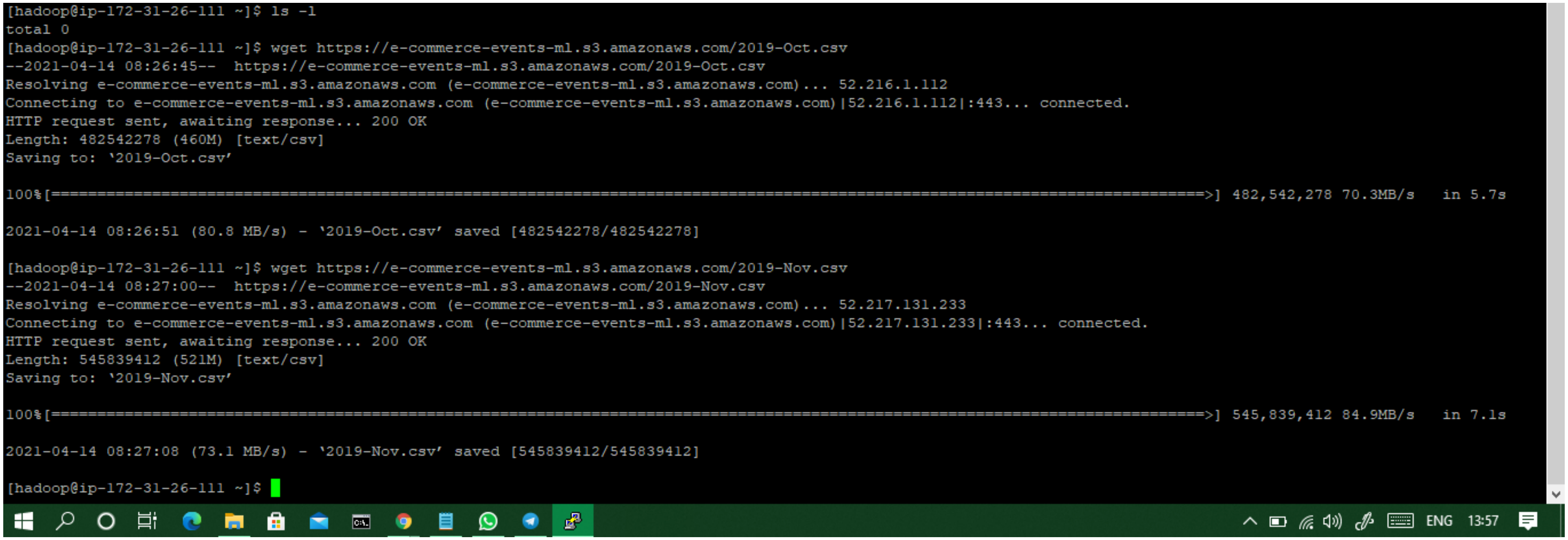
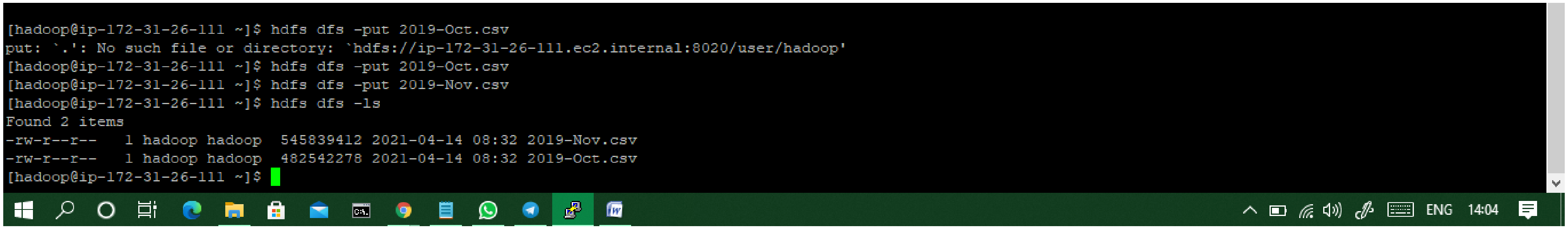
Importing the data to HDFS



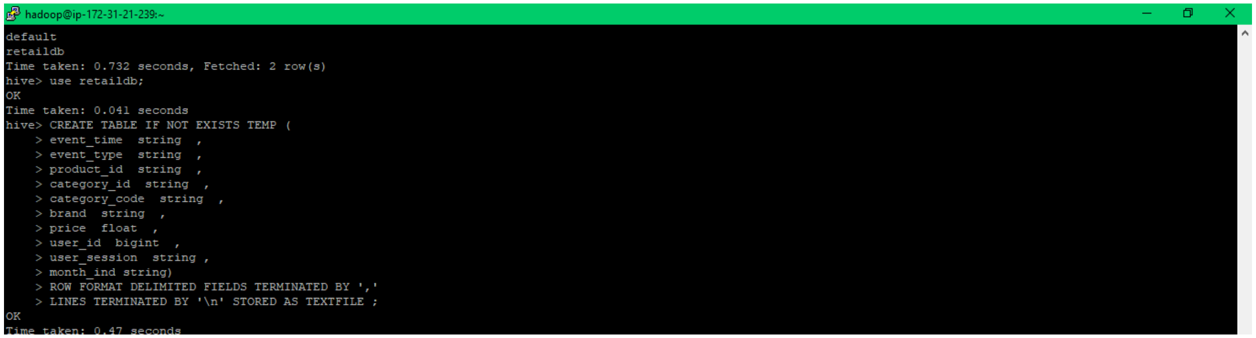
```
[hadoop@ip-172-31-26-111 ~]$ ls -l
total 0
[hadoop@ip-172-31-26-111 ~]$ wget https://e-commerce-events-ml.s3.amazonaws.com/2019-Oct.csv
--2021-04-14 08:26:45--  https://e-commerce-events-ml.s3.amazonaws.com/2019-Oct.csv
Resolving e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-ml.s3.amazonaws.com)... 52.216.1.112
Connecting to e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-ml.s3.amazonaws.com)|52.216.1.112|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 482542278 (460M) [text/csv]
Saving to: ‘2019-Oct.csv’

100%[===================================================================================================>] 482,542,278 70.3MB/s   in 5.7s

2021-04-14 08:26:51 (80.8 MB/s) - ‘2019-Oct.csv’ saved [482542278/482542278]

[hadoop@ip-172-31-26-111 ~]$ wget https://e-commerce-events-ml.s3.amazonaws.com/2019-Nov.csv
--2021-04-14 08:27:00--  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)... 52.217.131.233
Connecting to e-commerce-events-ml.s3.amazonaws.com (e-commerce-events-ml.s3.amazonaws.com)|52.217.131.233|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 545839412 (521M) [text/csv]
Saving to: ‘2019-Nov.csv’

100%[===================================================================================================>] 545,839,412 84.9MB/s   in 7.1s

2021-04-14 08:27:08 (73.1 MB/s) - ‘2019-Nov.csv’ saved [545839412/545839412]

[hadoop@ip-172-31-26-111 ~]$
```



```
[hadoop@ip-172-31-26-111 ~]$ hdfs dfs -put 2019-Oct.csv
put: `.': No such file or directory: `hdfs://ip-172-31-26-111.ec2.internal:8020/user/hadoop'
[hadoop@ip-172-31-26-111 ~]$ hdfs dfs -put 2019-Oct.csv
[hadoop@ip-172-31-26-111 ~]$ hdfs dfs -put 2019-Nov.csv
[hadoop@ip-172-31-26-111 ~]$ hdfs dfs -ls
Found 2 items
-rw-r--r--   1 hadoop hadoop  545839412 2021-04-14 08:32 2019-Nov.csv
-rw-r--r--   1 hadoop hadoop  482542278 2021-04-14 08:32 2019-Oct.csv
[hadoop@ip-172-31-26-111 ~]$
```

