HIVE CLASS 3 ASSIGNMENT

1) Download vechile sales data -> https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv

Downloaded.

2) Store raw data into hdfs location

First, created a folder/ directory with name ‘sales’ in hdfs location and then copied the data ‘sales_order_data.csv’ from local to hdfs location in ‘sales’ folder.

[cloudera@quickstart ~]$ **hdfs dfs -mkdir sales**

[cloudera@quickstart ~]$ **hdfs dfs -put /tmp/hive_assignment/sales_order_data.csv sales/**

[cloudera@quickstart ~]$ **hdfs dfs -ls sales/**

Found 1 items

-rw-r--r-- 1 cloudera cloudera 360233 2022-10-02 08:35 sales/sales_order_data.csv

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.

[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> **create database assignment;**

OK

Time taken: 0.244 seconds

hive> **use assignment;**

OK

Time taken: 0.051 seconds

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");** # To skip 1st row from source data 'sales_order_data'as it contains headers in 1st row.

OK

Time taken: 0.687 seconds

4) Load data from hdfs path into "sales_order_csv".

hive> **load data inpath 'sales/' into table sales_order_csv ;**

Loading data to table assignment.sales_order_csv

Table assignment.sales_order_csv stats: [numFiles=1, totalSize=360233]

OK

Time taken: 1.102 seconds

hive> **set hive.cli.print.header = true**; #To print headers on the top of columns.

