Bharathwin MA

205229105

Big Data Management and Analytics Lab Report

Q01: Display records who paid money through credit card?

```
hive> select * from Sales WHERE Payment='Credit card';
OK
631-41-3108     A     Yangon  Normal  Male    Home and lifestyle      46.33   7       16.2155 340.5255        03-03-2019      Credit card     16.2155 7.4
665-32-9167     A     Yangon  Member  Female  Health and beauty       36.26   2       3.626   76.146  01-10-2019      Credit card     3.626   7.2
692-92-5582     B     Mandalay        Member  Female  Food and beverages      54.84   3       8.226   172.746 2/20/2019       Credit card     8.226   5.9
656-95-9349     A     Yangon  Member  Female  Health and beauty       68.93   7       24.1255 506.6355        03-11-2019      Credit card     24.1255 4.6
765-26-6951     A     Yangon  Normal  Male    Sports and travel       72.61   6       21.783  457.443 01-01-2019      Credit card     21.783  6.9
329-62-1586     A     Yangon  Normal  Male    Food and beverages      54.67   3       8.2005  172.2105        1/21/2019       Credit card     8.2005  8.6
273-16-6619     B     Mandalay        Normal  Male    Home and lifestyle      33.2    2       3.32    69.72   3/15/2019       Credit card     3.32    4.4
227-03-5010     A     Yangon  Member  Female  Home and lifestyle      52.59   8       21.036  441.756 3/22/2019       Credit card     21.036  8.5
189-17-4241     A     Yangon  Normal  Female  Fashion accessories     87.67   2       8.767   184.107 03-10-2019      Credit card     8.767   7.7
871-79-8483     B     Mandalay        Normal  Male    Fashion accessories     94.13   5       23.5325 494.1825        2/25/2019       Credit card     23.5325 4.8
595-11-5460     A     Yangon  Normal  Male    Health and beauty       96.58   2       9.658   202.818 3/15/2019       Credit card     9.658   5.1
370-41-7321     B     Mandalay        Member  Male    Health and beauty       56.69   9       25.5105 535.7205        2/27/2019       Credit card     25.5105 8.4
669-54-1719     B     Mandalay        Member  Male    Electronic accessories  18.93   6       5.679   119.259 02-10-2019      Credit card     5.679   8.1
778-71-5554     C     Naypyitaw       Member  Male    Fashion accessories     15.43   1       0.7715  16.2015 1/25/2019       Credit card     0.7715  6.1
285-68-5083     C     Naypyitaw       Member  Female  Sports and travel       24.74   3       3.711   77.931  2/15/2019       Credit card     3.711   10.0
199-75-8169     A     Yangon  Member  Male    Sports and travel       15.81   10      7.905   166.005 03-06-2019      Credit card     7.905   8.6
393-65-2792     C     Naypyitaw       Normal  Male    Food and beverages      89.48   10      44.74   939.54  01-06-2019      Credit card     44.74   9.6
732-94-0499     C     Naypyitaw       Normal  Female  Electronic accessories  41.65   10      20.825  437.325 1/13/2019       Credit card     20.825  5.4
263-10-3913     C     Naypyitaw       Member  Male    Fashion accessories     49.04   9       22.068  463.428 01-09-2019      Credit card     22.068  8.6
381-20-0914     A     Yangon  Member  Female  Fashion accessories     20.01   9       9.0045  189.0945        01-12-2019      Credit card     9.0045  5.7
870-72-4431     C     Naypyitaw       Normal  Female  Health and beauty       99.19   6       29.757  624.897 1/21/2019       Credit card     29.757  5.5
787-56-0757     C     Naypyitaw       Member  Female  Food and beverages      80.36   4       16.072  337.512 2/23/2019       Credit card     16.072  8.3
633-44-8566     A     Yangon  Member  Male    Food and beverages      49.38   7       17.283  362.943 3/27/2019       Credit card     17.283  7.3
152-08-9985     B     Mandalay        Member  Male    Health and beauty       64.36   9       28.962  608.202 03-12-2019      Credit card     28.962  8.6
512-91-0811     C     Naypyitaw       Normal  Male    Health and beauty       89.75   1       4.4875  94.2375 02-06-2019      Credit card     4.4875  6.6
628-34-3388     C     Naypyitaw       Normal  Male    Fashion accessories     27.38   6       8.214   172.494 01-05-2019      Credit card     8.214   7.9
210-67-5886     C     Naypyitaw       Member  Female  Health and beauty       98.21   3       14.7315 309.3615        02-05-2019      Credit card     14.7315 7.8
237-01-6122     C     Naypyitaw       Member  Female  Home and lifestyle      80.79   9       36.3555 763.4655        02-01-2019      Credit card     36.3555 9.5
225-98-1496     C     Naypyitaw       Member  Female  Fashion accessories     27.02   3       4.053   85.113  03-02-2019      Credit card     4.053   7.1
638-60-7125     A     Yangon  Normal  Female  Electronic accessories  99.56   8       39.824  836.304 2/14/2019       Credit card     39.824  5.2
659-36-1684     C     Naypyitaw       Member  Male    Sports and travel       57.12   7       19.992  419.832 01-12-2019      Credit card     19.992  6.5
219-22-9386     B     Mandalay        Member  Male    Sports and travel       99.96   9       44.982  944.622 03-09-2019      Credit card     44.982  4.2
336-78-2147     C     Naypyitaw       Member  Male    Home and lifestyle      63.91   8       25.564  536.844 3/13/2019       Credit card     25.564  4.6
668-90-8900     A     Yangon  Normal  Female  Home and lifestyle      93.69   7       32.7915 688.6215        03-10-2019      Credit card     32.7915 4.5
189-08-9157     C     Naypyitaw       Normal  Female  Fashion accessories     31.73   9       14.2785 299.8485        01-08-2019      Credit card     14.2785 5.9
590-83-4591     B     Mandalay        Member  Male    Electronic accessories  87.45   6       26.235  550.935 2/17/2019       Credit card     26.235  8.8
700-81-1757     A     Yangon  Normal  Female  Electronic accessories  26.31   5       6.5775  138.1275        1/18/2019       Credit card     6.5775  8.8
```

