Lab Exercise 06: Analysis the Supermarket Sales data using Hive Query Language

This exercise try to Analysis the Supermarket Sales data using Hive Query Language.

Step 01: Display Available Database

Step 02: Create Database as “SuperMarket”

Step 03: Use “SuperMarket” Database

Step 04: Display available Tables

```
hive> show databases;
OK
default
Time taken: 0.021 seconds, Fetched: 1 row(s)
hive> create database supermarket;
OK
Time taken: 0.072 seconds
hive> use supermarket;
OK
Time taken: 0.022 seconds
hive> show tables;
OK
Time taken: 0.05 seconds
```

Step 05: Create Table “Sales” with following scheme

Step 06: Copy ‘supermarket_sales.csv’ into Hadoop local (/home/cloudera/)

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/
Found 6 items
-rw-r--r--   1 cloudera cloudera        259 2022-08-31 23:30 /user/cloudera/customer.csv
-rw-r--r--   1 cloudera cloudera        259 2022-08-31 23:29 /user/cloudera/customer.txt
```

Step 07: Load ‘supermarket_sales.csv’ data into table ‘Sales’

```
hive> load data local inpath '/home/cloudera/supermarket.csv' into table sales;
Loading data to table default.sales
Table default.sales stats: [numFiles=1, totalSize=107804]
OK
Time taken: 1.075 seconds
hive>
```

Step 08: Display the content in table ‘Sales’

```
303-96-2227     B       Mandalay        Normal  Female  Home and lifestyle      97.38   10      48.69   1022.49 NULL    Ewallet 48.69   4.4
        NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL
727-02-1313     A       Yangon  Member  Male    Food and beverages      31.84   1       1.592   33.432  NULL    Cash    1.592   7.7
        NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL
347-56-2442     A       Yangon  Normal  Male    Home and lifestyle      65.82   1       3.291   69.111  NULL    Cash    3.291   4.1
        NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL
849-09-3807     A       Yangon  Member  Female  Fashion accessories     88.34   7       30.919  649.299 NULL    Cash    30.919  6.6
Time taken: 0.094 seconds, Fetched: 4000 row(s)
hive>
```

Q01: Display records who paid money through credit card?

```
Time taken: 0.982 seconds, Fetched: 2000 row(s)
hive> select * from sales where payment == 'Credit card';
OK
NULL    A       Yangon  Normal  Male    Home and lifestyle      46.33   7       16.2155 340.5255        NULL    Credit card     16.2155 7.4
NULL    A       Yangon  Member  Female  Health and beauty       36.26   2       3.626   76.146  NULL    Credit card     3.626   7.2
NULL    B       Mandalay        Member  Female  Food and beverages      54.84   3       8.226   172.746 NULL    Credit card     8.226   5.9
NULL    A       Yangon  Member  Female  Health and beauty       68.93   7       24.1255 506.6355        NULL    Credit card     24.1255 4.6
NULL    A       Yangon  Normal  Male    Sports and travel       72.61   6       21.783  457.443 NULL    Credit card     21.783  6.9
```

Q02: Show all Invoice ID which table entry in has more than 9 rating.

```
hive> select invoiceid from sales1 where rating>9;
OK
750-67-8428
226-31-3081
636-48-8204
145-94-9061
```

```
hive> select avg(grossincome) from sales;
Query ID = cloudera_20220901001818_682283f0-d198-4a67-ab34-6e0c570f5efb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):

2022-09-01 00:18:36,098 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.01 sec
MapReduce Total cumulative CPU time: 4 seconds 10 msec
Ended Job = job_1659505365280_0029
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.01 sec   HDFS Read: 117465 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 10 msec
OK
15.37936899703741
Time taken: 28.999 seconds, Fetched: 1 row(s)
```

Q03: Find Average sale amount in the data set.

Q04: Calculate total gross income in the data set.

```
hive> select sum(grossincome) from sales;
Query ID = cloudera_20220901033030_cc32b575-8801-4524-b456-f1413100ebb2
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1659505365280_0035, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1659505365280_0035/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1659505365280_0035
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-01 03:30:42,494 Stage-1 map = 0%,  reduce = 0%
2022-09-01 03:30:52,190 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.13 sec
2022-09-01 03:31:00,660 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.65 sec
MapReduce Total cumulative CPU time: 3 seconds 650 msec
Ended Job = job_1659505365280_0035
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.65 sec   HDFS Read: 117191 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 650 msec
OK
15379.36899703741
Time taken: 27.706 seconds, Fetched: 1 row(s)
```

Q05: Find total sale amount under 'Sports and travel' product line.

