds 980 msec
Ended Job = job_1663407121153_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.98 sec HDFS Read: 37173 HDFS Write: 45 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 980 msec
OK
3516979.5    2003
4724162.6    2004
1791486.7    2005
Time taken: 23.600 seconds, Fetched: 3 row(s)
hive> set hive.cli.print.header=true;
hive> select round(sum(sales)),1,year_id from sales order .orc group by year_id;
Query ID = cloudera_20220917053535_92c01163-c2c9-4315-a220-209446da8cc4
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_1663407121153_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 05:35:27,859 Stage-1 map = 0%, reduce = 0%
2022-09-17 05:35:34,100 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.71 sec
2022-09-17 05:35:44,741 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.36 sec
MapReduce Total cumulative CPU time: 5 seconds 360 msec
Ended Job = job_1663407121153_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.36 sec HDFS Read: 37256 HDFS Write: 45 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 360 msec
OK
<0
3516979.5    2003
4724162.6    2004
1791486.7    2005
Time taken: 23.532 seconds, Fetched: 3 row(s)
hive> 
```

cloudera@quickstart:~ [Cloudera Live : Welco...]

80°F Partly sunny

ENG IN 18:05 17-09-2022

Calculate total sales per year

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System Sat Sep 17, 7:08 AM cloudera

cloudera@quickstart:~

```

File Edit View Search Terminal Help
Total MapReduce CPU Time Spent: 14 seconds 190 msec
OK
quantity_ordered      productcode
1774      518 2322
Time taken: 67.2 seconds, Fetched: 1 row(s)
hive> select sum(quantityordered) as quantity_ordered,productline from sales_order_.orc group by productline order by quantity_ordered desc limit 1;
Query ID = cloudera_20220917070707_0b35575e-b51e-456b-898b-a5f1c24eba8
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_1663407121153_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:48:11,000 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:54,702 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.44 sec
2022-09-17 07:08:01,089 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.13 sec
MapReduce Total cumulative CPU time: 4 seconds 130 msec
Ended Job = job_1663407121153_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_1663407121153_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-17 07:08:15,399 Stage-2 map = 0%, reduce = 0%
2022-09-17 07:08:21,678 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 0.98 sec
2022-09-17 07:08:21,678 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.4 sec
MapReduce Total cumulative CPU time: 2 seconds 400 msec
Ended Job = job_1663407121153_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.13 sec HDFS Read: 28591 HDFS Write: 311 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.4 sec HDFS Read: 5231 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 538 msec
OK
quantity_ordered      productline
33992      Classic Cars
Time taken: 45.474 seconds, Fetched: 1 row(s)
hive> 
```

cloudera@quickstart:~ [Cloudera Live : Welco...]

ENG IN 19:38 17-09-2022

Find a product for which maximum orders were placed

```

cloudera@quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sat Sep 17, 7:13 AM cloudera
cloudera@quickstart:~>