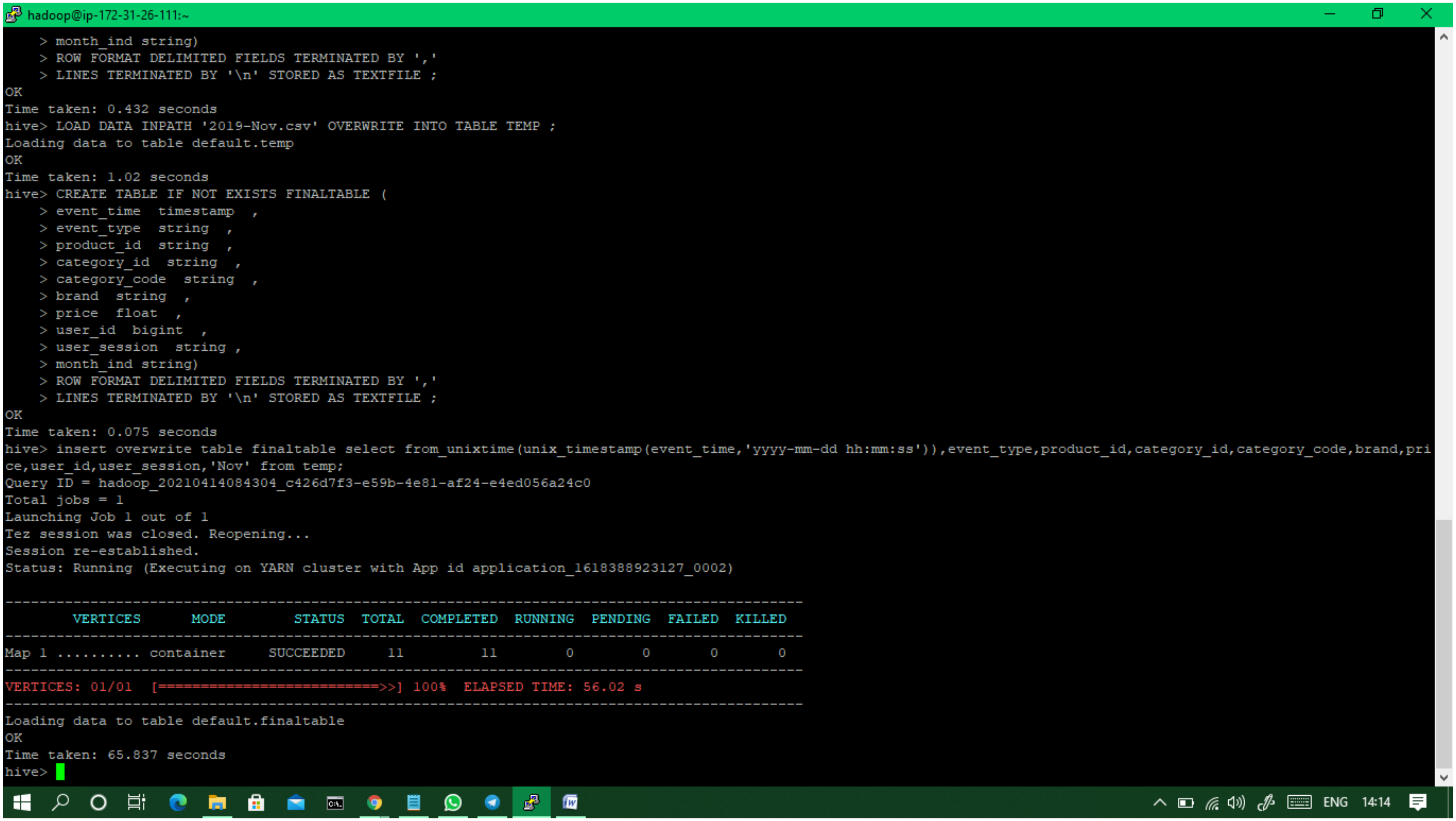
Creating Hive Tables

Create TEMP table to load data



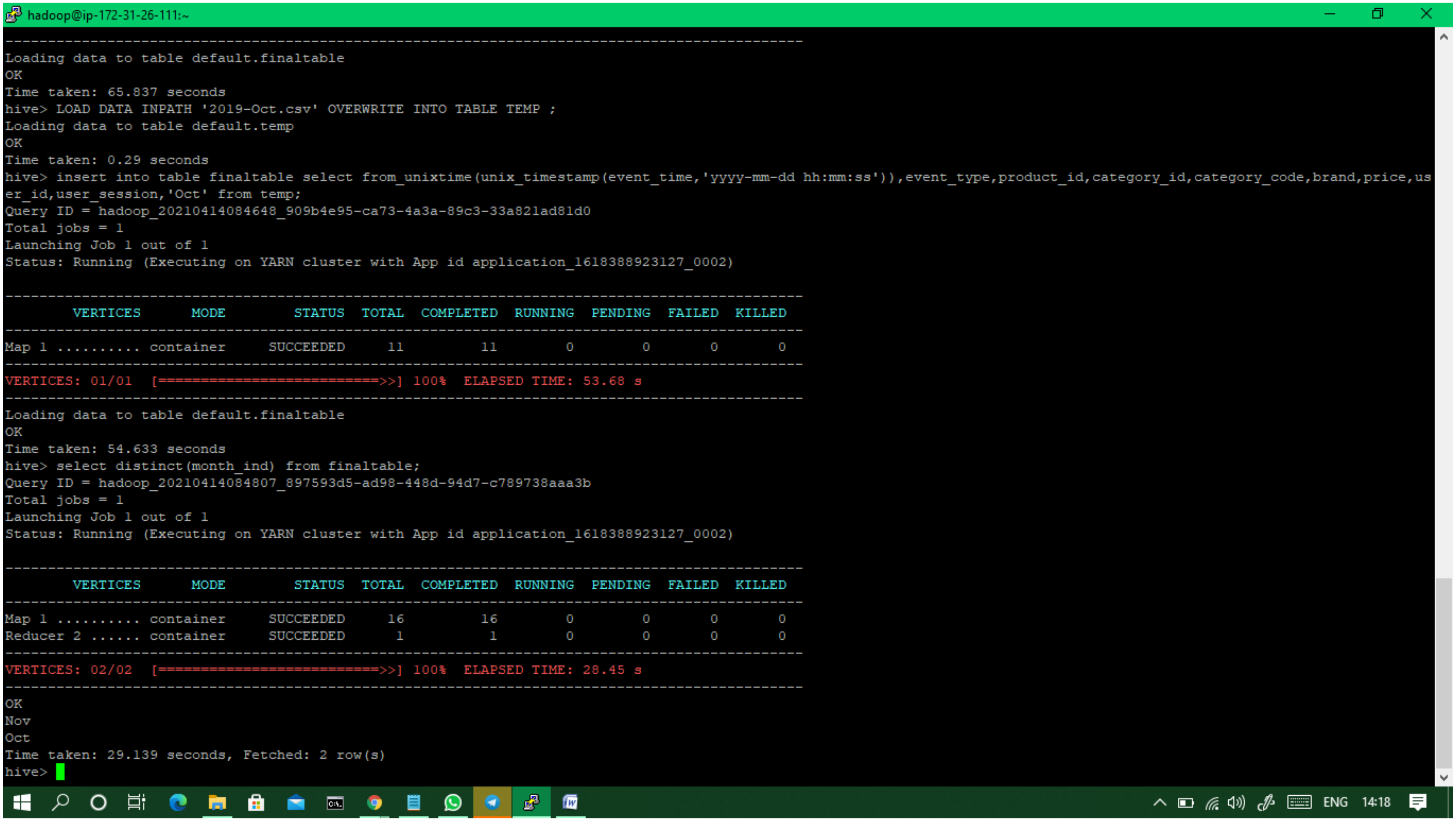
```
default
retaildb
Time taken: 0.732 seconds, Fetched: 2 row(s)
hive> use retaildb;
OK
Time taken: 0.041 seconds
hive> CREATE TABLE IF NOT EXISTS TEMP (
    > event_time  string  ,
    > event_type  string  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string ,
    > month_ind string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS TEXTFILE ;
OK
Time taken: 0.47 seconds
```

Create Finaltable to load data from temp table for both month combined. Below Screenshot contain load steps for November month data



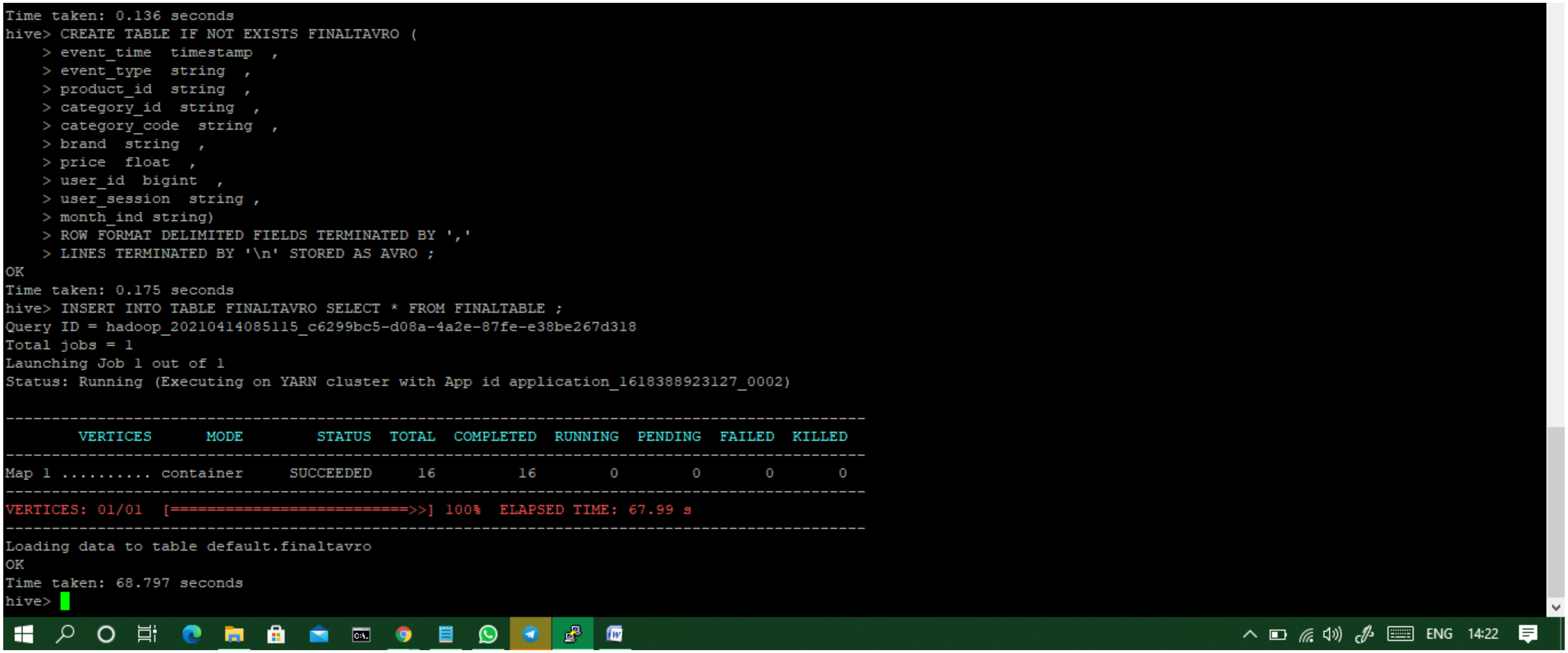
```
    > month_ind string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS TEXTFILE ;
OK
Time taken: 0.432 seconds
hive> LOAD DATA INPATH '2019-Nov.csv' OVERWRITE INTO TABLE TEMP ;
Loading data to table default.temp
OK
Time taken: 1.02 seconds
hive> CREATE TABLE IF NOT EXISTS FINALTABLE (
    > event_time  timestamp  ,
    > event_type  string  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string ,
    > month_ind string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS TEXTFILE ;
OK
Time taken: 0.075 seconds
hive> insert overwrite table finaltable select from_unixtime(unix_timestamp(event_time,'yyyy-mm-dd hh:mm:ss')),event_type,product_id,category_id,category_code,brand,pri
ce,user_id,user_session,'Nov' from temp;
Query ID = hadoop_20210414084304_c426d7f3-e59b-4e81-af24-e4ed056a24c0
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1618388923127_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     11         11        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 56.02 s
----------------------------------------------------------------------------------------------
Loading data to table default.finaltable
OK
Time taken: 65.837 seconds
hive>
```

