E-Commerce Assignment

Submitted By: Tom Mathews

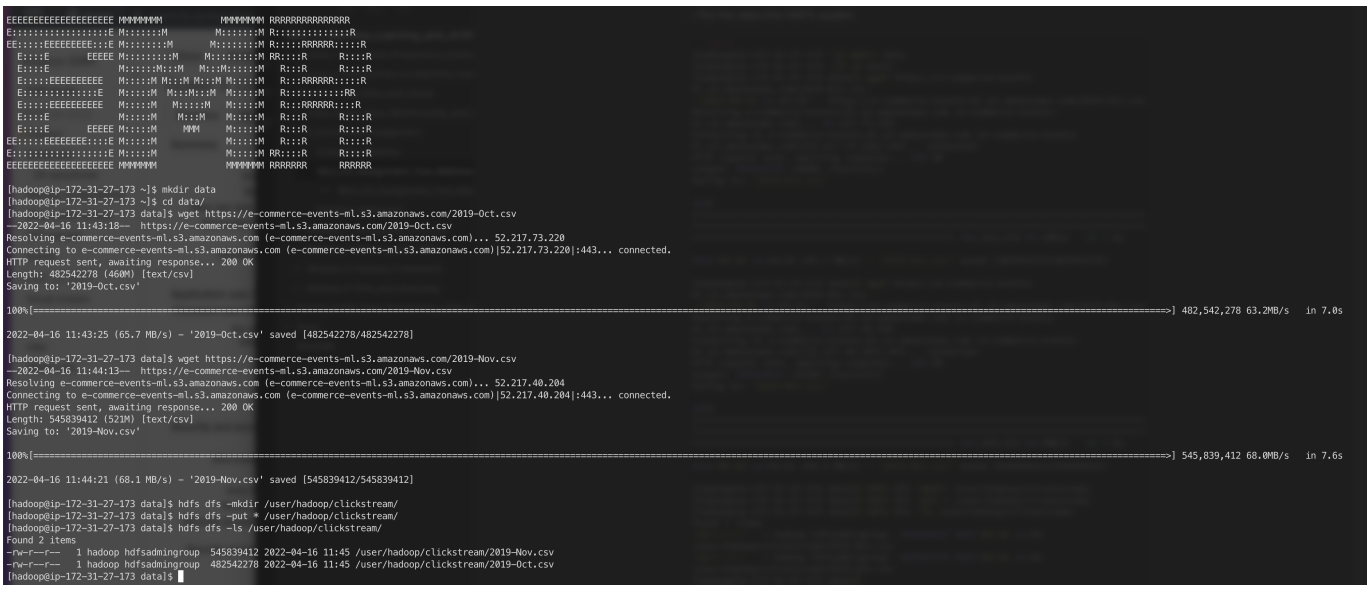
Setup

Get Data into HDFS System

- Download the data to EMR cluster.
- Put the data into HDFS system.

```
[hadoop@ip-172-31-27-173 ~]$ mkdir data
[hadoop@ip-172-31-27-173 ~]$ cd data/
[hadoop@ip-172-31-27-173 data]$ wget https://e-commerce-events-
ml.s3.amazonaws.com/2019-Oct.csv
--2022-04-16 11:43:18--  https://e-commerce-events-ml.s3.amazonaws.com/2019-
Oct.csv
Resolving e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-
ml.s3.amazonaws.com)... 52.217.73.220
Connecting to e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-
ml.s3.amazonaws.com)|52.217.73.220|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 482542278 (460M) [text/csv]
Saving to: '2019-Oct.csv'

100%
[==========================================================================
==========================================================================
==========================================================>] 482,542,278 63.2MB/s
in 7.0s

2022-04-16 11:43:25 (65.7 MB/s) - '2019-Oct.csv' saved [482542278/482542278]

[hadoop@ip-172-31-27-173 data]$ wget https://e-commerce-events-
ml.s3.amazonaws.com/2019-Nov.csv
--2022-04-16 11:44:13--  https://e-commerce-events-ml.s3.amazonaws.com/2019-
Nov.csv
Resolving e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-
ml.s3.amazonaws.com)... 52.217.40.204
Connecting to e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-
ml.s3.amazonaws.com)|52.217.40.204|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 545839412 (521M) [text/csv]
Saving to: '2019-Nov.csv'

100%
[==========================================================================
==========================================================================
==========================================================>] 545,839,412 68.0MB/s
in 7.6s

2022-04-16 11:44:21 (68.1 MB/s) - '2019-Nov.csv' saved [545839412/545839412]

[hadoop@ip-172-31-27-173 data]$ hdfs dfs -mkdir /user/hadoop/clickstream/
[hadoop@ip-172-31-27-173 data]$ hdfs dfs -put * /user/hadoop/clickstream/
[hadoop@ip-172-31-27-173 data]$ hdfs dfs -ls /user/hadoop/clickstream/
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup  545839412 2022-04-16 11:45
```

```
/user/hadoop/clickstream/2019-Nov.csv
-rw-r--r--   1 hadoop hdfsadmingroup  482542278 2022-04-16 11:45
/user/hadoop/clickstream/2019-Oct.csv
[hadoop@ip-172-31-27-173 data]$
```



Create Tables and Load Data to Hive

- Start Hive session using beeline.
- Create database and use the database.
- Create two tables for october and november data.
- Load data into the two tables from HDFS system.

```
[hadoop@ip-172-31-22-74 data]$ beeline -u jdbc:hive2://localhost:10000/default
-n hadoop
Connecting to jdbc:hive2://localhost:10000/default
Connected to: Apache Hive (version 2.3.9-amzn-0)
Driver: Hive JDBC (version 2.3.9-amzn-0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 2.3.9-amzn-0 by Apache Hive
0: jdbc:hive2://localhost:10000/default> create database clickstream_db;
No rows affected (3.735 seconds)
0: jdbc:hive2://localhost:10000/default> use clickstream_db;
No rows affected (0.143 seconds)

0: jdbc:hive2://localhost:10000/default> create table clickstream_oct
. . . . . .> (
. . . . . .> event_time string,
. . . . . .> event_type string,
. . . . . .> product_id string,
. . . . . .> category_id string,
. . . . . .> category_code string,
. . . . . .> brand string,
. . . . . .> price float,
. . . . . .> user_id bigint,
. . . . . .> user_session string
. . . . . .> )
. . . . . .> ROW FORMAT DELIMITED
. . . . . .> FIELDS TERMINATED BY ','
. . . . . .> LINES TERMINATED BY '\n'
. . . . . .>
tblproperties("skip.header.line.count"="1");
No rows affected (0.121 seconds)
0: jdbc:hive2://localhost:10000/default> create table clickstream_nov
. . . . . .> (
. . . . . .> event_time string,
. . . . . .> event_type string,
. . . . . .> product_id string,
. . . . . .> category_id string,
. . . . . .> category_code string,
. . . . . .> brand string,
. . . . . .> price float,
. . . . . .> user_id bigint,
. . . . . .> user_session string
. . . . . .> )
. . . . . .> ROW FORMAT DELIMITED
. . . . . .> FIELDS TERMINATED BY ','
. . . . . .> LINES TERMINATED BY '\n'
. . . . . .>
tblproperties("skip.header.line.count"="1");
```

```
No rows affected (0.147 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
+------------------+
|     tab_name     |
+------------------+
| clickstream_nov  |
| clickstream_oct  |
+------------------+
2 rows selected (0.189 seconds)
0: jdbc:hive2://localhost:10000/default> LOAD DATA INPATH
'/user/hadoop/clickstream/2019-Oct.csv' INTO TABLE clickstream_oct;
No rows affected (0.426 seconds)
0: jdbc:hive2://localhost:10000/default> LOAD DATA INPATH
'/user/hadoop/clickstream/2019-Nov.csv' INTO TABLE clickstream_nov;
No rows affected (0.456 seconds)
0: jdbc:hive2://localhost:10000/default> select * from clickstream_oct limit 5;
+-----------------------------+-----------------------------+-----------------
----------+-------------------------------+---------------------------------+---
--------------------+-----------------------+--------------------------+-----
-----------------------------------+
| clickstream_oct.event_time  | clickstream_oct.event_type  |
clickstream_oct.product_id  | clickstream_oct.category_id  |
clickstream_oct.category_code  | clickstream_oct.brand  | clickstream_oct.price
| clickstream_oct.user_id  |     clickstream_oct.user_session      |
+-----------------------------+-----------------------------+-----------------
----------+-------------------------------+---------------------------------+---
--------------------+-----------------------+--------------------------+-----
-----------------------------------+
| 2019-10-01 00:00:00 UTC     | cart                        | 5773203
| 1487580005134238553         |                               | runail
| 2.62                  | 463240011               | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  |
| 2019-10-01 00:00:03 UTC     | cart                        | 5773353
| 1487580005134238553         |                               | runail
| 2.62                  | 463240011               | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  |
| 2019-10-01 00:00:07 UTC     | cart                        | 5881589
| 2151191071051219817         |                               | lovely
| 13.48                 | 429681830               | 49e8d843-adf3-428b-a2c3-
fe8bc6a307c9  |
| 2019-10-01 00:00:07 UTC     | cart                        | 5723490
| 1487580005134238553         |                               | runail
| 2.62                  | 463240011               | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  |
| 2019-10-01 00:00:15 UTC     | cart                        | 5881449
| 1487580013522845895         |                               | lovely
| 0.56                  | 429681830               | 49e8d843-adf3-428b-a2c3-
fe8bc6a307c9  |
```

```
+----------------------------+----------------------------+-----------------
----------+-----------------------------+-------------------------------+---
--------------------+-----------------------+--------------------------+-----
---------------------------------+
5 rows selected (0.335 seconds)
0: jdbc:hive2://localhost:10000/default> select * from clickstream_nov limit 5;
+----------------------------+----------------------------+-----------------
----------+-----------------------------+-------------------------------+---
--------------------+-----------------------+--------------------------+-----
---------------------------------+
| clickstream_nov.event_time  | clickstream_nov.event_type  |
clickstream_nov.product_id  | clickstream_nov.category_id  |
clickstream_nov.category_code  | clickstream_nov.brand  | clickstream_nov.price
| clickstream_nov.user_id  |     clickstream_nov.user_session      |
+----------------------------+----------------------------+-----------------
----------+-----------------------------+-------------------------------+---
--------------------+-----------------------+--------------------------+-----
---------------------------------+
| 2019-11-01 00:00:02 UTC     | view                       | 5802432
| 1487580009286598681          |                              |
| 0.32                  | 562076640               | 09fafd6c-6c99-46b1-834f-
33527f4de241  |
| 2019-11-01 00:00:09 UTC     | cart                       | 5844397
| 1487580006317032337          |                              |
| 2.38                  | 553329724               | 2067216c-31b5-455d-a1cc-
af0575a34ffb  |
| 2019-11-01 00:00:10 UTC     | view                       | 5837166
| 1783999064103190764          |                              | pnb
| 22.22                 | 556138645               | 57ed222e-a54a-4907-9944-
5a875c2d7f4f  |
| 2019-11-01 00:00:11 UTC     | cart                       | 5876812
| 1487580010100293687          |                              | jessnail
| 3.16                  | 564506666               | 186c1951-8052-4b37-adce-
dd9644b1d5f7  |
| 2019-11-01 00:00:24 UTC     | remove_from_cart           | 5826182
| 1487580007483048900          |                              |
| 3.33                  | 553329724               | 2067216c-31b5-455d-a1cc-
af0575a34ffb  |
+----------------------------+----------------------------+-----------------
----------+-----------------------------+-------------------------------+---
--------------------+-----------------------+--------------------------+-----
---------------------------------+
5 rows selected (0.217 seconds)
```