File Edit View Search Terminal Help
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1663407121153_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-17 07:08:10,215 Stage-2 map = 0%, reduce = 0%
2022-09-17 07:08:15,399 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.98 sec
2022-09-17 07:08:21,678 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.4 sec
MapReduce Total cumulative CPU time: 2 seconds 408 msec
Ended Job = job_1663407121153_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.13 sec HDFS Read: 28591 HDFS Write: 311 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.4 sec HDFS Read: 5231 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 530 msec
OK
quantity_ordered      productline
33992    Classic Cars
Time taken: 45.474 seconds, Fetched: 1 row(s)
hive> select sum(sales) as total_sales_overall_by_quarter,qtr_id from sales_order_orc group by qtr_id;
Query ID = cloudera_20220917071212_2f6dc3b-5d11-4a54-945b-d3e190c3623d
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=<numbers>
Starting Job = job_1663407121153_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:12:54,657 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:13:03,293 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.22 sec
2022-09-17 07:13:16,395 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.96 sec
MapReduce Total cumulative CPU time: 4 seconds 960 msec
Ended Job = job_1663407121153_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.96 sec HDFS Read: 37281 HDFS Write: 81 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 960 msec
OK
total_sales_overall_by_quarter qtr_id
2358017 7265814805 1
2048120 3629174805 2
1758910 888959061 3
3874780 010925293 4
Time taken: 32.578 seconds, Fetched: 4 row(s)
hive> 
```

Calculate
 the total
 sales for
 each
 quarter
 quarters as a whole

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Sat Sep 17, 7:15 AM cloudera
cloudera@quickstart:~[File Edit View Search Terminal Help
.set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
.set mapreduce.job.reduces=<number>
Starting Job = job_1663407121153_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:08:10,215 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:08:15,399 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.98 sec
2022-09-17 07:08:21,678 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.4 sec
MapReduce Total cumulative CPU time: 2 seconds 400 msec
Ended Job = job_1663407121153_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.13 sec HDFS Read: 2891 HDFS Write: 311 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.4 sec HDFS Read: 5231 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 530 msec
OK
quantities_ordered      productline
33992  Classic Cars
Time taken: 45.474 seconds, Fetched: 1 row(s)
hive> select sum(sales) as total_sales_overall_by_quarter,qtr_id from sales_order_orc group by qtr_id;
Query ID = cloudera_20220917071212_2f6dc3b-5d11-4a54-945b-d3e190c3623d
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_1663407121153_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:12:54,657 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:13:03,293 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.22 sec
2022-09-17 07:13:16,900 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.96 sec
MapReduce Total cumulative CPU time: 4 seconds 960 msec
Ended Job = job_1663407121153_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.96 sec HDFS Read: 37281 HDFS Write: 81 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 960 msec
OK
total_sales_overall_by_quarter  qtr_id
2359817.726501468   1
2048120.3029174805  2
1758910.808959961   3
3874788.018925293   4
Time taken: 32.578 seconds, Fetched: 4 row(s)
hive> [cloudera@quickstart:~] [Cloudera Live : Welco...]
88°F Partly cloudy
19:45 17-09-2022

```

Calculate the total sales for each quarter by year:

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Sat Sep 17, 7:17 AM cloudera
cloudera@quickstart:~[File Edit View Search Terminal Help
.set hive.exec.reducers.max=<number>
2022-09-17 07:13:16,395 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.96 sec
MapReduce Total cumulative CPU time: 4 seconds 960 msec
Ended Job = job_1663407121153_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.96 sec HDFS Read: 37281 HDFS Write: 81 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 960 msec
OK
total_sales_overall_by_quarter  qtr_id
2359817.726501465   1
2048120.3029174805  2
1758910.808959961   3
3874788.018925293   4
Time taken: 32.578 seconds, Fetched: 4 row(s)
hive> select sum(sales) as total_sales_overall_by_quarter,qtr_id,year_id from sales_order_orc group by year_id,qtr_id;
Query ID = cloudera_20220917071616_457caf9-6666-4317-bf1f-c71cf539b0fc
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_1663407121153_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:16:32,238 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:16:43,124 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.6 sec
2022-09-17 07:16:58,523 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.12 sec
MapReduce Total cumulative CPU time: 7 seconds 120 msec
Ended Job = job_1663407121153_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.12 sec HDFS Read: 38003 HDFS Write: 253 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 120 msec
OK
total_sales_overall_by_quarter  qtr_id  year_id
443496.600750898    1    2002
562309.231067579    2    2003
649514.5415039062   3    2003
1868005.094177246   4    2003
833730.6786499823   1    2004
766260.7305297852   2    2004
1109996.207459347   3    2004
2044774.308864069   4    2004
1071904.2589932637  1    2005
719494.3505859375  2    2005
Time taken: 38.177 seconds, Fetched: 10 row(s)
hive> [cloudera@quickstart:~] [Cloudera Live : Welco...]
88°F Partly cloudy
19:47 17-09-2022

```

In

which

quarter

sales

was

minimum

quarters as a whole