Below screenshot contain data load for October month and query to confirm data load in Finaltable



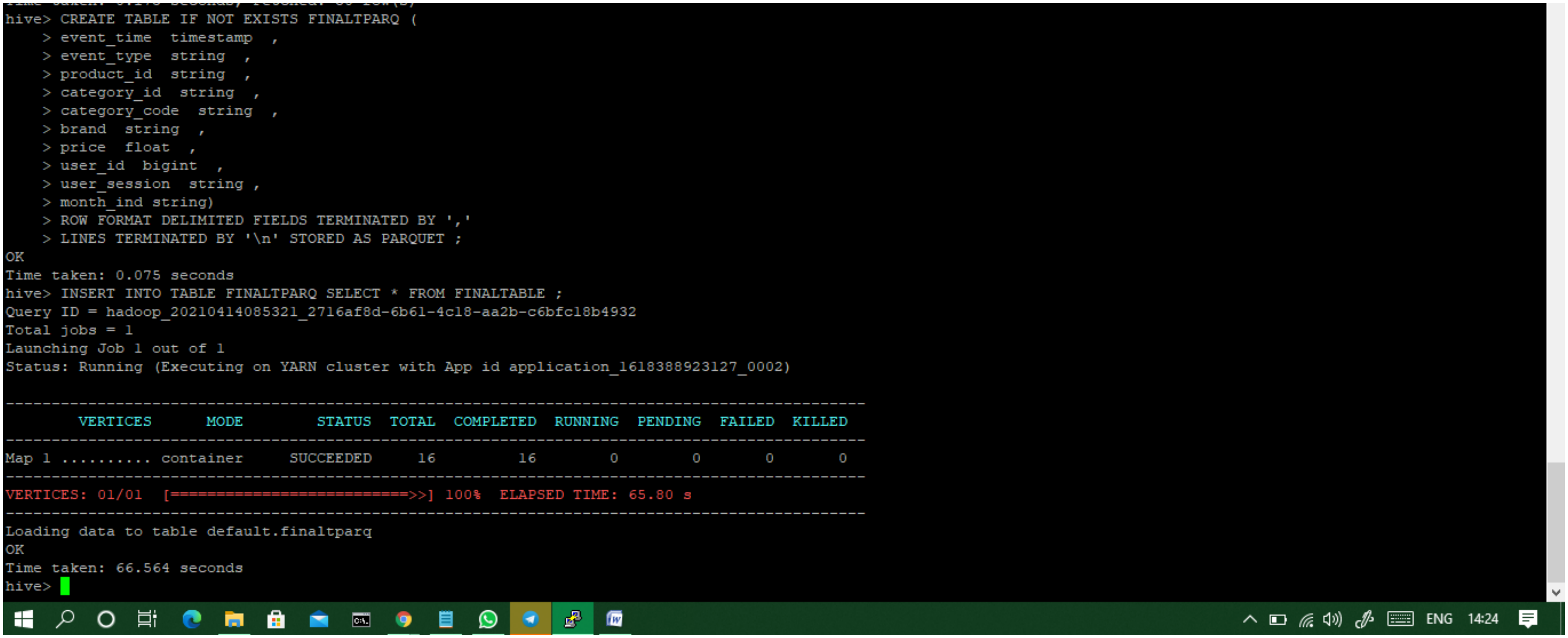
```
----------------------------------------------------------------------------------------------
Loading data to table default.finaltable
OK
Time taken: 65.837 seconds
hive> LOAD DATA INPATH '2019-Oct.csv' OVERWRITE INTO TABLE TEMP ;
Loading data to table default.temp
OK
Time taken: 0.29 seconds
hive> insert into table finaltable select from_unixtime(unix_timestamp(event_time,'yyyy-mm-dd hh:mm:ss')),event_type,product_id,category_id,category_code,brand,price,us
er_id,user_session,'Oct' from temp;
Query ID = hadoop_20210414084648_909b4e95-ca73-4a3a-89c3-33a821ad81d0
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     11         11        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 53.68 s
----------------------------------------------------------------------------------------------
Loading data to table default.finaltable
OK
Time taken: 54.633 seconds
hive> select distinct(month_ind) from finaltable;
Query ID = hadoop_20210414084807_897593d5-ad98-448d-94d7-c789738aaa3b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 28.45 s
----------------------------------------------------------------------------------------------
OK
Nov
Oct
Time taken: 29.139 seconds, Fetched: 2 row(s)
hive>
```