Create Merged Table

- Create new table with one extra column "month".

- Insert data to new table from above tables and evaluate the data for month column.

```
0: jdbc:hive2://localhost:10000/default> create table clickstream
. . . . . .> (
. . . . . .> event_time string,
. . . . . .> event_type string,
. . . . . .> product_id string,
. . . . . .> category_id string,
. . . . . .> category_code string,
. . . . . .> brand string,
. . . . . .> price float,
. . . . . .> user_id bigint,
. . . . . .> user_session string,
. . . . . .> month int
. . . . . .> );
No rows affected (0.12 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
+------------------+
|     tab_name     |
+------------------+
| clickstream      |
| clickstream_nov  |
| clickstream_oct  |
+------------------+
3 rows selected (0.058 seconds)
0: jdbc:hive2://localhost:10000/default> insert into clickstream select *,
month(event_time) month from clickstream_oct;
No rows affected (43.944 seconds)
0: jdbc:hive2://localhost:10000/default> insert into clickstream select *,
month(event_time) month from clickstream_nov;
No rows affected (35.121 seconds)
0: jdbc:hive2://localhost:10000/default> select * from clickstream limit 5;
+--------------------------+-------------------------+-------------------------
+--------------------------+----------------------------+--------------------+-
------------------+----------------------+-----------------------------------
---+--------------------+
|  clickstream.event_time  | clickstream.event_type  | clickstream.product_id
| clickstream.category_id  | clickstream.category_code  | clickstream.brand  |
clickstream.price  | clickstream.user_id  |       clickstream.user_session
| clickstream.month  |
+--------------------------+-------------------------+-------------------------
+--------------------------+----------------------------+--------------------+-
------------------+----------------------+-----------------------------------
---+--------------------+
| 2019-10-01 00:00:00 UTC  | cart                    | 5773203
| 1487580005134238553      |                            | runail             |
2.62               | 463240011            | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  | 10                 |
| 2019-10-01 00:00:03 UTC  | cart                    | 5773353
```

```
| 1487580005134238553      |                                  | runail              |
2.62              | 463240011           | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  | 10                |
| 2019-10-01 00:00:07 UTC  | cart                    | 5881589
| 2151191071051219817      |                                  | lovely              |
13.48             | 429681830           | 49e8d843-adf3-428b-a2c3-
fe8bc6a307c9  | 10                |
| 2019-10-01 00:00:07 UTC  | cart                    | 5723490
| 1487580005134238553      |                                  | runail              |
2.62              | 463240011           | 26dd6e6e-4dac-4778-8d2c-
92e149dab885  | 10                |
| 2019-10-01 00:00:15 UTC  | cart                    | 5881449
| 1487580013522845895      |                                  | lovely              |
0.56              | 429681830           | 49e8d843-adf3-428b-a2c3-
fe8bc6a307c9  | 10                |
+--------------------------+-------------------------+-----------------------
+--------------------------+---------------------------+-------------------+-
------------------+---------------------+-----------------------------------
---+--------------------+
5 rows selected (0.434 seconds)
0: jdbc:hive2://localhost:10000/default> select count(*) from clickstream;
+----------+
|   _c0    |
+----------+
| 8738120  |
+----------+
1 row selected (0.163 seconds)
```

Create Table in Parquet Format

- Create new table stored in parquet format with partitions on "event_type" and "month columns".
- Insert data from the "clickstream" table into "clickstream_parq" table.

```
0: jdbc:hive2://localhost:10000/default> create table clickstream_parq
. . . . . . . . . . . . . . . . . . . . .> (
. . . . . . . . . . . . . . . . . . . . .> event_time string,
. . . . . . . . . . . . . . . . . . . . .> product_id string,
. . . . . . . . . . . . . . . . . . . . .> category_id string,
. . . . . . . . . . . . . . . . . . . . .> category_code string,
. . . . . . . . . . . . . . . . . . . . .> brand string,
. . . . . . . . . . . . . . . . . . . . .> price float,
. . . . . . . . . . . . . . . . . . . . .> user_id bigint,
. . . . . . . . . . . . . . . . . . . . .> user_session string
. . . . . . . . . . . . . . . . . . . . .> )
. . . . . . . . . . . . . . . . . . . . .> partitioned by (event_type string,
month int)
. . . . . . . . . . . . . . . . . . . . .> stored as parquet;
No rows affected (0.101 seconds)
0: jdbc:hive2://localhost:10000/default> describe clickstream_parq;
+--------------------------+----------------------+----------------------+
|         col_name         |       data_type      |        comment       |
+--------------------------+----------------------+----------------------+
| event_time               | string               |                      |
| product_id               | string               |                      |
| category_id              | string               |                      |
| category_code            | string               |                      |
| brand                    | string               |                      |
| price                    | float                |                      |
| user_id                  | bigint               |                      |
| user_session             | string               |                      |
| event_type               | string               |                      |
| month                    | int                  |                      |
|                          | NULL                 | NULL                 |
| # Partition Information  | NULL                 | NULL                 |
| # col_name               | data_type            | comment              |
|                          | NULL                 | NULL                 |
| event_type               | string               |                      |
| month                    | int                  |                      |
+--------------------------+----------------------+----------------------+
16 rows selected (0.106 seconds)
0: jdbc:hive2://localhost:10000/default> set hive.exec.dynamic.partition=true;
No rows affected (0.012 seconds)
0: jdbc:hive2://localhost:10000/default> set
hive.exec.dynamic.partition.mode=nonstrict;
No rows affected (0.006 seconds)
0: jdbc:hive2://localhost:10000/default> insert into table clickstream_parq
. . . . . . . . . . . . . . . . . . . . .> partition(event_type, month)
. . . . . . . . . . . . . . . . . . . . .> select event_time, product_id,
category_id, category_code, brand, price, user_id, user_session, event_type,
month from clickstream;
```

```
No rows affected (132.432 seconds)
0: jdbc:hive2://localhost:10000/default> select * from clickstream_parq limit
5;
+------------------------------+------------------------------+---------------
--------------+----------------------------------+-------------------------+---
---------------------+---------------------------+---------------------------
----------+------------------------------+-------------------------+
| clickstream_parq.event_time  | clickstream_parq.product_id  |
clickstream_parq.category_id  | clickstream_parq.category_code  |
clickstream_parq.brand  | clickstream_parq.price  | clickstream_parq.user_id  |
clickstream_parq.user_session     | clickstream_parq.event_type  |
clickstream_parq.month  |
+------------------------------+------------------------------+---------------
--------------+----------------------------------+-------------------------+---
---------------------+---------------------------+---------------------------
----------+------------------------------+-------------------------+
| 2019-10-25 22:25:07 UTC      | 5858562                      |
1658462125284131265           |                                  | uskusi
| 3.43                  | 519746611                | 38f808c1-429c-d338-
107c-905c75d68f44  | cart                         | 10                       |
| 2019-10-26 21:31:36 UTC      | 5694155                      |
1487580007952810971           |                                  | yoko
| 17.78                 | 503106786                | 64e1c679-7c64-4c3a-
8fce-e1f7146cadf1  | cart                         | 10                       |
| 2019-10-26 11:22:26 UTC      | 5723406                      |
1487580012927254698           |                                  |
| 14.29                 | 564348885                | 75a43dc8-2be6-4036-
965e-7ff2b2140fd4  | cart                         | 10                       |
| 2019-10-25 22:25:00 UTC      | 5867962                      |
1487580005671109489           |                                  | masura
| 3.95                  | 564200530                | ea04a4b2-7d5c-4e0e-
9a44-6edc451a1f7b  | cart                         | 10                       |
| 2019-10-26 11:22:27 UTC      | 5737912                      |
1487580008984608779           |                                  | kosmekka
| 10.79                 | 564343710                | 0fc46e11-ff20-459d-
ab5b-fe85625351df  | cart                         | 10                       |
+------------------------------+------------------------------+---------------
--------------+----------------------------------+-------------------------+---
---------------------+---------------------------+---------------------------
----------+------------------------------+-------------------------+
5 rows selected (1.454 seconds)
0: jdbc:hive2://localhost:10000/default> select count(*) from clickstream_parq;
+----------+
|   _c0    |
+----------+
| 8738120  |
+----------+
1 row selected (0.219 seconds)
```

```
0: jdbc:hive2://localhost:10000/default> select count(*) from clickstream;
+----------+
|   _c0    |
+----------+
| 8738120  |
+----------+
1 row selected (0.454 seconds)
```

```
0: jdbc:hive2://localhost:10000/default> create table clickstream_parq
. . . . . .> (
. . . . . .> event_time string,
. . . . . .> product_id string,
. . . . . .> category_id string,
. . . . . .> category_code string,
. . . . . .> brand string,
. . . . . .> price float,
. . . . . .> user_id bigint,
. . . . . .> user_session string
. . . . . .> )
. . . . . .> partitioned by (event_type string, month int)
. . . . . .> stored as parquet;
No rows affected (0.127 seconds)
0: jdbc:hive2://localhost:10000/default> describe clickstream_parq;
+-------------------------+-----------------------+-----------------------+
|        col_name         |       data_type       |        comment        |
+-------------------------+-----------------------+-----------------------+
| event_time              | string                |                       |
| product_id              | string                |                       |
| category_id             | string                |                       |
| category_code           | string                |                       |
| brand                   | string                |                       |
| price                   | float                 |                       |
| user_id                 | bigint                |                       |
| user_session            | string                |                       |
| event_type              | string                |                       |
| month                   | int                   |                       |
|                         | NULL                  | NULL                  |
| # Partition Information | NULL                  | NULL                  |
| # col_name              | data_type             | comment               |
|                         | NULL                  | NULL                  |
| event_type              | string                |                       |
| month                   | int                   |                       |
+-------------------------+-----------------------+-----------------------+
16 rows selected (0.158 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Question 1

Qn: Find the total revenue generated due to the purchases made in October.

Ans:

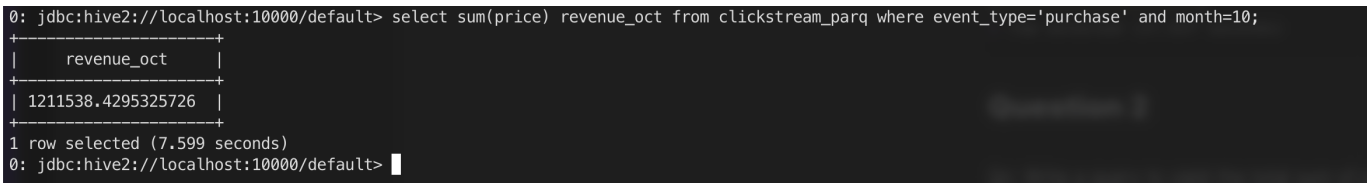
Query:

```
select sum(price) revenue_oct from clickstream_parq where event_type='purchase'
and month=10;
```

Output:

```
+---------------------+
|     revenue_oct     |
+---------------------+
| 1211538.4295325726  |
+---------------------+
1 row selected (7.599 seconds)
```

Screenshot:



Question 2

Qn: Write a query to yield the total sum of purchases per month in a single output.

Ans:

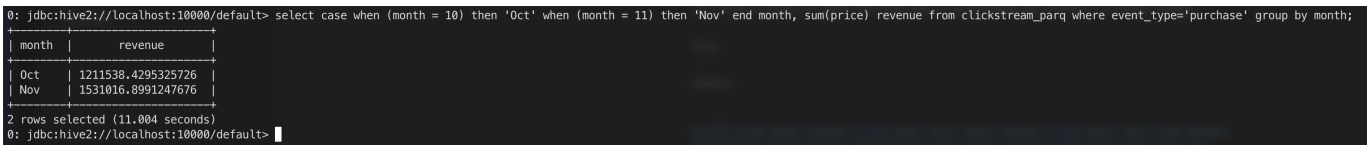
Query:

```
select case when (month = 10) then 'Oct' when (month = 11) then 'Nov' end
month, sum(price) revenue from clickstream_parq where event_type='purchase'
group by month;
```

Output:

```
+--------+---------------------+
| month  |       revenue       |
+--------+---------------------+
| Oct    | 1211538.4295325726  |
| Nov    | 1531016.8991247676  |
+--------+---------------------+
2 rows selected (11.004 seconds)
```

Screenshot:



Question 3

Qn: Write a query to find the change in the revenue generated due to purchases made from October to November.

Ans:

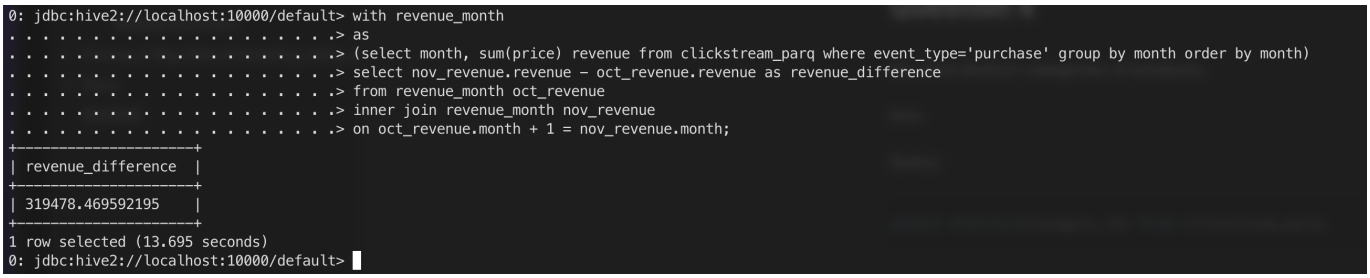
Query:

```
with revenue_month
as
(select month, sum(price) revenue from clickstream_parq where
event_type='purchase' group by month order by month)
select nov_revenue.revenue - oct_revenue.revenue as revenue_difference
from revenue_month oct_revenue
inner join revenue_month nov_revenue
on oct_revenue.month + 1 = nov_revenue.month;
```

Output:

```
+----------------------+
| revenue_difference   |
+----------------------+
| 319478.469592195     |
+----------------------+
1 row selected (13.695 seconds)
```

Screenshot:



Question 4

Qn: Find distinct categories of products.

Ans:

Query:

```
select distinct(category_id) from clickstream_parq;
```

Output:

category_id
1487580004882580302
1487580004916134735
1487580005025186644
1487580005134238553
1487580005176181595
1487580005293622112
1487580005318787937
1487580005343953762
1487580005461394279
1487580005486560104
1487580005570446188
1487580005595612013
1487580005629166447
1487580005687886706
1487580005754995573
1487580005855658874
1487580005880824699
1487580005998265217
1487580006015042434
1487580006073762693
1487580006098928518
1487580006174425994
1487580006199591819
1487580006216369036
1487580006300255120
1487580006317032337
1487580006509970331
1487580006526747548
1487580006551913373
1487580006585467807
1487580006644188066
1487580006711296933
1487580006820348840
1487580006845514665
1487580006870680490
1487580006895846315
1487580007004898224
1487580007139115958
1487580007172670392
1487580007189447609
1487580007214613434
1487580007256556476
1487580007281722301
1487580007399162817

14875800074327172 50
14875800075921008 09
14875800076592096 76
14875800076759868 93
14875800077011527 18
14875800077179299 35
14875800078353704 53
14875800078521476 70
14875800078773134 95
14875800078940907 12
14875800079360337 54
14875800079528109 71
14875800080115312 30
14875800080283084 47
14875800080702514 89
14875800080870287 06
14875800081625261 82
14875800082464122 66
14875800084477388 66
14875800084729046 91
14875800085232363 41
14875800085903452 08
14875800086574540 75
14875800086742312 92
14875800087581163 52
14875800087748935 69
14875800089091113 03
14875800089258885 20
14875800089510543 45
14875800089846087 79
14875800090517176 46
14875800091188265 13
14875800092614328 56
14875800092865986 81
14875800093117645 06
14875800093369303 31
14875800093620961 56
14875800094459822 39
14875800096473088 39
14875800096724746 64
14875800097395835 30
14875800097983037 88
14875800098234696 13
14875800098905785 79
14875800099744645 62
14875800101002936 87
14875800101254595 12
14875800101925683 79