hive> **select * from sales_order_csv limit 10;**

OK

sales_order_csv.ordernumber sales_order_csv.quantityordered sales_order_csv.priceeach sales_order_csv.orderlinenumber
sales_order_csv.sales sales_order_csv.status sales_order_csv.qtr_id sales_order_csv.month_id
sales_order_csv.year_id sales_order_csv.productline sales_order_csv.msrp sales_order_csv.productcode
sales_order_csv.phone sales_order_csv.city sales_order_csv.state sales_order_csv.postalcode
sales_order_csv.country sales_order_csv.territory sales_order_csv.contactlastname
sales_order_csv.contactfirstname sales_order_csv.dealsize

```
10107   30      95.7    2       2871.0  Shipped 1       2       2003    Motorcycles     95
        S10_1678 2125557818     NYC     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     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           75508   France  EMEA    Da Cunha Daniel Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95
        S10_1678 6265557265     Pasadena CA     90003   USA     NA      Young   Julie   Medium
10159   49      100.0   14      5205.27 Shipped 4       10      2003    Motorcycles     95
        S10_1678 6505551386     San Francisco   CA              USA     NA      Brown   Julie
        Medium
10168   36      96.66   1       3479.76 Shipped 4       10      2003    Motorcycles     95
        S10_1678 6505556809     Burlingame      CA      94217   USA     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              USA     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: 0.201 seconds, Fetched: 10 row(s)

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: 0.291 seconds

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

hive> **from sales_order_csv insert overwrite table sales_order_orc select *;**

Query ID = cloudera_20221002093131_04110d2d-a759-49bc-8e8f-1c2ed4c3aad4

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_1664695311825_0029, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664695311825_0029/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664695311825_0029

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2022-10-02 09:31:38,404 Stage-1 map = 0%, reduce = 0%

2022-10-02 09:31:53,668 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.79 sec

MapReduce Total cumulative CPU time: 4 seconds 790 msec

Ended Job = job_1664695311825_0029

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/assignment.db/sales_order_orc/.hive-staging_hive_2022-10-02_09-31-19_805_5680173471525832207-1/-ext-10000

Loading data to table assignment.sales_order_orc

Table assignment.sales_order_orc stats: [numFiles=1, numRows=2823, totalSize=37548, rawDataSize=3153291]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1   Cumulative CPU: 4.79 sec   HDFS Read: 367310 HDFS Write: 37637 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 790 msec

OK

sales_order_csv.ordernumber sales_order_csv.quantityordered sales_order_csv.priceeach sales_order_csv.orderlinenumber
sales_order_csv.sales sales_order_csv.status sales_order_csv.qtr_id sales_order_csv.month_id
sales_order_csv.year_id sales_order_csv.productline sales_order_csv.msrp sales_order_csv.productcode
sales_order_csv.phone sales_order_csv.city sales_order_csv.state sales_order_csv.postalcode
sales_order_csv.country sales_order_csv.territory sales_order_csv.contactlastname
sales_order_csv.contactfirstname sales_order_csv.dealsize

Time taken: 39.215 seconds

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 sales_order_orc.productcode
sales_order_orc.phone sales_order_orc.city sales_order_orc.state sales_order_orc.postalcode
sales_order_orc.country 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 NA 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 75508 France EMEA Da Cunha Daniel Medium

10145 45 83.26 6 3746.7 Shipped 3 8 2003 Motorcycles 95
S10_1678 6265557265 Pasadena CA 90003 USA NA Young Julie Medium

10159 49 100.0 14 5205.27 Shipped 4 10 2003 Motorcycles 95
S10_1678 6505551386 San Francisco CA USA NA Brown Julie
Medium

10168 36 96.66 1 3479.76 Shipped 4 10 2003 Motorcycles 95
S10_1678 6505556809 Burlingame CA 94217 USA 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 USA 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 DominiqueMedium

Time taken: 0.133 seconds, Fetched: 10 row(s)

a) Calculate total sales per year.

hive> **select YEAR_ID as Year, sum(SALES) as Sales_Per_Year from sales_order_orc group by YEAR_ID;**

Query ID = cloudera_20221002192626_45e37c96-0ff4-4324-8e0c-374a27d4933f

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_1664762051968_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0014/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0014

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2022-10-02 19:26:22,888 Stage-1 map = 0%, reduce = 0%

2022-10-02 19:26:34,124 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.67 sec