hive> select * from sales where payment='Credit card';

OK

631-41-3108   A   Yangon Normal Male   Home and lifestyle   46.33   7   16.2155
    340.5255   03-03-2019   Credit card   16.2155   7.4

665-32-9167   A   Yangon Member   Female Health and beauty   36.26   2   3.626
    76.146 01-10-2019   Credit card   3.626   7.2

692-92-5582   B   Mandalay   Member   Female Food and beverages   54.84   3
    8.226   172.746   2/20/2019   Credit card   8.226   5.9

656-95-9349   A   Yangon Member   Female Health and beauty   68.93   7
    24.1255   506.6355   03-11-2019   Credit card   24.1255   4.6

765-26-6951   A   Yangon Normal Male   Sports and travel   72.61   6   21.783
    457.443   01-01-2019   Credit card   21.783 6.9

329-62-1586   A   Yangon Normal Male   Food and beverages   54.67   3   8.2005
    172.2105   1/21/2019   Credit card   8.2005 8.6

273-16-6619   B   Mandalay   Normal Male   Home and lifestyle   33.2   2   3.32
    69.72   3/15/2019   Credit card   3.32   4.4

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

```
hive> select InvoiceID from sales where rating > 9;
OK
750-67-8428
226-31-3081
636-48-8204
145-94-9061
132-32-9879
326-78-5178
120-06-4233
285-68-5083
347-34-2234
109-28-2512
393-65-2792
287-21-9091
362-58-8315
865-92-6136
212-62-1842
861-77-0145
237-01-6122
354-39-5160
575-30-8091
338-65-2210
458-41-1477
685-64-1609
214-17-6927
400-89-4171
307-85-2293
423-57-2993
316-55-4634
608-27-6295
175-54-2529
342-65-4817
400-60-7251
286-43-6208
831-07-6050
426-39-2418
672-51-8681
389-25-3394
```

hive> select InvoiceID from sales where rating > 9;

OK

750-67-8428

226-31-3081

636-48-8204

145-94-9061

132-32-9879

326-78-5178

120-06-4233

285-68-5083

347-34-2234

109-28-2512

393-65-2792

Time taken: 0.332 seconds, Fetched: 151 row(s)

Q03: Find Average sale amount in the data set.

```
hive> select avg(total) from sales;
Query ID = cloudera_20211022030303_64ef5663-61ac-4be0-8354-14bfb263fcce
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_1634891597995_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:03:39,380 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:03:53,816 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.94 sec
2021-10-22 03:04:05,588 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.71 sec
MapReduce Total cumulative CPU time: 4 seconds 710 msec
Ended Job = job_1634891597995_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.71 sec   HDFS Read: 117241 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 710 msec
OK
322.9667492055893
Time taken: 63.357 seconds, Fetched: 1 row(s)
hive>
```