```
| 1487580010268065854  |
| 1487580010318397504  |
| 1487580010335174721  |
| 1487580010360340546  |
| 1487580010427449413  |
| 1487580010561667147  |
| 1487580010645553231  |
| 1487580010662330448  |
+----------------------+
|     category_id      |
+----------------------+
| 1487580010695884882  |
| 1487580010737827924  |
| 1487580010779770966  |
| 1487580010821714008  |
| 1487580010846879833  |
| 1487580010872045658  |
| 1487580010897211483  |
| 1487580010922377308  |
| 1487580010955931741  |
| 1487580010989486174  |
| 1487580011031429216  |
| 1487580011140481125  |
| 1487580011157258342  |
| 1487580011182424167  |
| 1487580011199201384  |
| 1487580011283087468  |
| 1487580011308253293  |
| 1487580011333419118  |
| 1487580011408916594  |
| 1487580011425693811  |
| 1487580011476025461  |
| 1487580011517968503  |
| 1487580011534745720  |
| 1487580011585077370  |
| 1487580011652186237  |
| 1487580011702517887  |
| 1487580011786403970  |
| 1487580011828347011  |
| 1487580011853512836  |
| 1487580011903844485  |
| 1487580012021285001  |
| 1487580012071616651  |
| 1487580012172279951  |
| 1487580012205834384  |
| 1487580012432326804  |
| 1487580012507824279  |
| 1487580012549767321  |
```

| 1487580012574933146 |
| 1487580012591710363 |
| 1487580012642042013 |
| 1487580012683985055 |
| 1487580012734316706 |
| 1487580012759482531 |
| 1487580012809814181 |
| 1487580012876923048 |
| 1487580012927254698 |
| 1487580012952420523 |
| 1487580013086638258 |
| 1487580013145358517 |
| 1487580013170524342 |
| 1487580013212467384 |
| 1487580013296353468 |
| 1487580013321519293 |
| 1487580013338296510 |
| 1487580013430571202 |
| 1487580013539623112 |
| 1487580013615120588 |
| 1487580013657063630 |
| 1487580013690618064 |
| 1487580013782892757 |
| 1487580013824835799 |
| 1487580013841613016 |
| 1487580013858390233 |
| 1487580013900333275 |
| 1487580014009385185 |
| 1487580014042939619 |
| 1495705810754339379 |
| 1511892746070131099 |
| 1525995662934540829 |
| 1543705961326182546 |
| 1588205336064425989 |
| 1597770225539875791 |
| 1605161575889502297 |
| 1645114480121610699 |
| 1648815651034235876 |
| 1660394382047576308 |
| 1715102762414375164 |
| 1720400165430363096 |
| 1725504706412807026 |
| 1752742610406999036 |
| 1752742615205281895 |
| 1759024699007828783 |
| 1761186209054327497 |
| 1783999063574708423 |
| 1783999064036081896 |

| 1783999064103190764 |
| 1783999064136745198 |
| 1783999067156644376 |
| 1783999068909863670 |
| 1783999072407912622 |
| 1783999073792033086 |
| 1783999076535108182 |
| 1791442703716712702 |
| 1797122099068797194 |
| 1805953852441101266 |
| 1810470908326838736 |
| 1814592470537732872 |
| 1842735764450837316 |
| 1842735764509557576 |
| 1842735768846467167 |

category_id
1858245586344477369
1889472915104072007
1891434214553813877
1897124478404526487
1897124495995438074
1911999801088541491
1911999801642189621
1911999884991397970
1911999948073730325
1913479463425802701
1921723506584715388
1924049106385240809
1924049110428549877
1926797403503985079
1977575787259232875
1977786601392047073
1987724780047958178
2022516588854378951
2022622168218599898
2029731308699124089
2060156961931919712
2069804424703771380
2071303198680810125
2084144451428549153
2089259162625114209
2093602042093240877
2094448780651791052
2106514244437541443
2106514244487873093
2114584564549550293

| 2115334439910245200 |
| 2134354342373753638 |
| 2134354356349173879 |
| 2145935122136826354 |
| 2151191059827262021 |
| 2151191070908613477 |
| 2151191070984110951 |
| 2151191071378375538 |
| 2151191075757228942 |
| 2166295400451933025 |
| 2193074740552270669 |
| 2193074740686488401 |
| 2195085255034011676 |
| 2195085258272014535 |
| 1487580004832248652 |
| 1487580004857414477 |
| 1487580004966466385 |
| 1487580004983243602 |
| 1487580005008409427 |
| 1487580005050352469 |
| 1487580005067129686 |
| 1487580005092295511 |
| 1487580005268456287 |
| 1487580005369119587 |
| 1487580005385896804 |
| 1487580005411062629 |
| 1487580005427839846 |
| 1487580005511725929 |
| 1487580005528503146 |
| 1487580005553668971 |
| 1487580005654332272 |
| 1487580005671109489 |
| 1487580005713052531 |
| 1487580005796938615 |
| 1487580005897601916 |
| 1487580005922767741 |
| 1487580005939544958 |
| 1487580006056985476 |
| 1487580006132482952 |
| 1487580006157648777 |
| 1487580006241534861 |
| 1487580006350586771 |
| 1487580006409307030 |
| 1487580006434472855 |
| 1487580006451250072 |
| 1487580006484804506 |
| 1487580006610633632 |
| 1487580006627410849 |

```
| 1487580006744851367  |
| 1487580006937789357  |
| 1487580006979732399  |
| 1487580007021675441  |
| 1487580007046841266  |
| 1487580007072007091  |
| 1487580007097172916  |
| 1487580007113950133  |
| 1487580007306888126  |
| 1487580007365608384  |
| 1487580007457883075  |
| 1487580007483048900  |
| 1487580007508214725  |
| 1487580007524991942  |
| 1487580007550157767  |
| 1487580007575323592  |
| 1487580007634043851  |
| 1487580007759872977  |
| 1487580007776650194  |
| 1487580007910867929  |
| 1487580007986365405  |
| 1487580008053474272  |
+----------------------+
|     category_id      |
+----------------------+
| 1487580008112194531  |
| 1487580008128971748  |
| 1487580008145748965  |
| 1487580008187692007  |
| 1487580008204469224  |
| 1487580008221246441  |
| 1487580008263189483  |
| 1487580008288355308  |
| 1487580008313521133  |
| 1487580008422573041  |
| 1487580008565179383  |
| 1487580008607122425  |
| 1487580008699397117  |
| 1487580008800059394  |
| 1487580008816836611  |
| 1487580008858779653  |
| 1487580009009774604  |
| 1487580009026551821  |
| 1487580009076883471  |
| 1487580009143992338  |
| 1487580009177546772  |
| 1487580009202712597  |
| 1487580009227878422  |
```

14875800093872619 81
14875800094711480 64
14875800094963138 89
14875800095214797 14
14875800095466455 39
14875800095718113 64
14875800096053657 97
14875800096221430 14
14875800097647493 55
14875800098570240 46
14875800100247962 12
14875800100499620 37
14875800100751278 62
14875800101506253 37
14875800102177342 04
14875800102429000 29
14875800102932316 79
14875800103771177 63
14875800106036101 89
14875800106287760 14
14875800107210507 07
14875800107546051 41
14875800107965481 83
14875800110062633 91
14875800110985380 83
14875800111153153 00
14875800112243672 09
14875800112411444 26
14875800113837507 69
14875800115011912 86
14875800115599115 45
14875800116018545 87
14875800116270204 12
14875800116773520 62
14875800117528495 37
14875800119457875 26
14875800119709533 51
14875800119961191 76
14875800120967824 76
14875800121219483 01
14875800121471141 26
14875800122310002 09
14875800123736065 46
14875800124574926 29
14875800124826584 54
14875800125246014 96
14875800126168761 88
14875800126588192 30

```
| 1487580012717539489  |
| 1487580012784648356  |
| 1487580012851757223  |
| 1487580012902088873  |
| 1487580012969197740  |
| 1487580012994363565  |
| 1487580013011140782  |
| 1487580013027917999  |
| 1487580013053083824  |
| 1487580013069861041  |
| 1487580013128581300  |
| 1487580013229244601  |
| 1487580013254410426  |
| 1487580013279576251  |
| 1487580013363462335  |
| 1487580013388628160  |
| 1487580013413793985  |
| 1487580013472514244  |
| 1487580013489291461  |
| 1487580013506068678  |
| 1487580013522845895  |
| 1487580013564788937  |
| 1487580013581566154  |
| 1487580013640286413  |
| 1487580013732561106  |
| 1487580013749338323  |
| 1487580013799669974  |
| 1487580013917110492  |
| 1487580013933887709  |
+----------------------+
|     category_id      |
+----------------------+
| 1487580013950664926  |
| 1487580013992607968  |
| 1487580014093271270  |
| 1495705810662064688  |
| 1495705810704007729  |
| 1495705810729173554  |
| 1516331853567492962  |
| 1526733091857498510  |
| 1532652067498229876  |
| 1542195323827388674  |
| 1547480590851244887  |
| 1554383493545328915  |
| 1558526315760452545  |
| 1559261858748170632  |
| 1584505206060614455  |
| 1597769965795017114  |
```

16029436818730523 86
16044270947569504 59
16044271451389302 10
16280996264332499 31
16384561190661005 10
16488158438967220 76
16577220393870298 86
16584621252841312 65
17151027737473843 34
17527426066992341 59
17527426176966985 37
17590246989826629 57
17839990633146615 46
17839990641702996 32
17839990671818102 04
17839990688679206 26
17839990711999529 17
17839990713257820 53
17839990723324151 42
17839990723659695 78
17839990737584786 50
17839990765518853 99
17914428493848911 69
17914428959919973 90
18016352235914532 92
18043835825761813 09
18059539656789200 77
18196939590818862 39
18427357588053038 37
18427357604998027 45
18710189599275097 50
18914343518501603 81
18971244690176737 88
18971244892593856 29
19217234917201023 87
19240492875540072 23
19333643447209828 71
19334722867534240 63
19363270505497895 32
19371690730077562 69
19443582582233504 94
19582785512076746 74
19625251189282578 18
19625251265031708 58
19625254622991485 72
19775757754732388 31
19828602443797630 42
19828602635728981 12

```
| 1998040849203594085  |
| 1998040850688377703  |
| 1998040852064109417  |
| 2007399943458784057  |
| 2013754353822728372  |
| 2018287324474901238  |
| 2018395024110125980  |
| 2020400296164851991  |
| 2027602240612598357  |
| 2028340285225828822  |
| 2029082628195353599  |
| 2035665444290953519  |
| 2055161088059638328  |
| 2055368408169447599  |
| 2068966806634103136  |
| 2069171133327868014  |
| 2069804417665728971  |
| 2095736144888071137  |
| 2121383893343929118  |
| 2130081478220972046  |
| 2140803113261466607  |
| 2141560642253881670  |
| 2151191059751764547  |
| 2151191071051219817  |
| 2151191071118328683  |
| 2154396123597373922  |
| 2155132423103316327  |
| 2164688961165852944  |
| 2177933350667289121  |
| 2187686850687140020  |
| 2187790129827939246  |
| 2193074740493550411  |
| 2193074740619379535  |
| 2195085255117897760  |
| 2195085255176618020  |
| 2195085258339123402  |
+----------------------+
500 rows selected (25.925 seconds)
```

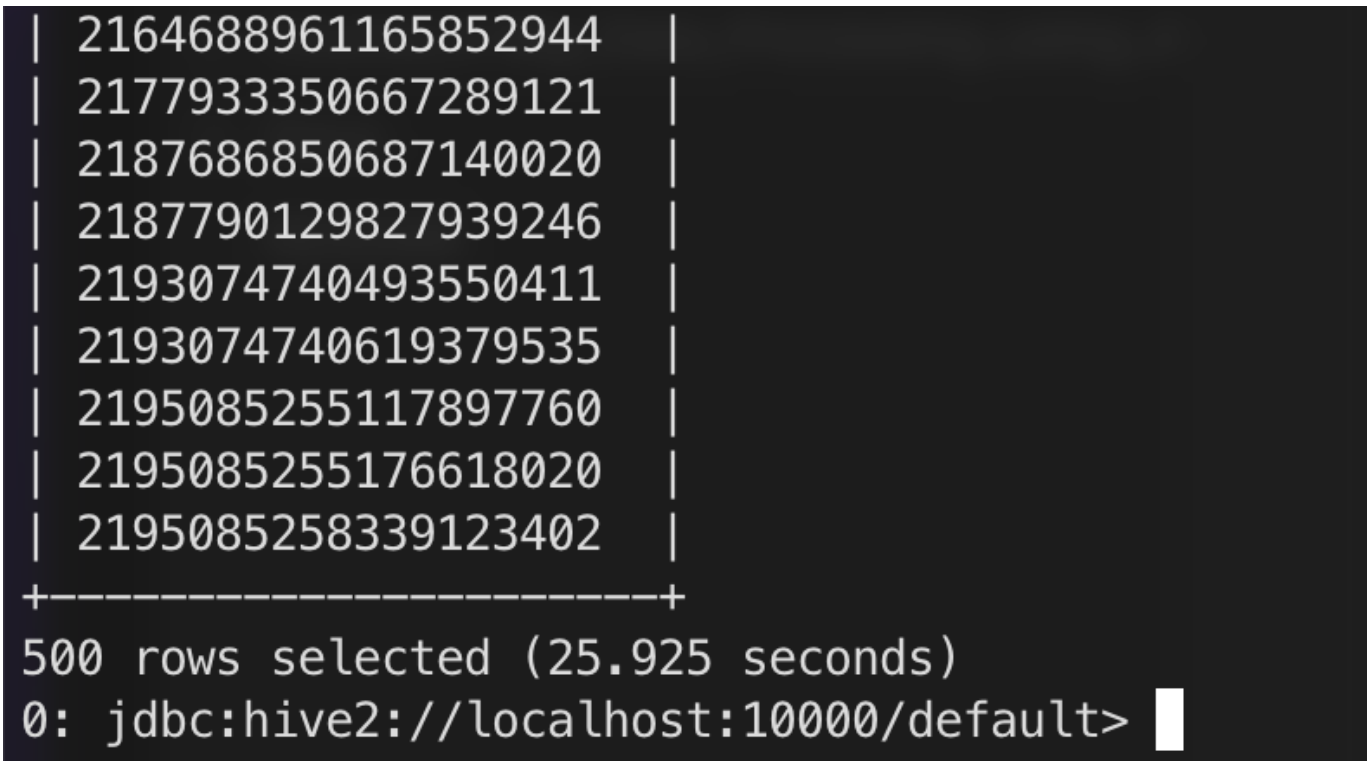
Screenshot:

```
0: jdbc:hive2://localhost:10000/default> select distinct(category_id) from clickstream_parq;
+----------------------+
|     category_id      |
+----------------------+
| 1487580004882580302  |
| 1487580004916134735  |
| 1487580005025186644  |
| 1487580005134238553  |
| 1487580005176181595  |
| 1487580005293622112  |
| 1487580005318787937  |
| 1487580005343953762  |
| 1487580005461394279  |
| 1487580005486560104  |
| 1487580005570446188  |
| 1487580005595612013  |
| 1487580005629166447  |
| 1487580005687886706  |
| 1487580005754995573  |
| 1487580005855658874  |
| 1487580005880824699  |
| 1487580005998265217  |
| 1487580006015042434  |
| 1487580006073762693  |
| 1487580006098928518  |
| 1487580006174425994  |
| 1487580006199591819  |
| 1487580006216369036  |
| 1487580006300255120  |
| 1487580006317032337  |
| 1487580006509970331  |
| 1487580006526747548  |
| 1487580006551913373  |
| 1487580006585467807  |
| 1487580006644188066  |
| 1487580006711296933  |
| 1487580006820348840  |
| 1487580006845514665  |
| 1487580006870680490  |
| 1487580006895846315  |
| 1487580007004898224  |
| 1487580007139115958  |
| 1487580007172670392  |
| 1487580007189447609  |
| 1487580007214613434  |
```

Screenshot:

```
| 1936327050549789532  |
| 1937169073007756269  |
| 1944358258223350494  |
| 1958278551207674674  |
| 1962525118928257818  |
| 1962525126503170858  |
```

1962525462299148572	
1977575775473238831	
1982860244379763042	
1982860263572898112	
1998040849203594085	
1998040850688377703	
1998040852064109417	
2007399943458784057	
2013754353822728372	
2018287324474901238	
2018395024110125980	
2020400296164851991	
2027602240612598357	
2028340285225828822	
2029082628195353599	
2035665444290953519	
2055161088059638328	
2055368408169447599	
2068966806634103136	
2069171133327868014	
2069804417665728971	
2095736144888071137	
2121383893343929118	
2130081478220972046	
2140803113261466607	
2141560642253881670	
2151191059751764547	
2151191071051219817	
2151191071118328683	
2154396123597373922	
2155132423103316327	



Question 5

Qn: Find the total number of products available under each category.

Ans:

Query:

```
select category_id, count(distinct(product_id)) product_count from
clickstream_parq group by category_id;
```

Output:

category_id	product_count
1487580004857414477	543
1487580004916134735	585
1487580004966466385	9
1487580004983243602	43
1487580005008409427	460
1487580005025186644	1
1487580005050352469	505
1487580005092295511	653
1487580005134238553	471
1487580005268456287	402
1487580005293622112	94
1487580005343953762	121
1487580005486560104	26
1487580005528503146	230
1487580005553668971	622
1487580005595612013	1229
1487580005654332272	1
1487580005687886706	9
1487580005754995573	438
1487580005796938615	1
1487580005855658874	19
1487580005880824699	7
1487580005939544958	7
1487580006015042434	2
1487580006073762693	14
1487580006174425994	8
1487580006317032337	877
1487580006484804506	46
1487580006509970331	34
1487580006551913373	59
1487580006627410849	5
1487580006644188066	72
1487580006711296933	5
1487580006744851367	3
1487580006820348840	10
1487580006895846315	95
1487580007021675441	28
1487580007072007091	29
1487580007097172916	8
1487580007113950133	6
1487580007139115958	5
1487580007189447609	17
1487580007214613434	15
1487580007256556476	56

14875800072817223O1	82
14875800073068881 26	60
14875800073991628 17	27
14875800074830489 00	85
14875800075082147 25	10
14875800075921008 09	23
14875800076340438 51	185
14875800076592096 76	99
14875800077011527 18	28
14875800077179299 35	472
14875800078353704 53	33
14875800078521476 70	197
14875800078940907 12	42
14875800079360337 54	87
14875800079528109 71	124
14875800079863654 05	12
14875800081121945 31	132
14875800081289717 48	1
14875800081457489 65	382
14875800081625261 82	24
14875800082044692 24	2
14875800082464122 66	967
14875800084477388 66	185
14875800085232363 41	9
14875800085651793 83	1
14875800086071224 25	239
14875800086574540 75	189
14875800089091113 03	13
14875800089258885 20	10
14875800090265518 21	68
14875800090768834 71	17
14875800091439923 38	180
14875800091775467 72	15
14875800092278784 22	16
14875800093620961 56	58
14875800094459822 39	257
14875800095214797 14	66
14875800095718113 64	14
14875800096053657 97	97
14875800096473088 39	1
14875800096724746 64	4
14875800097647493 55	2
14875800098234696 13	9
14875800098570240 46	1
14875800100499620 37	3
14875800100751278 62	1
14875800101002936 87	469
14875800101254595 12	92

1487580010150625337	1
1487580010192568379	12
1487580010217734204	40
1487580010242900029	47
1487580010268065854	65
1487580010293231679	46
1487580010318397504	87
1487580010377117763	19

category_id	product_count
1487580010427449413	1
1487580010603610189	5
1487580010645553231	61
1487580010662330448	11
1487580010695884882	42
1487580010779770966	50
1487580010846879833	14
1487580010955931741	6
1487580010989486174	4
1487580011115315300	6
1487580011157258342	16
1487580011182424167	16
1487580011199201384	4
1487580011224367209	1
1487580011308253293	85
1487580011333419118	2
1487580011425693811	76
1487580011476025461	1
1487580011501191286	77
1487580011534745720	136
1487580011702517887	350
1487580011853512836	113
1487580011903844485	18
1487580011945787526	5
1487580011996119176	48
1487580012373606546	182
1487580012432326804	66
1487580012457492629	16
1487580012482658454	8
1487580012507824279	3
1487580012549767321	1
1487580012574933146	183
1487580012642042013	83
1487580012658819230	2
1487580012759482531	4
1487580012809814181	4
1487580012851757223	5

1487580012927254698	117
1487580012952420523	22
1487580012969197740	78
1487580012994363565	114
1487580013027917999	28
1487580013053083824	138
1487580013069861041	126
1487580013145358517	101
1487580013212467384	53
1487580013321519293	38
1487580013338296510	42
1487580013363462335	1
1487580013413793985	56
1487580013472514244	5
1487580013564788937	14
1487580013581566154	127
1487580013732561106	45
1487580013782892757	1
1487580013824835799	21
1487580013858390233	125
1487580013950664926	226
1495705810662064688	24
1495705810704007729	49
1495705810754339379	1
1511892746070131099	236
1516331853567492962	36
1525995662934540829	273
1526733091857498510	198
1532652067498229876	6
1542195323827388674	163
1543705961326182546	6
1547480590851244887	4
1559261858748170632	11
1588205336064425989	13
1602943681873052386	233
1658462125284131265	707
1660394382047576308	3
1720400165430363096	7
1752742606699234159	109
1752742615205281895	13
1761186209054327497	53
1783999064136745198	53
1783999067181810204	8
1783999068867920626	40
1783999068909863670	167
1783999071199952917	120
1783999072332415142	388
1783999072407912622	24

1783999073758478650	102
1783999073792033086	81
1783999076535108182	58
1783999076551885399	43
1791442849384891169	7
1791442895991997390	49
1797122099068797194	16
1804383582576181309	6
1810470908326838736	3
1842735764450837316	1
1842735764509557576	20
1842735768846467167	17
1891434214553813877	92
1897124469017673788	120
1897124489259385629	16

category_id	product_count
1911999801088541491	100
1911999948073730325	20
1913479463425802701	6
1924049287554007223	21
1926797403503985079	7
1933364344720982871	98
1936327050549789532	3
1937169073007756269	170
1958278551207674674	38
1962525118928257818	10
1962525126503170858	10
1977575787259232875	49
1982860263572898112	57
1987724780047958178	8
1998040849203594085	16
1998040852064109417	6
2007399943458784057	76
2013754353822728372	74
2022622168218599898	2
2029082628195353599	11
2035665444290953519	117
2055161088059638328	27
2060156961931919712	3
2068966806634103136	14
2069171133327868014	1
2069804417665728971	196
2084144451428549153	230
2089259162625114209	89
2093602042093240877	24
2094448780651791052	48

2095736144888071137	77
2106514244437541443	25
2106514244487873093	30
2114584564549550293	86
2115334439910245200	73
2121383893343929118	34
2130081478220972046	6
2134354342373753638	148
2145935122136826354	7
2151191059751764547	29
2151191059827262021	5
2151191070908613477	20
2151191071051219817	233
2151191071118328683	36
2151191075757228942	4
2155132423103316327	3
2177933350667289121	142
2187790129827939246	15
2193074740686488401	4
2195085255034011676	66
2195085255117897760	13
2195085258272014535	6
2195085258339123402	1
1487580004832248652	283
1487580004882580302	148
1487580005067129686	88
1487580005176181595	18
1487580005318787937	39
1487580005369119587	3
1487580005385896804	211
1487580005411062629	329
1487580005427839846	512
1487580005461394279	781
1487580005511725929	615
1487580005570446188	7
1487580005629166447	71
1487580005671109489	963
1487580005713052531	302
1487580005897601916	42
1487580005922767741	2
1487580005998265217	3
1487580006056985476	3
1487580006098928518	10
1487580006132482952	76
1487580006157648777	24
1487580006199591819	17
1487580006216369036	2
1487580006241534861	36

1487580006300255120	102
1487580006350586771	85
1487580006409307030	41
1487580006434472855	18
1487580006451250072	16
1487580006526747548	45
1487580006585467807	6
1487580006610633632	4
1487580006845514665	14
1487580006870680490	6
1487580006937789357	62
1487580006979732399	8
1487580007004898224	18
1487580007046841266	14
1487580007172670392	9
1487580007365608384	79
1487580007432717250	188
1487580007457883075	65
1487580007524991942	12
1487580007550157767	101
1487580007575323592	391
1487580007675986893	2322

category_id	product_count
1487580007759872977	28
1487580007776650194	32
1487580007877313495	24
1487580007910867929	117
1487580008011531230	35
1487580008028308447	10
1487580008053474272	9
1487580008070251489	9
1487580008087028706	19
1487580008187692007	112
1487580008221246441	18
1487580008263189483	455
1487580008288355308	211
1487580008313521133	202
1487580008422573041	25
1487580008472904691	23
1487580008590345208	7
1487580008674231292	51
1487580008699397117	78
1487580008758116352	66
1487580008774893569	88
1487580008800059394	329
1487580008816836611	27

14875800088587796 53	22	
14875800089510543 45	16	
14875800089846087 79	48	
14875800090097746 04	1	
14875800090517176 46	186	
14875800091188265 13	15	
14875800092027125 97	52	
14875800092614328 56	43	
14875800092865986 81	430	
14875800093117645 06	132	
14875800093369303 31	34	
14875800093872619 81	80	
14875800094711480 64	142	
14875800094963138 89	41	
14875800095466455 39	8	
14875800096221430 14	20	
14875800097395835 30	10	
14875800097983037 88	9	
14875800098905784 79	1	
14875800099744645 62	2	
14875800100247962 12	1	
14875800103351747 21	21	
14875800103603405 46	2	
14875800105616671 47	111	
14875800106287760 14	75	
14875800107210507 07	12	
14875800107378279 24	7	
14875800107546051 41	141	
14875800107965481 83	97	
14875800108217140 08	81	
14875800108720456 58	142	
14875800108972114 83	1	
14875800109223773 08	18	
14875800110062633 91	3	
14875800110314292 16	9	
14875800110985380 83	16	
14875800111404811 25	1	
14875800112411444 26	28	
14875800112830874 68	113	
14875800113837507 69	1	
14875800114089165 94	87	
14875800115179685 03	15	
14875800115599115 45	3	
14875800115850773 70	866	
14875800116018545 87	20	
14875800116270204 12	255	
14875800116521862 37	360	
14875800116773520 62	113	

1487580011752849537	4
1487580011786403970	17
1487580011828347011	1
1487580011970953351	27
1487580012021285001	60
1487580012071616651	2
1487580012096782476	357
1487580012121948301	97
1487580012147114126	16
1487580012172279951	23
1487580012205834384	34
1487580012231000209	1
1487580012524601496	139
1487580012591710363	55
1487580012616876188	17
1487580012683985055	83
1487580012717539489	39
1487580012734316706	11
1487580012784648356	14
1487580012876923048	3
1487580012902088873	55
1487580013011140782	65
1487580013086638258	33
1487580013128581300	68
1487580013170524342	114
1487580013229244601	146
1487580013254410426	116
1487580013279576251	297
1487580013296353468	67

category_id	product_count
1487580013388628160	125
1487580013430571202	59
1487580013489291461	64
1487580013506068678	68
1487580013522845895	215
1487580013539623112	51
1487580013615120588	18
1487580013640286413	134
1487580013657063630	88
1487580013690618064	49
1487580013749338323	74
1487580013799669974	22
1487580013841613016	1820
1487580013900333275	9
1487580013917110492	93
1487580013933887709	4

1487580013992607968	18
1487580014009385185	29
1487580014042939619	6
1487580014093271270	15
1495705810729173554	72
1554383493545328915	2
1558526315760452545	60
1584505206060614455	16
1597769965795017114	114
1597770225539875791	160
1604427094756950459	56
1604427145138930210	3
1605161575889502297	22
1628099626433249931	9
1638456119066100510	407
1645114480121610699	42
1648815651034235876	51
1648815843896722076	3
1657722039387029886	10
1715102762414375164	1
1715102773747384334	9
1725504706412807026	8
1752742610406999036	12
1752742617696698537	5
1759024698982662957	1
1759024699007828783	41
1783999063314661546	156
1783999063574708423	10
1783999064036081896	4
1783999064103190764	41
1783999064170299632	3
1783999067156644376	477
1783999071325782053	64
1783999072365969578	63
1791442703716712702	2
1801635223591453292	9
1805953852441101266	7
1805953965678920077	1
1814592470537732872	34
1819693959081886239	108
1842735758805303837	118
1842735760499802745	116
1858245586344477369	4
1871018959927509750	232
1889472915104072007	25
1891434351850160381	34
1897124478404526487	91
1897124495995438074	6

```
| 1911999801642189621  | 44             |
| 1911999884991397970  | 6              |
| 1921723491720102387  | 84             |
| 1921723506584715388  | 16             |
| 1924049106385240809  | 175            |
| 1924049110428549877  | 323            |
| 1933472286753424063  | 176            |
| 1944358258223350494  | 53             |
| 1962525462299148572  | 20             |
| 1977575775473238831  | 2              |
| 1977786601392047073  | 83             |
| 1982860244379763042  | 53             |
| 1998040850688377703  | 4              |
| 2018287324474901238  | 4              |
| 2018395024110125980  | 75             |
| 2020400296164851991  | 2              |
| 2022516588854378951  | 4              |
| 2027602240612598357  | 1              |
| 2028340285225828822  | 6              |
| 2029731308699124089  | 57             |
| 2055368408169447599  | 26             |
| 2069804424703771380  | 1              |
| 2071303198680810125  | 15             |
| 2134354356349173879  | 2              |
| 2140803113261466607  | 52             |
| 2141560642253881670  | 17             |
| 2151191070984110951  | 30             |
| 2151191071378375538  | 161            |
| 2154396123597373922  | 10             |
| 2164688961165852944  | 48             |
| 2166295400451933025  | 1              |
| 2187686850687140020  | 2              |
| 2193074740493550411  | 4              |
| 2193074740552270669  | 23             |
| 2193074740619379535  | 6              |
| 2195085255176618020  | 56             |
+----------------------+----------------+
500 rows selected (31.702 seconds)
```

Screenshot:

```
0: jdbc:hive2://localhost:10000/default> select category_id, count(distinct(product_id)) product_count from clickstream_parq group by category_id;
+----------------------+----------------+
|     category_id      | product_count  |
+----------------------+----------------+
| 1487580004857414477  | 543            |
| 1487580004916134735  | 585            |
| 1487580004966466385  | 9              |
| 1487580004983243602  | 43             |
| 1487580005008409427  | 460            |
| 1487580005025186644  | 1              |
| 1487580005050352469  | 505            |
| 1487580005092295511  | 653            |
| 1487580005134238553  | 471            |
| 1487580005268456287  | 402            |
| 1487580005293622112  | 94             |
| 1487580005343953762  | 121            |
| 1487580005486560104  | 26             |
| 1487580005528503146  | 230            |
| 1487580005553668971  | 622            |
| 1487580005595612013  | 1229           |
| 1487580005654332272  | 1              |
| 1487580005687886706  | 9              |
| 1487580005754995573  | 438            |
| 1487580005796938615  | 1              |
| 1487580005855658874  | 19             |
| 1487580005880824699  | 7              |
| 1487580005939544958  | 7              |
| 1487580006015042434  | 2              |
| 1487580006073762693  | 14             |
| 1487580006174425994  | 8              |
| 1487580006317032337  | 877            |
| 1487580006484804506  | 46             |
| 1487580006509970331  | 34             |
| 1487580006551913373  | 59             |
| 1487580006627410849  | 5              |
| 1487580006644188066  | 72             |
| 1487580006711296933  | 5              |
| 1487580006744851367  | 3              |
| 1487580006820348840  | 10             |
| 1487580006895846315  | 95             |
| 1487580007021675441  | 28             |
| 1487580007072007091  | 29             |
| 1487580007097172916  | 8              |
| 1487580007113950133  | 6              |
| 1487580007139115958  | 5              |
| 1487580007189447609  | 17             |
| 1487580007214613434  | 15             |
| 1487580007256556476  | 56             |
| 1487580007281722301  | 82             |
```

Screenshot:

```
| 1814592470537732872  | 34             |
| 1819693959081886239  | 108            |
| 1842735758805303837  | 118            |
| 1842735760499802745  | 116            |
| 1858245586344477369  | 4              |
| 1871018959927509750  | 232            |
| 1889472915104072007  | 25             |
| 1891434351850160381  | 34             |
| 1897124478404526487  | 91             |
| 1897124495995438074  | 6              |
| 1911999801642189621  | 44             |
| 1911999884991397970  | 6              |
| 1921723491720102387  | 84             |
| 1921723506584715388  | 16             |
| 1924049106385240809  | 175            |
| 1924049110428549877  | 323            |
| 1933472286753424063  | 176            |
| 1944358258223350494  | 53             |
| 1962525462299148572  | 20             |
| 1977575775473238831  | 2              |
| 1977786601392047073  | 83             |
| 1982860244379763042  | 53             |
| 1998040850688377703  | 4              |
| 2018287324474901238  | 4              |
| 2018395024110125980  | 75             |
| 2020400296164851991  | 2              |
| 2022516588854378951  | 4              |
| 2027602240612598357  | 1              |
| 2028340285225828822  | 6              |
| 2029731308699124089  | 57             |
| 2055368408169447599  | 26             |
| 2069804424703771380  | 1              |
| 2071303198680810125  | 15             |
| 2134354356349173879  | 2              |
| 2140803113261466607  | 52             |
| 2141560642253881670  | 17             |
| 2151191070984110951  | 30             |
| 2151191071378375538  | 161            |
| 2154396123597373922  | 10             |
| 2164688961165852944  | 48             |
| 2166295400451933025  | 1              |
| 2187686850687140020  | 2              |
| 2193074740493550411  | 4              |
| 2193074740552270669  | 23             |
| 2193074740619379535  | 6              |
| 2195085255176618020  | 56             |
+----------------------+----------------+
500 rows selected (31.702 seconds)
0: jdbc:hive2://localhost:10000/default>
```

Question 6

Qn: Which brand had the maximum sales in October and November combined?

Ans:

Query:

```
select brand, sum(price) revenue from clickstream_parq where
event_type='purchase' group by brand order by revenue desc limit 2;
```

Output:

```
+---------+---------------------+
|  brand  |       revenue       |
+---------+---------------------+
|         | 1094188.2993474863  |
| runail  | 148297.93996394053  |
+---------+---------------------+
2 rows selected (9.624 seconds)
```

Screenshot:

PS: The brand that made most revenue is "UNKNOWN" (might mean a collection of smaller lesser known brands) and the brand that made the most revenue is "RUNAIL"

Question 7

Qn: Which brands increased their sales from October to November?

Ans:

Query:

```
with revenue_month
as
(select brand, month, sum(price) revenue from clickstream_parq where
event_type='purchase' and brand <> '' group by month, brand order by month)
select nov_revenue.brand as brand,  (nov_revenue.revenue - oct_revenue.revenue)
as revenue_difference
from revenue_month oct_revenue
inner join revenue_month nov_revenue
on oct_revenue.month + 1 = nov_revenue.month
and oct_revenue.brand = nov_revenue.brand
where (nov_revenue.revenue - oct_revenue.revenue) > 0;
```

Output:

brand	revenue_difference
aura	93.56000328063965
batiste	101.76999497413635
benovy	2850.349985599518
bioaqua	455.2300034761429
blixz	24.450002551078796
bluesky	258.2899710536003
bodyton	4.3000054359436035
bpw.style	3265.290124952793
browxenna	585.3599421977997
candy	264.419993519783
chi	179.67000150680542
concept	2348.2599958777428
cosima	0.6999996304512024
cosmoprofi	6214.179965496063
cristalinas	157.32000541687012
cutrin	68.24999809265137
de.lux	1115.810008585453
deoproce	12.329999446868896
depilflax	96.70999884605408
dizao	126.38000643253326
ecolab	951.4500054121017
egomania	68.57000255584717
elizavecca	133.76999950408936
elskin	56.55999952554703
eos	98.26999711990356
estelare	27.060003489255905
finish	132.00000047683716
freedecor	4250.020043194294
gehwol	468.6100044250488
godefroy	23.899999141693115
happyfons	289.66999769210815
haruyama	2962.2199823260307
insight	278.26000452041626
irisk	1354.0799748450518
italwax	2859.130049407482
jaguar	8.539989948272705
jas	338.4699912816286
kapous	2165.91998565197
keen	199.27000188827515
kerasys	94.29000055789948
kims	302.0000066757202
kiss	395.77999687194824
kocostar	284.08000469207764
koelcia	57.25

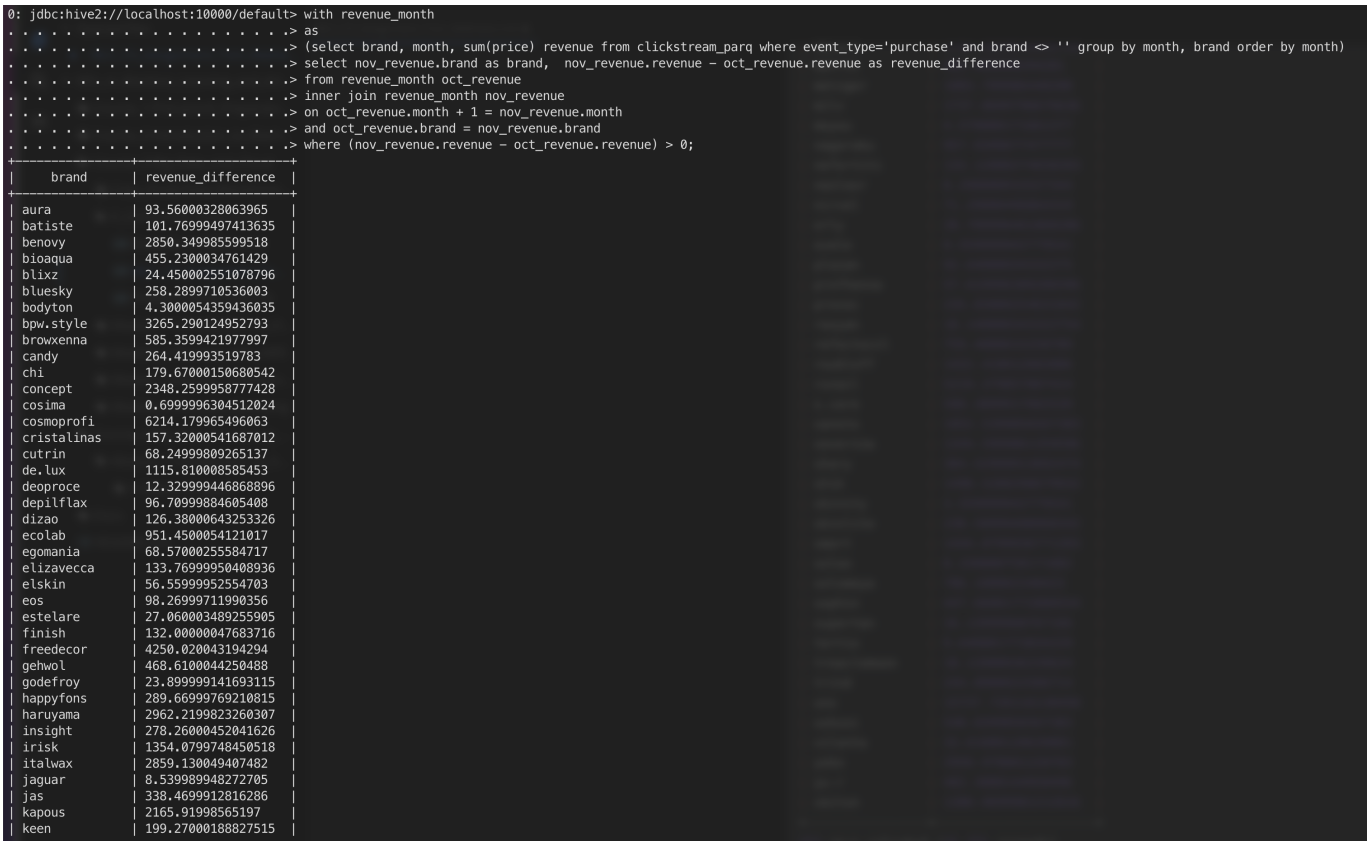
kosmekka	631.9300060272217
laboratorium	66.01999807357788
ladykin	44.920000076293945
latinoil	135.07000207901
limoni	487.69998824596405
mane	193.4699993133545
marutaka-foot	60.11000061035156
mavala	37.28000020980835
miskin	135.02999925613403
missha	856.4500023126602
nitrile	315.3999910354614
oniq	1416.240023612976
osmo	116.72999906539917
polarus	5358.210015535355
profepil	24.659998178482056
protokeratin	255.5400047302246
rosi	764.520025730133
staleks	3355.8799889683723
strong	9474.640277385712
swarovski	1155.2299975156784
veraclara	21.09999930858612
airnails	572.6200138628483
art-visage	905.0899993181229
artex	1596.6100018024445
balbcare	57.04999780654907
beautix	1729.0000777244568
beauty-free	1228.6899927854538
beautyblender	30.670001983642578
beauugreen	256.8400115966797
biore	29.659997940063477
carmex	98.27999782562256
coifin	525.4900169372559
domix	1537.1199771165848
ecocraft	200.79000186920166
ellips	360.19000577926636
enjoy	95.22000217437744
entity	239.54999166727066
estel	2385.9200018048286
f.o.x	1953.0500071048737
farmavita	454.59999895095825
farmona	150.9699878692627
fedua	211.42999839782715
fly	10.030001163482666
foamie	45.44999980926514
freshbubble	183.63999772071838
glysolid	21.860000491142273
grace	1.6900005340576172
grattol	36027.169416069984

greymy	460.279993057251
igrobeauty	131.40999913215637
ingarden	10404.820005103946
inm	63.18999981880188
jessnail	7057.389890432358
joico	1309.5800108909607
kaaral	673.6400128602982
kamill	18.47999930381775

brand	revenue_difference
kaypro	2387.3599891662598
kinetics	611.0100679397583
koelf	84.55999374389648
konad	70.83999770879745
lador	387.9200019836426
levissime	857.8099956512451
levrana	1420.5399911999702
lianail	10501.399931311607
likato	44.90999746322632
lovely	3234.6799781620502
lowence	324.91000270843506
marathon	2992.3500514030457
markell	1065.6799969673157
masura	1792.3899676203728
matrix	483.489999294281
metzger	1083.709986448288
milv	1737.0699796676636
moyou	4.570000171661377
nagaraku	957.9399677477777
nefertiti	133.12000274658203
neoleor	8.290000915527344
nirvel	71.29000496864319
orly	28.709996461868286
ovale	0.559999942779541
plazan	92.64000034332275
profhenna	57.619996309280396
provoc	235.83000254631042
rasyan	10.140000343322754
refectocil	759.4000015258789
roubloff	1422.410012602806
runail	5219.379857007414
s.care	500.3899917602539
sanoto	1052.5399856567383
severina	1344.5999861359596
shary	304.52999913692474
shik	1498.5200290679932
skinity	3.559999942779541