2022-10-02 19:26:47,506 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.91 sec

MapReduce Total cumulative CPU time: 6 seconds 910 msec

Ended Job = job_1664762051968_0014

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.91 sec HDFS Read: 36883 HDFS Write: 70 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 910 msec

OK

year sales_per_year

2003 3516979.547241211

2004 4724162.593383789

2005 1791486.7086791992

Time taken: 38.332 seconds, Fetched: 3 row(s)

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

hive> **select productcode from (select productcode, count(ordernumber) as total_orders from sales_order_orc group by productcode order by total_orders desc limit 1) tab;**

Query ID = cloudera_20221003003131_1709d3c1-4413-4ba3-836d-30b7bfa25bd4

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_1664762051968_0106, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0106/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0106

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2022-10-03 00:31:51,401 Stage-1 map = 0%, reduce = 0%

2022-10-03 00:32:02,434 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.76 sec

2022-10-03 00:32:14,653 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.09 sec

MapReduce Total cumulative CPU time: 7 seconds 90 msec

Ended Job = job_1664762051968_0106

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_1664762051968_0107, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0107/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0107

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2022-10-03 00:32:31,869 Stage-2 map = 0%, reduce = 0%

2022-10-03 00:32:41,950 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.18 sec

2022-10-03 00:32:54,785 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.18 sec

MapReduce Total cumulative CPU time: 5 seconds 180 msec

Ended Job = job_1664762051968_0107

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 7.09 sec   HDFS Read: 28829 HDFS Write: 3071 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 5.18 sec   HDFS Read: 8110 HDFS Write: 9 SUCCESS

Total MapReduce CPU Time Spent: 12 seconds 270 msec

OK

productcode

S18_3232

Time taken: 77.694 seconds, Fetched: 1 row(s)

c) Calculate the total sales for each quarter.

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

Query ID = cloudera_20221002192525_0bc6c5c1-5f70-4c85-a5a2-dd5105d9e8e8

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_1664762051968_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0013/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0013

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2022-10-02 19:25:36,997 Stage-1 map = 0%, reduce = 0%

2022-10-02 19:25:49,496 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.1 sec

2022-10-02 19:26:02,988 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.27 sec

MapReduce Total cumulative CPU time: 7 seconds 270 msec

Ended Job = job_1664762051968_0013

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.27 sec HDFS Read: 37069 HDFS Write: 81 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 270 msec

OK

quarter total_sales

1 2350817.726501465

2 2048120.3029174805

3 1758910.808959961

4 3874780.010925293

Time taken: 40.71 seconds, Fetched: 4 row(s)

d) In which quarter sales was minimum.

hive> **select Qtr_id as Quarter_with_min_sales from (select qtr_id, sum(sales) from sales_order_orc group by qtr_id order by qtr_id limit 1)tab;**

Query ID = cloudera_20221002235656_db2cdf26-e44b-4245-b34b-713d39355ec3

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_1664762051968_0101, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0101/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1664762051968_0101
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-10-02 23:56:48,865 Stage-1 map = 0%,  reduce = 0%
2022-10-02 23:56:59,858 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.43 sec
2022-10-02 23:57:13,089 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.48 sec
MapReduce Total cumulative CPU time: 6 seconds 480 msec
Ended Job = job_1664762051968_0101
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_1664762051968_0102, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0102/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1664762051968_0102
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-02 23:57:27,975 Stage-2 map = 0%,  reduce = 0%
2022-10-02 23:57:38,140 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.94 sec
2022-10-02 23:57:51,899 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.13 sec
MapReduce Total cumulative CPU time: 6 seconds 130 msec
Ended Job = job_1664762051968_0102
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.48 sec   HDFS Read: 36273 HDFS Write: 168 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 6.13 sec   HDFS Read: 4611 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 610 msec
OK
quarter_with_min_sales
1
```