```
cloudera-quickstart-vm-5.13.0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera
cloudera@quickstart:~ Sat Sep 17, 7:26 AM cloudera
File Edit View Search Terminal Help
Total MapReduce CPU Time Spent: 11 seconds 190 msec
OK
total_sales_overall_by_quarter qtr_id year_id
1758910.808959961 3
Time Taken: 62.831 seconds, Fetched: 1 row(s)
hive> select sum(sales) as total_sales_overall_by_quarter,qtr_id,year_id from sales_order_orc group by year_id,qtr_id order by total_sales_overall_by_quarter asc limit 1;
Query ID = cloudera_20220917072525_bf404f49-03ae-46e6-b8f3-2ea295a4b66f
Total jobs: 1
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 mapred.job.reduces=<number>
Starting Job = job_1663407121153_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:25:48,187 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:25:59,395 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.44 sec
2022-09-17 07:26:12,798 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.29 sec
MapReduce Total cumulative CPU time: 7 seconds 296 msec
Ended Job = job_1663407121153_0015
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 mapred.job.reduces=<number>
Starting Job = job_1663407121153_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0016
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-17 07:26:32,624 Stage-2 map = 0%, reduce = 0%
2022-09-17 07:26:44,815 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.01 sec
2022-09-17 07:26:44,815 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.75 sec
MapReduce Total cumulative CPU time: 4 seconds 750 msec
Ended Job = job_1663407121153_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.29 sec HDFS Read: 37025 HDFS Write: 386 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.75 sec HDFS Read: 5584 HDFS Write: 25 SUCCESS
Total MapReduce CPU time: 12 seconds 48 msec
OK
total_sales_overall_by_quarter qtr_id year_id
445094.8097583008 1 2003
Time taken: 71.719 seconds, Fetched: 1 row(s)
hive>
```



In which quarter sales ws minimum quarter by year