Creating AVRO table and loading the data



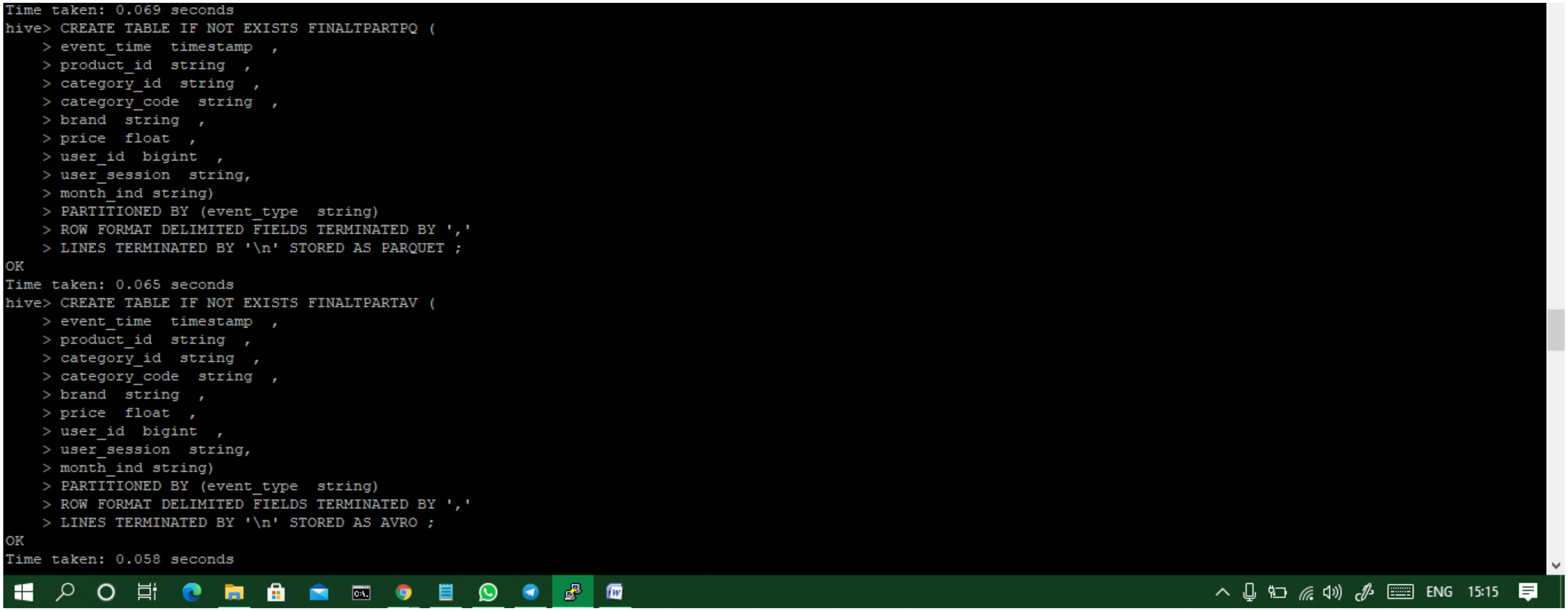
```
Time taken: 0.136 seconds
hive> CREATE TABLE IF NOT EXISTS FINALTAVRO (
    > event_time  timestamp  ,
    > event_type  string  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string ,
    > month_ind string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS AVRO ;
OK
Time taken: 0.175 seconds
hive> INSERT INTO TABLE FINALTAVRO SELECT * FROM FINALTABLE ;
Query ID = hadoop_20210414085115_c6299bc5-d08a-4a2e-87fe-e38be267d318
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 67.99 s
----------------------------------------------------------------------------------------------
Loading data to table default.finaltavro
OK
Time taken: 68.797 seconds
hive> 
```