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

hive> **select min(case when h=1 then country end)max_sale_country, min(case when l=1 then country end)min_sale_country from (select country, sum(sales), row_number() over(order by sum(sales))l, row_number() over(order by sum(sales) desc)h from sales_order_orc group by country)tab;**

```
Query ID = cloudera_20221002234646_1311248f-62cf-4172-ac9f-3896497487c5
Total jobs = 4
Launching Job 1 out of 4
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_1664762051968_0095, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0095/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1664762051968_0095
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-10-02 23:47:10,778 Stage-1 map = 0%,  reduce = 0%
2022-10-02 23:47:22,151 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.77 sec
2022-10-02 23:47:38,859 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.0 sec
MapReduce Total cumulative CPU time: 6 seconds 0 msec
Ended Job = job_1664762051968_0095
Launching Job 2 out of 4
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_1664762051968_0096, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0096/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1664762051968_0096
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-10-02 23:47:54,260 Stage-2 map = 0%,  reduce = 0%
2022-10-02 23:48:05,191 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.11 sec
2022-10-02 23:48:18,470 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.07 sec
MapReduce Total cumulative CPU time: 6 seconds 70 msec
Ended Job = job_1664762051968_0096
Launching Job 3 out of 4
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_1664762051968_0097, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0097/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0097

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1

2022-10-02 23:48:34,159 Stage-3 map = 0%, reduce = 0%

2022-10-02 23:48:46,253 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.45 sec

2022-10-02 23:49:01,547 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec

MapReduce Total cumulative CPU time: 6 seconds 310 msec

Ended Job = job_1664762051968_0097

Launching Job 4 out of 4

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_1664762051968_0098, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0098/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0098

Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1

2022-10-02 23:49:17,991 Stage-4 map = 0%, reduce = 0%

2022-10-02 23:49:29,276 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 2.22 sec

2022-10-02 23:49:41,472 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 5.76 sec

MapReduce Total cumulative CPU time: 5 seconds 760 msec

Ended Job = job_1664762051968_0098

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.0 sec HDFS Read: 37053 HDFS Write: 716 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.07 sec HDFS Read: 6817 HDFS Write: 735 SUCCESS

Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 8451 HDFS Write: 125 SUCCESS

Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 5.76 sec HDFS Read: 5156 HDFS Write: 12 SUCCESS

Total MapReduce CPU Time Spent: 24 seconds 140 msec

OK

max_sale_country min_sale_country

USA Ireland

Time taken: 166.051 seconds, Fetched: 1 row(s)

f) Calculate quartelry sales for each city.

hive> **select city, qtr_id as quarter, sum(sales) as quarterly_sales from sales_order_orc group by city, qtr_id;**

Query ID = cloudera_20221002192424_6df35ceb-33db-4ae4-8ebc-5084eb47efea

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_1664762051968_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0012/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0012

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2022-10-02 19:24:42,615 Stage-1 map = 0%, reduce = 0%

2022-10-02 19:24:53,998 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.21 sec

2022-10-02 19:25:09,239 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.89 sec

MapReduce Total cumulative CPU time: 7 seconds 890 msec

Ended Job = job_1664762051968_0012

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.89 sec HDFS Read: 39073 HDFS Write: 5283 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 890 msec

OK

city	quarter	quarterly_sales
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
Boras	3	53941.68981933594
Boras	4	48710.92053222656
Boston	2	74994.240234375
Boston	3	15344.640014648438
Boston	4	63730.7802734375
Brickhaven	1	31474.7802734375
Brickhaven	2	7277.35009765625
Brickhaven	3	114974.53967285156
Brickhaven	4	11528.52978515625
Bridgewater	2	75778.99060058594
Bridgewater	4	26115.800537109375
Brisbane	1	16118.479858398438
Brisbane	3	34100.030029296875
Bruxelles	1	18800.089721679688
Bruxelles	2	8411.949829101562
Bruxelles	3	47760.479736328125
Burbank	1	37850.07958984375
Burbank	4	8234.559936523438
Burlingame	1	13529.570190429688
Burlingame	3	42031.83020019531
Burlingame	4	65221.67004394531
Cambridge	1	21782.699951171875
Cambridge	2	14380.920043945312
Cambridge	3	48828.71942138672
Cambridge	4	54251.659912109375
Charleroi	1	16628.16015625
Charleroi	2	1711.260009765625
Charleroi	3	1637.199951171875
Charleroi	4	13463.480224609375
Chatswood	2	43971.429931640625
Chatswood	3	69694.40002441406
Chatswood	4	37905.14990234375
Cowes	1	26906.68017578125
Cowes	4	51334.15966796875
Dublin	1	38784.470458984375
Dublin	3	18971.959838867188
Espoo	1	51373.49072265625

Espoo	2	31018.230102539062
Espoo	3	31569.430053710938
Frankfurt	1	48698.82922363281
Frankfurt	4	36472.76025390625
Gensve	1	50432.549560546875
Gensve	3	67281.00903320312
Glen Waverly	2	14378.089965820312