```
hive> select sum(grossincome) from sales1 where productline = 'Sports and travel';
Query ID = cloudera_20220901003434_4f53a4d2-2a3a-4240-825a-3ed7d3ae4746
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1659505365280_0031, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1659505365280_0031,
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1659505365280_0031
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-01 00:34:38,799 Stage-1 map = 0%,  reduce = 0%
2022-09-01 00:34:46,211 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.37 sec
2022-09-01 00:34:54,792 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.92 sec
MapReduce Total cumulative CPU time: 3 seconds 920 msec
Ended Job = job_1659505365280_0031
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.92 sec   HDFS Read: 226120 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 920 msec
OK
5249.793011307716
Time taken: 25.735 seconds, Fetched: 1 row(s)
```

Q06: Find maximum sales amount in which payment in ‘Ewallet’payment

```
hive> select max(grossincome)from sales1 where payment = 'Ewallet';
Query ID = cloudera_20220901003636_bf38932e-f6cb-4c8c-a81f-32f3794be4c5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1659505365280_0032, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1659505365280_0032/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1659505365280_0032
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-01 00:36:31,938 Stage-1 map = 0%,  reduce = 0%
2022-09-01 00:36:39,528 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.38 sec
2022-09-01 00:36:49,037 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.13 sec
MapReduce Total cumulative CPU time: 4 seconds 130 msec
Ended Job = job_1659505365280_0032
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.13 sec   HDFS Read: 225960 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 130 msec
OK
49.26
Time taken: 26.721 seconds, Fetched: 1 row(s)
```

Q07:Count number of sales which is has >5 quantity.

```
hive> select count(grossincome)from sales where quantity>5;
Query ID = cloudera_20220901003939_6fc9e2dc-ba2e-4c7b-89f9-572cdf0379bb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1659505365280_0033, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1659505365280_0033/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1659505365280_0033
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-01 00:39:29,737 Stage-1 map = 0%,  reduce = 0%
2022-09-01 00:39:37,302 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.29 sec
2022-09-01 00:39:45,758 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.03 sec
MapReduce Total cumulative CPU time: 4 seconds 30 msec
Ended Job = job_1659505365280_0033
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.03 sec   HDFS Read: 117817 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 30 msec
OK
496
Time taken: 25.423 seconds, Fetched: 1 row(s)
```

Q08: Find minimum and maximum sale total amount

```
hive> select min(grossincome) AS min_sales,max(grossincome) AS max_sales from sales1;
Query ID = cloudera_20220901025757_db219a9c-0b72-4851-95e4-a345fbc46400
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1659505365280_0034, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1659505365280_0034/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1659505365280_0034
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-09-01 02:57:36,765 Stage-1 map = 0%,  reduce = 0%
2022-09-01 02:57:44,113 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.47 sec
2022-09-01 02:57:51,496 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.05 sec
MapReduce Total cumulative CPU time: 3 seconds 50 msec
Ended Job = job_1659505365280_0034
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.05 sec   HDFS Read: 225515 HDFS Write: 13 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
0.5085  49.65
Time taken: 25.248 seconds, Fetched: 1 row(s)
```

Q09: List the product line which has more than 8 rating

```
hive> select productline,rating from sales where rating>8;
OK
Health and beauty       9.1
Electronic accessories  9.6
Health and beauty       8.4
Food and beverages      8.2
Food and beverages      8.6
Electronic accessories  9.9
Home and lifestyle      8.5
Food and beverages      9.6
Electronic accessories  9.5
Health and beauty       8.4
Electronic accessories  8.1
```

Q10: Display the records who has paid through cash and female customer.

```
hive> select * from sales1 where payment = 'Cash' and gender = 'Female';
OK
226-31-3081     C       Naypyitaw       Normal  Female  Electronic accessories  15.28   5       3.82    80.22   NULL    Cash    3.82    9.6
829-34-3910     A       Yangon  Normal  Female  Health and beauty       71.38   10      35.69   749.49  NULL    Cash    35.69   5.7
299-46-1805     B       Mandalay        Member  Female  Sports and travel       93.72   6       28.116  590.436 NULL    Cash    28.116  4.5
145-94-9061     B       Mandalay        Normal  Female  Food and beverages      88.36   5       22.09   463.89  NULL    Cash    22.09   9.6
354-25-5821     B       Mandalay        Member  Female  Sports and travel       69.12   6       20.736  435.456 NULL    Cash    20.736  5.6
228-96-1411     C       Naypyitaw       Member  Female  Food and beverages      98.7    8       39.48   829.08  NULL    Cash    39.48   7.6
132-32-9879     B       Mandalay        Member  Female  Electronic accessories  93.96   4       18.792  394.632 NULL    Cash    18.792  9.5
162-48-8011     A       Yangon  Member  Female  Food and beverages      44.59   5       11.1475 234.0975        NULL    Cash    11.1475 8.5
399-46-5918     C       Naypyitaw       Normal  Female  Electronic accessories  85.98   8       34.392  722.232 NULL    Cash    34.392  8.2
```