Creating Parquet table and load data into it



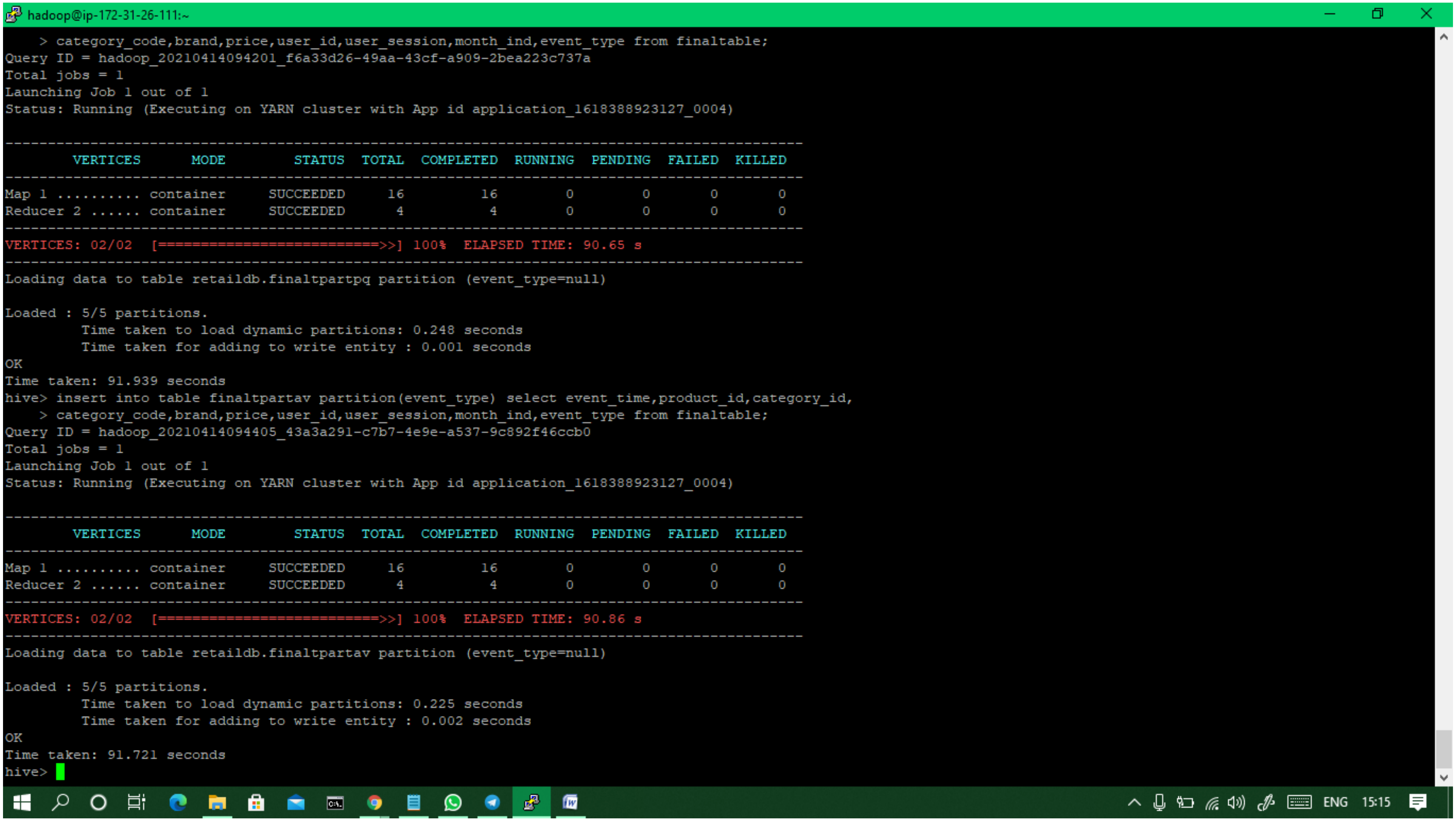
```
Time taken: 0.175 seconds, Fetched: 60 row(s)
hive> CREATE TABLE IF NOT EXISTS FINALTPARQ (
    > event_time  timestamp  ,
    > event_type  string  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string ,
    > month_ind string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS PARQUET ;
OK
Time taken: 0.075 seconds
hive> INSERT INTO TABLE FINALTPARQ SELECT * FROM FINALTABLE ;
Query ID = hadoop_20210414085321_2716af8d-6b61-4c18-aa2b-c6bfc18b4932
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 65.80 s
----------------------------------------------------------------------------------------------
Loading data to table default.finaltparq
OK
Time taken: 66.564 seconds
hive>
```

Create Partition table in Parquet and AVRO format



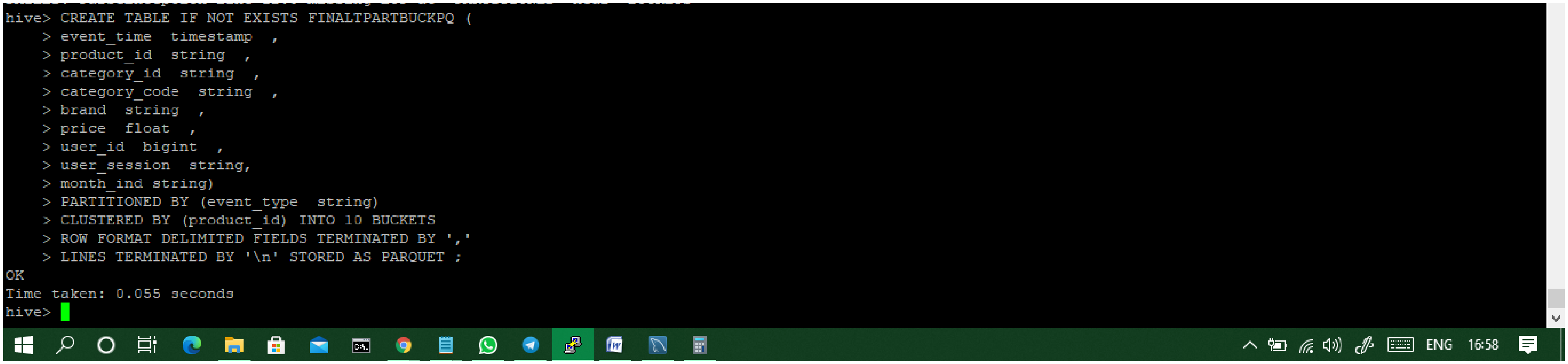
```
Time taken: 0.069 seconds
hive> CREATE TABLE IF NOT EXISTS FINALTPARTPQ (
    > event_time  timestamp  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string,
    > month_ind string)
    > PARTITIONED BY (event_type  string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS PARQUET ;
OK
Time taken: 0.065 seconds
hive> CREATE TABLE IF NOT EXISTS FINALTPARTAV (
    > event_time  timestamp  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string,
    > month_ind string)
    > PARTITIONED BY (event_type  string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS AVRO ;
OK
Time taken: 0.058 seconds
```