```
cloudera-quickstart-vm-5.13.0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Search Terminal Help
Applications Places System cloudera
cloudera@quickstart:~ Sat Sep 17, 7:43 AM cloudera
File Edit View Search Terminal Help
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.job.reduces=<number>
Starting Job = job_1663407121153_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-17 07:41:23,128 Stage-2 map = 0%, reduce = 0%, Cumulative CPU 1.64 sec
2022-09-17 07:41:31,877 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1663407121153_0020
Launching Job 4 out of 5
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 mapred.job.reduces=<number>
Starting Job = job_1663407121153_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0021
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-17 07:41:47,794 Stage-3 map = 0%, reduce = 0%
2022-09-17 07:41:56,657 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.47 sec
2022-09-17 07:42:04,303 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 4.58 sec
MapReduce Total cumulative CPU time: 4 seconds 580 msec
Ended Job = job_1663407121153_0021
Launching Job 5 out of 5
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1663407121153_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0022
Hadoop job information for Stage-4: number of mappers: 2; number of reducers: 0
2022-09-17 07:42:15,687 Stage-4 map = 0%, reduce = 0%
2022-09-17 07:42:26,568 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 6.39 sec
MapReduce Total cumulative CPU time: 6 seconds 390 msec
Ended Job = job_1663407121153_0022
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.45 sec HDFS Read: 37254 HDFS Write: 716 SUCCESS
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 7.05 sec HDFS Read: 37262 HDFS Write: 716 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.67 sec HDFS Read: 4733 HDFS Write: 129 SUCCESS
Stage-Stage-5: Map: 1 Reduce: 1 Cumulative CPU: 4.58 sec HDFS Read: 4737 HDFS Write: 125 SUCCESS
Stage-Stage-3: Map: 2 Cumulative CPU: 6.39 sec HDFS Read: 6338 HDFS Write: 48 SUCCESS
Total MapReduce CPU time: 29 seconds 140 msec
OK
  ul.sum 1          w1.country
57756.43029785156  Ireland
3627982.825744629  USA
Time taken: 146.9 seconds, Fetched: 2 row(s)
hive>
```



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

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sat Sep 17, 7:48 AM cloudera
cloudera@quickstart:~>

File Edit View Search Terminal Help
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.reducers=<number>
Starting Job = job_1663407121153_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0023
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-17 07:45:26,448 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:45:44,028 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.14 sec
2022-09-17 07:45:44,028 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.94 sec
MapReduce Total cumulative CPU time: 4 seconds 940 msec
Ended Job = job_1663407121153_0023
MapReduce Job Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.94 sec HDFS Read: 39355 HDFS Write: 5283 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 940 msec
OK
qtr_id quarterly sum of sales city
1 100955.5498946795 Aarhus
2 6166.799804675 Allentown
3 71930.61041259766 Allentown
4 44948.729736328125 Allentown
2 4219.2001953125 Barcelona
4 10000.60000000000 Barcelona
1 50181.20000000000 Bergamo
81774.40008544922 Bergamo
3 16363.099975505938 Bergen
4 95277.17993164062 Bergen
1 31666.72021484375 Boras
3 53941.65981933594 Boras
4 48710.9205322256 Boras
2 20000.20000000000 Boston
3 15344.04000000000 Boston
4 63730.7802734375 Boston
1 31474.7802734375 Brickhaven
2 7277.35009765625 Brickhaven
3 114974.53967285156 Brickhaven
4 11528.52978515625 Brickhaven
2 75778.96000000000 Brighewater
4 20100.00000000000 Brighewater
1 20100.00000000000 Brisbane
3 34100.030829296875 Bruxelles
1 18800.089721679688 Bruxelles
2 8411.949829101562 Bruxelles
3 47760.479736328125 Bruxelles
1 37850.07958984375 Burbank
4 8234.559936523438 Burbank

```

Calculate quartely sales for each city

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sat Sep 17, 7:56 AM cloudera
cloudera@quickstart:~>

File Edit View Search Terminal Help
2022-09-17 07:54:27,295 Stage-1 map = 0%, reduce = 0%
2022-09-17 07:54:37,311 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.01 sec
2022-09-17 07:54:52,262 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.47 sec
MapReduce Total cumulative CPU time: 9 seconds 479 msec
Ended Job = job_1663407121153_0024
Launching Job 2 out of 3
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.reducers=<number>
Starting Job = job_1663407121153_0025, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0025/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0025
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-09-17 07:55:04,805 Stage-2 map = 0%, reduce = 0%
2022-09-17 07:55:12,384 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2022-09-17 07:55:22,410 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.91 sec
MapReduce Total cumulative CPU time: 3 seconds 910 msec
Ended Job = job_1663407121153_0025
Launching Job 3 out of 3
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.reducers=<number>
Starting Job = job_1663407121153_0026, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1663407121153_0026/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1663407121153_0026
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-09-17 07:55:34,558 Stage-3 map = 0%, reduce = 0%
2022-09-17 07:55:42,218 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.82 sec
2022-09-17 07:55:50,218 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 5.64 sec
MapReduce Total cumulative CPU time: 5 seconds 640 msec
Ended Job = job_1663407121153_0026
MapReduce Job Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.47 sec HDFS Read: 37415 HDFS Write: 937 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.91 sec HDFS Read: 5055 HDFS Write: 937 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 5.64 sec HDFS Read: 8713 HDFS Write: 61 SUCCESS
Total MapReduce CPU Time Spent: 19 seconds 20 msec
OK
sub1.year_id sub1.month_id sub1.sum1 sub1.rank
2003 11 1029837.66 1
2004 11 1089848.01 1
2005 5 457861.06 1
Time taken: 104.465 seconds, Fetched: 3 row(s)
hive> 
```

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