Q04: Calculate total gross income in the data set.

```
hive> select SUM(grossincome) from sales;
Query ID = cloudera_20211022031919_7d57fbc8-60f4-4ade-a0a7-83563cb94383
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_1634891597995_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:19:28,406 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:19:40,437 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.94 sec
2021-10-22 03:19:49,951 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.01 sec
MapReduce Total cumulative CPU time: 4 seconds 10 msec
Ended Job = job_1634891597995_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.01 sec   HDFS Read: 117050 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 10 msec
OK
15379.36899703741
Time taken: 33.279 seconds, Fetched: 1 row(s)
```

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

```
hive> select SUM(total) from sales WHERE Productline='Sports and travel';
Query ID = cloudera_20211022033131_bbbc9bcb-83d2-4e69-8b88-f55db357479e
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_1634891597995_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:31:33,979 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:31:42,606 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.31 sec
2021-10-22 03:31:52,128 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.43 sec
MapReduce Total cumulative CPU time: 4 seconds 430 msec
Ended Job = job_1634891597995_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.43 sec   HDFS Read: 118146 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 430 msec
OK
55122.82658100128
Time taken: 28.873 seconds, Fetched: 1 row(s)
hive>
```

Q06: Find maximum sales amount in which payment in 'Ewallet' payment:

```
hive> select max(total) from sales where payment='Ewallet';
Query ID = cloudera_20211022033333_c282eeca-d3f9-46b7-8129-db7ce7c69839
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_1634891597995_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:33:25,333 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:33:34,817 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.1 sec
2021-10-22 03:33:45,422 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.59 sec
MapReduce Total cumulative CPU time: 4 seconds 590 msec
Ended Job = job_1634891597995_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.59 sec   HDFS Read: 117902 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 590 msec
OK
1034.46
Time taken: 31.796 seconds, Fetched: 1 row(s)
hive>
```

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

```
hive> select count(total) from sales where Quantity > 5;
Query ID = cloudera_20211022033737_2ef18ff7-6330-43b4-bda2-211a1887458c
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_1634891597995_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:38:05,070 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:38:13,609 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.28 sec
2021-10-22 03:38:23,105 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.43 sec
MapReduce Total cumulative CPU time: 4 seconds 430 msec
Ended Job = job_1634891597995_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.43 sec   HDFS Read: 117674 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 430 msec
OK
496
Time taken: 28.897 seconds, Fetched: 1 row(s)
hive>
```

Q08: Find minimum and maximum sale total amount:

```
hive> select min(total) from sales;
Query ID = cloudera_20211022033939_26639f27-9871-403a-8755-ff4d5f46e537
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_1634891597995_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:39:57,780 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:40:06,283 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.76 sec
2021-10-22 03:40:16,783 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.71 sec
MapReduce Total cumulative CPU time: 3 seconds 710 msec
Ended Job = job_1634891597995_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.71 sec   HDFS Read: 116847 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 710 msec
OK
10.6785
Time taken: 28.622 seconds, Fetched: 1 row(s)
hive>
```

```
hive> select max(total) from sales;
Query ID = cloudera_20211022034141_abe86e21-d909-4c07-a09d-ee47adc68395
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_1634891597995_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1634891597995_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1634891597995_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-22 03:41:16,936 Stage-1 map = 0%,  reduce = 0%
2021-10-22 03:41:24,327 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.78 sec
2021-10-22 03:41:33,866 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.19 sec
MapReduce Total cumulative CPU time: 4 seconds 190 msec
Ended Job = job_1634891597995_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.19 sec   HDFS Read: 116847 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 190 msec
OK
1042.65
Time taken: 26.997 seconds, Fetched: 1 row(s)
hive>
```

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

```
hive> select Productline from sales where rating > 8;
OK
Health and beauty
Electronic accessories
Health and beauty
Food and beverages
Food and beverages
Electronic accessories
Home and lifestyle
Food and beverages
Electronic accessories
Health and beauty
Electronic accessories
Food and beverages
Food and beverages
Electronic accessories
Electronic accessories
Sports and travel
Sports and travel
Sports and travel
Fashion accessories
Food and beverages
Electronic accessories
Home and lifestyle
Fashion accessories
Food and beverages
Fashion accessories
Sports and travel
Health and beauty
Food and beverages
Food and beverages
Electronic accessories
Fashion accessories
```

hive> select Productline from sales where rating > 8;

OK

Health and beauty

Electronic accessories

Health and beauty

Food and beverages

Food and beverages

Electronic accessories

Home and lifestyle