Load data into Parquet and AVRO partition table



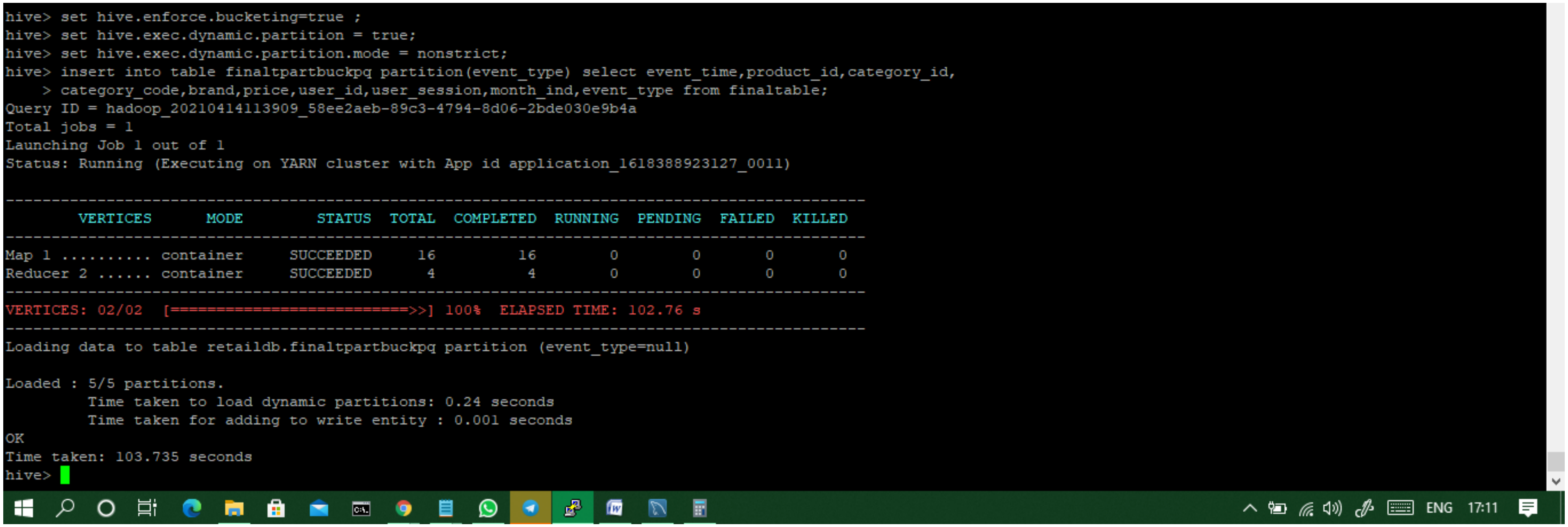
```
    > category_code,brand,price,user_id,user_session,month_ind,event_type from finaltable;
Query ID = hadoop_20210414094201_f6a33d26-49aa-43cf-a909-2bea223c737a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      4          4        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 90.65 s
----------------------------------------------------------------------------------------------
Loading data to table retaildb.finaltpartpq partition (event_type=null)

Loaded : 5/5 partitions.
         Time taken to load dynamic partitions: 0.248 seconds
         Time taken for adding to write entity : 0.001 seconds
OK
Time taken: 91.939 seconds
hive> insert into table finaltpartav partition(event_type) select event_time,product_id,category_id,
    > category_code,brand,price,user_id,user_session,month_ind,event_type from finaltable;
Query ID = hadoop_20210414094405_43a3a291-c7b7-4e9e-a537-9c892f46ccb0
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      4          4        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 90.86 s
----------------------------------------------------------------------------------------------
Loading data to table retaildb.finaltpartav partition (event_type=null)

Loaded : 5/5 partitions.
         Time taken to load dynamic partitions: 0.225 seconds
         Time taken for adding to write entity : 0.002 seconds
OK
Time taken: 91.721 seconds
hive>
```

Create Partition and Bucket under Parquet format table



```
hive> CREATE TABLE IF NOT EXISTS FINALTPARTBUCKPQ (
    > event_time  timestamp  ,
    > product_id  string  ,
    > category_id  string  ,
    > category_code  string  ,
    > brand  string  ,
    > price  float  ,
    > user_id  bigint  ,
    > user_session  string,
    > month_ind string)
    > PARTITIONED BY (event_type  string)
    > CLUSTERED BY (product_id) INTO 10 BUCKETS
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n' STORED AS PARQUET ;
OK
Time taken: 0.055 seconds
hive>
```

Load data into Parquet table partitioned and bucketed

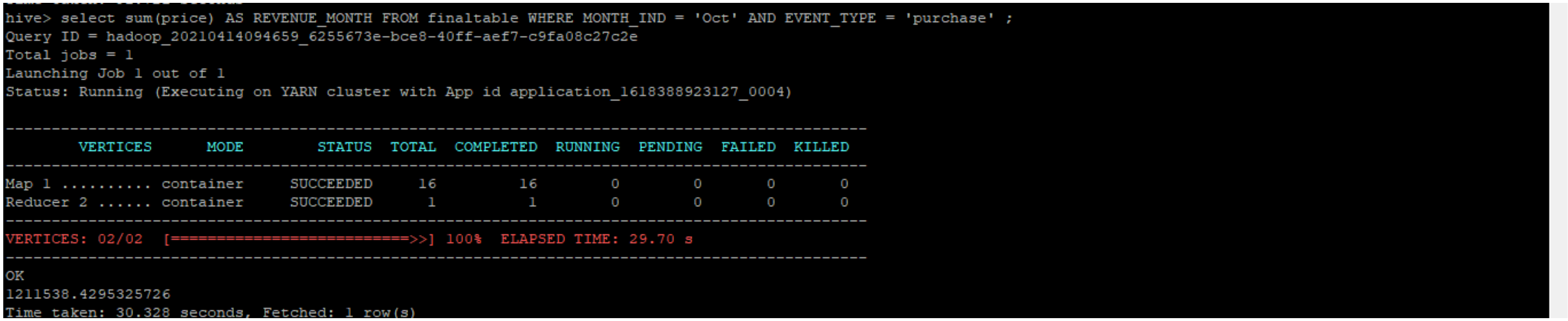


```
hive> set hive.enforce.bucketing=true ;
hive> set hive.exec.dynamic.partition = true;
hive> set hive.exec.dynamic.partition.mode = nonstrict;
hive> insert into table finaltpartbuckpq partition(event_type) select event_time,product_id,category_id,
    > category_code,brand,price,user_id,user_session,month_ind,event_type from finaltable;
Query ID = hadoop_20210414113909_58ee2aeb-89c3-4794-8d06-2bde030e9b4a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      4          4        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 102.76 s
----------------------------------------------------------------------------------------------
Loading data to table retaildb.finaltpartbuckpq partition (event_type=null)

Loaded : 5/5 partitions.
         Time taken to load dynamic partitions: 0.24 seconds
         Time taken for adding to write entity : 0.001 seconds
OK
Time taken: 103.735 seconds
hive>
```