Glen Waverly	3	12334.819580078125
Glen Waverly	4	37878.54992675781
Glendale	1	3987.199951171875
Glendale	2	20350.949768066406
Glendale	3	7600.1201171875
Glendale	4	34485.49987792969
Graz	1	8775.159912109375
Graz	4	43488.740234375
Helsinki	1	26422.819458007812
Helsinki	3	42744.0595703125
Helsinki	4	42083.499755859375
Kobenhavn	1	58871.110107421875
Kobenhavn	2	62091.880615234375
Kobenhavn	4	24078.610107421875
Koln	4	100306.58020019531
Las Vegas	2	33847.61975097656
Las Vegas	3	34453.84973144531
Las Vegas	4	14449.609741210938
Lille	1	20178.1298828125
Lille	4	48874.28088378906
Liverpool	2	91211.0595703125
Liverpool	4	26797.210083007812
London	1	8477.219970703125
London	2	32376.29052734375
London	4	83970.029296875
Los Angeles	1	23889.320068359375
Los Angeles	4	24159.14013671875
Lule	1	9748.999755859375
Lule	4	66005.8798828125
Lyon	1	101339.13977050781
Lyon	4	41535.11022949219

Madrid	1	357668.4899291992
Madrid	2	339588.0513305664
Madrid	3	69714.09008789062
Madrid	4	315580.80963134766
Makati City	1	55245.02014160156
Makati City	4	38770.71032714844
Manchester	1	51017.919860839844
Manchester	4	106789.88977050781
Marseille	1	2317.43994140625
Marseille	2	52481.840087890625
Marseille	4	20136.859985351562
Melbourne	1	49637.57067871094
Melbourne	2	60135.84033203125
Melbourne	4	91221.99914550781
Minato-ku	1	38191.38977050781
Minato-ku	2	26482.700256347656
Minato-ku	4	55888.65026855469
Montreal	2	58257.50012207031
Montreal	4	15947.290405273438
Munich	3	34993.92004394531
NYC	1	32647.809814453125
NYC	2	165100.33947753906
NYC	3	63027.92004394531
NYC	4	300011.6999511719
Nantes	1	59617.39978027344
Nantes	2	60344.990173339844
Nantes	3	61310.880126953125
Nantes	4	23031.589599609375
Nashua	1	12133.25
Nashua	4	119552.04949951172
New Bedford	1	48578.95935058594
New Bedford	3	45738.38952636719
New Bedford	4	113557.509765625
New Haven	2	36973.309814453125
New Haven	4	42498.760498046875
Newark	1	8722.1201171875
Newark	2	74506.06909179688
North Sydney	1	65012.41955566406

North Sydney	3	47191.76013183594
North Sydney	4	41791.949462890625
Osaka	1	50490.64013671875
Osaka	2	17114.43017578125
Oslo	3	34145.47021484375
Oslo	4	45078.759765625
Oulu	1	49055.40026855469
Oulu	2	17813.40008544922
Oulu	3	37501.580322265625
Paris	1	71494.17944335938
Paris	2	80215.4203491211
Paris	3	27798.480102539062
Paris	4	89436.60034179688
Pasadena	1	44273.359436035156
Pasadena	3	55776.119873046875
Pasadena	4	4512.47998046875
Philadelphia	1	27398.820434570312
Philadelphia	2	7287.240234375
Philadelphia	4	116503.07043457031
Reggio Emilia	2	41509.94006347656
Reggio Emilia	3	56421.650390625
Reggio Emilia	4	44669.740478515625
Reims	1	52029.07043457031
Reims	2	18971.959716796875
Reims	3	15146.31982421875
Reims	4	48895.59014892578
Salzburg	2	98104.24005126953
Salzburg	3	6693.2802734375
Salzburg	4	45001.10986328125
San Diego	1	87489.23010253906
San Francisco	1	72899.19995117188
San Francisco	4	151459.4805908203
San Jose	2	160010.27026367188
San Rafael	1	267315.2586669922
San Rafael	2	7261.75
San Rafael	3	216297.40063476562
San Rafael	4	163983.64880371094
Sevilla	4	54723.621154785156

Singapore 1 28395.18994140625

Singapore 2 92033.77014160156

Singapore 3 90250.07995605469

Singapore 4 77809.37023925781

South Brisbane 1 21730.029907226562

South Brisbane 3 10640.290161132812

South Brisbane 4 27098.800048828125

Stavern 1 54701.999755859375

Stavern 4 61897.19006347656

Strasbourg 2 80438.47985839844

Torino 3 94117.25988769531

Toulouse 1 15139.1201171875

Toulouse 3 17251.08056640625

Toulouse 4 38098.240234375

Tsawassen 2 31302.500244140625

Tsawassen 3 43332.349609375

Vancouver 4 75238.91955566406

Versailles 1 5759.419921875

Versailles 4 59074.90026855469

White Plains 4 85555.98962402344

Time taken: 42.229 seconds, Fetched: 182 row(s)

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

hive> **select year, month from (select year_id as year, month_id as month, sum(quantityordered) as total_quantity, dense_rank() over(partition by year_id order by sum(quantityordered) desc) as r from sales_order_orc group by year_id, month_id)tab where r=1;**

Query ID = cloudera_20221002191818_f0f6cca5-ec5e-472c-b5a3-6c1d94fdd69d

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_1664762051968_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0010/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0010

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2022-10-02 19:19:09,263 Stage-1 map = 0%, reduce = 0%

2022-10-02 19:19:20,590 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.71 sec

2022-10-02 19:19:34,144 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.09 sec

MapReduce Total cumulative CPU time: 7 seconds 90 msec

Ended Job = job_1664762051968_0010

Launching Job 2 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_1664762051968_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1664762051968_0011/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1664762051968_0011

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2022-10-02 19:19:49,432 Stage-2 map = 0%, reduce = 0%

2022-10-02 19:19:58,321 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.1 sec

2022-10-02 19:20:11,706 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.36 sec

MapReduce Total cumulative CPU time: 6 seconds 360 msec

Ended Job = job_1664762051968_0011

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 7.09 sec   HDFS Read: 29519 HDFS Write: 792 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 6.36 sec   HDFS Read: 8673 HDFS Write: 23 SUCCESS

Total MapReduce CPU Time Spent: 13 seconds 450 msec

OK

year	month
2003	11
2004	11
2005	5

Time taken: 75.164 seconds, Fetched: 3 row(s)