```
| skinlite        | 238.50999408960342  |
| smart           | 1444.8799936771393  |
| soleo           | 8.330000758171082   |
| solomeya        | 786.100002348423    |
| sophin          | 447.66001772880554  |
| supertan        | 16.13999968767166   |
| tertio          | 9.640001773834229   |
| treaclemoon     | 18.12000036239624   |
| trind           | 244.8900022506714   |
| uno             | 15737.720116138458  |
| uskusi          | 548.0399856567383   |
| vilenta         | 33.61000120639801   |
| yoko            | 2950.970001220703   |
| yu-r            | 402.3000144958496   |
| zeitun          | 1300.9699981212616  |
+-----------------+---------------------+
152 rows selected (15.371 seconds)
```

Screenshot:



```
0: jdbc:hive2://localhost:10000/default> with revenue_month
. . . . . .> as
. . . . . .> (select brand, month, sum(price) revenue from clickstream_parq where event_type='purchase' and brand <> '' group by month, brand order by month)
. . . . . .> select nov_revenue.brand as brand,  nov_revenue.revenue - oct_revenue.revenue as revenue_difference
. . . . . .> from revenue_month oct_revenue
. . . . . .> inner join revenue_month nov_revenue
. . . . . .> on oct_revenue.month + 1 = nov_revenue.month
. . . . . .> and oct_revenue.brand = nov_revenue.brand
. . . . . .> where (nov_revenue.revenue - oct_revenue.revenue) > 0;
+---------------+---------------------+
|     brand     | revenue_difference  |
+---------------+---------------------+
| aura          | 93.56000328063965   |
| batiste       | 101.76999497413635  |
| benovy        | 2850.349985599518   |
| bioaqua       | 455.2300034761429   |
| blixz         | 24.450002551078796  |
| bluesky       | 258.2899710536003   |
| bodyton       | 4.3000054359436035  |
| bpw.style     | 3265.290124952793   |
| browxenna     | 585.3599421977997   |
| candy         | 264.419993519783    |
| chi           | 179.67000150680542  |
| concept       | 2348.2599958777428  |
| cosima        | 0.6999996304512024  |
| cosmoprofi    | 6214.179965496063   |
| cristalinas   | 157.32000541687012  |
| cutrin        | 68.24999809265137   |
| de.lux        | 1115.810008585453   |
| deoproce      | 12.329999446868896  |
| depilflax     | 96.70999884605408   |
| dizao         | 126.38000643253326  |
| ecolab        | 951.4500054121017   |
| egomania      | 68.57000255584717   |
| elizavecca    | 133.76999950408936  |
| elskin        | 56.55999952554703   |
| eos           | 98.26999711990356   |
| estelare      | 27.060003489255905  |
| finish        | 132.00000047683716  |
| freedecor     | 4250.020043194294   |
| gehwol        | 468.6100044250488   |
| godefroy      | 23.899999141693115  |
| happyfons     | 289.66999769210815  |
| haruyama      | 2962.2199823260307  |
| insight       | 278.26000452041626  |
| irisk         | 1354.0799748450518  |
| italwax       | 2859.130049407482   |
| jaguar        | 8.539989948272705   |
| jas           | 338.4699912816286   |
| kapous        | 2165.91998565197    |
| keen          | 199.27000188827515  |
```

Screenshot:

levissime	857.8099956512451
levrana	1420.5399911999702
lianail	10501.399931311607
likato	44.90999746322632
lovely	3234.6799781620502
lowence	324.91000270843506
marathon	2992.3500514030457
markell	1065.6799969673157
masura	1792.3899676203728
matrix	483.489999294281
metzger	1083.709986448288
milv	1737.0699796676636
moyou	4.570000171661377
nagaraku	957.9399677477777
nefertiti	133.12000274658203
neoleor	8.290000915527344
nirvel	71.29000496864319
orly	28.709996461868286
ovale	0.559999942779541
plazan	92.64000034332275
profhenna	57.619996309280396
provoc	235.83000254631042
rasyan	10.140000343322754
refectocil	759.4000015258789
roubloff	1422.410012602806
runail	5219.379857007414
s.care	500.3899917602539
sanoto	1052.5399856567383
severina	1344.5999861359596
shary	304.52999913692474
shik	1498.5200290679932
skinity	3.559999942779541
skinlite	238.50999408960342
smart	1444.8799936771393
soleo	8.330000758171082
solomeya	786.100002348423
sophin	447.66001772880554
supertan	16.13999968767166
tertio	9.640001773834229
treaclemoon	18.12000036239624
trind	244.8900022506714
uno	15737.720116138458
uskusi	548.0399856567383
vilenta	33.61000120639801
yoko	2950.970001220703
yu-r	402.3000144958496
zeitun	1300.9699981212616

152 rows selected (15.371 seconds)
0: jdbc:hive2://localhost:10000/default>

Question 8

Qn: Your company wants to reward the top 10 users of its website with a Golden Customer plan. Write a query to generate a list of top 10 users who spend the most on purchases.

Ans:

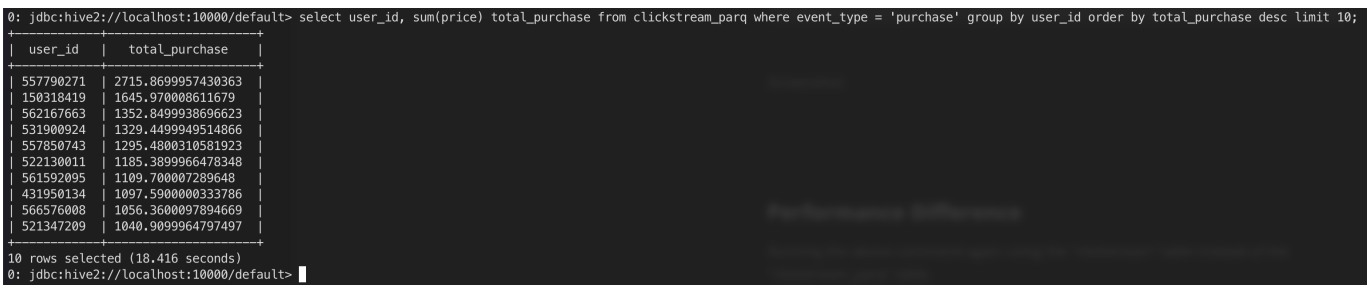
Query:

```
select user_id, sum(price) total_purchase from clickstream_parq where
event_type = 'purchase' group by user_id order by total_purchase desc limit 10;
```

Output:

```
+------------+---------------------+
|  user_id   |   total_purchase    |
+------------+---------------------+
| 557790271  | 2715.8699957430363  |
| 150318419  | 1645.970008611679   |
| 562167663  | 1352.8499938696623  |
| 531900924  | 1329.4499949514866  |
| 557850743  | 1295.4800310581923  |
| 522130011  | 1185.3899966478348  |
| 561592095  | 1109.700007289648   |
| 431950134  | 1097.5900000333786  |
| 566576008  | 1056.3600097894669  |
| 521347209  | 1040.9099964797497  |
+------------+---------------------+
10 rows selected (18.416 seconds)
```

Screenshot:



Performance Difference

Running the above command again using the "clickstream" table instead of the "clickstream_parq" table.

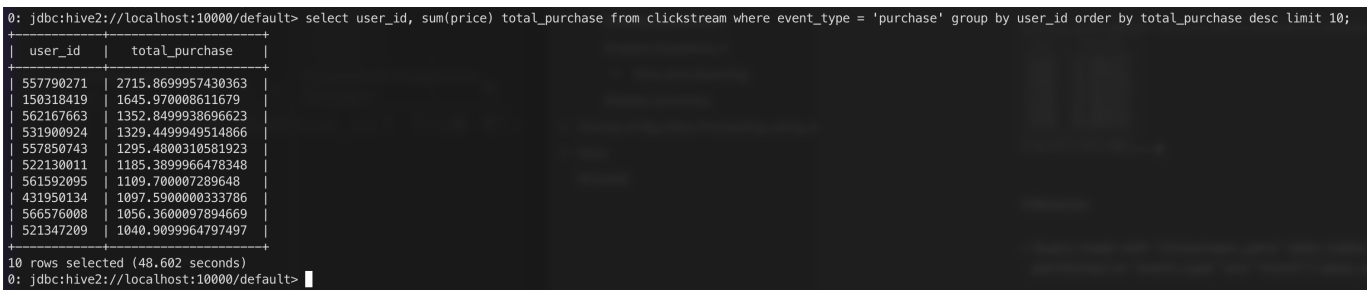
Query:

```
select user_id, sum(price) total_purchase from clickstream where event_type =
'purchase' group by user_id order by total_purchase desc limit 10;
```

Output:

```
|  user_id   |   total_purchase    |
+------------+---------------------+
| 557790271  | 2715.8699957430363  |
| 150318419  | 1645.970008611679   |
| 562167663  | 1352.8499938696623  |
| 531900924  | 1329.4499949514866  |
| 557850743  | 1295.4800310581923  |
| 522130011  | 1185.3899966478348  |
| 561592095  | 1109.700007289648   |
| 431950134  | 1097.5900000333786  |
| 566576008  | 1056.3600097894669  |
| 521347209  | 1040.9099964797497  |
+------------+---------------------+
10 rows selected (48.602 seconds)
```

Screenshot:



- Query made with "clickstream_parq" table (table stored in PARQUET format and partitioned on "event_type" and "month") takes about **18.416 seconds**.
- Query made with "clickstream" table (table just with raw data) takes about **48.602 seconds**.
- We can clearly infer that query made on the performance optimised table is significantly lower than the regular table.