Optimization techniques

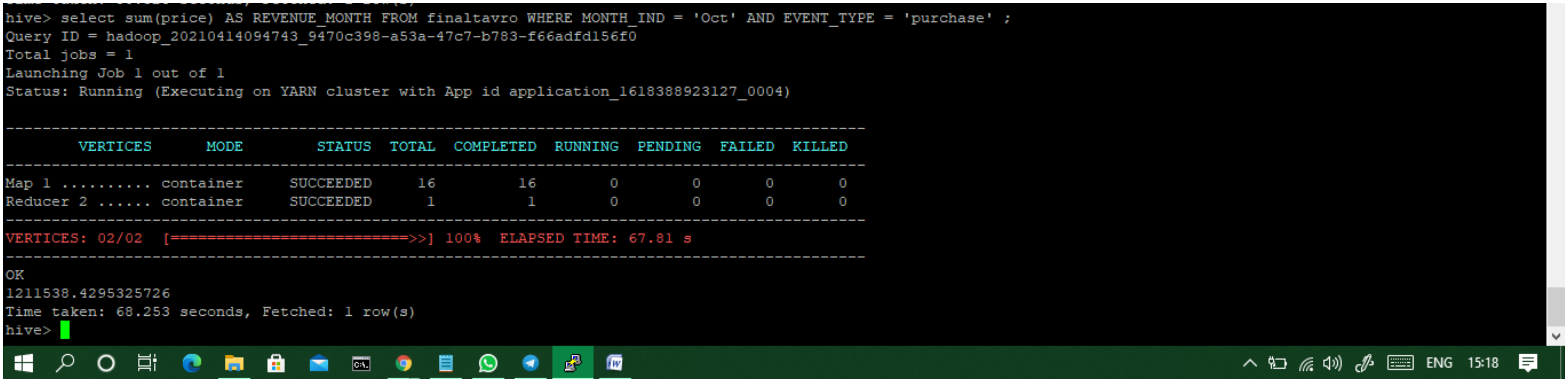
Outcome in seconds for 1st query from different tables

Query Output from text format raw table :-



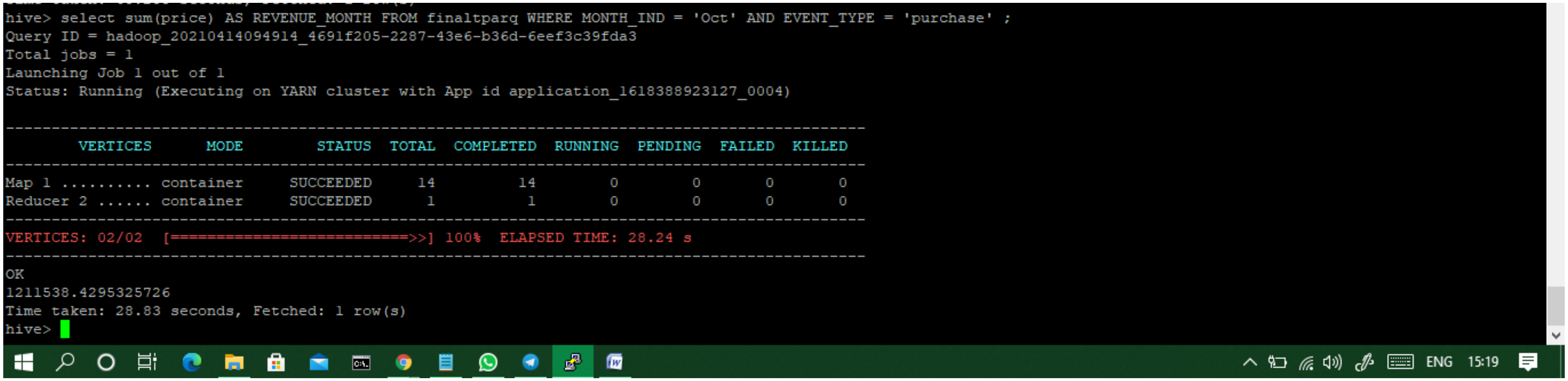
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltable WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094659_6255673e-bce8-40ff-aef7-c9fa08c27c2e
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 29.70 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 30.328 seconds, Fetched: 1 row(s)
```

From AVRO format table :-



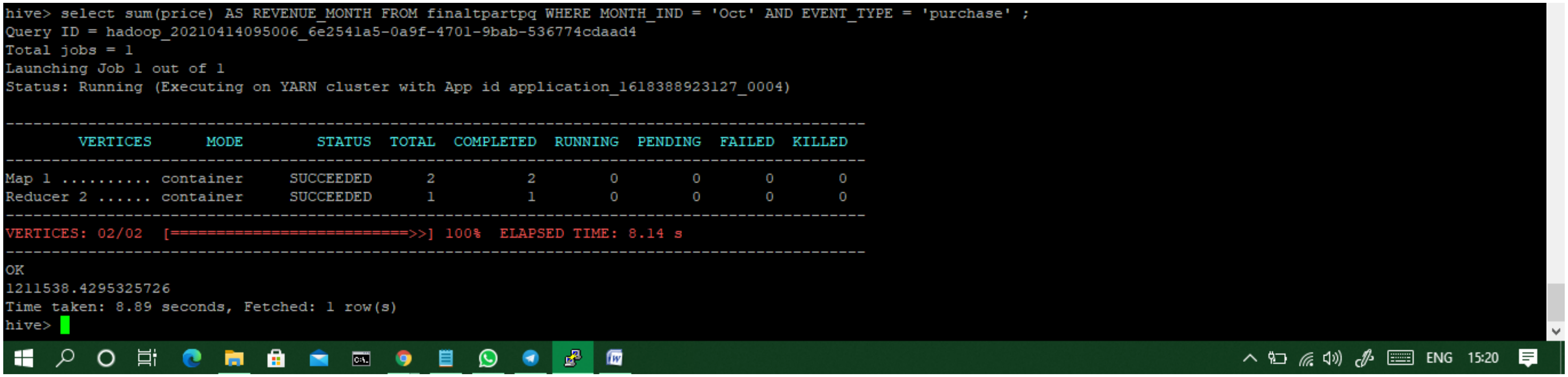
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltavro WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094743_9470c398-a53a-47c7-b783-f66adfd156f0
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 67.81 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 68.253 seconds, Fetched: 1 row(s)
hive>
```

From Parquet format table :-



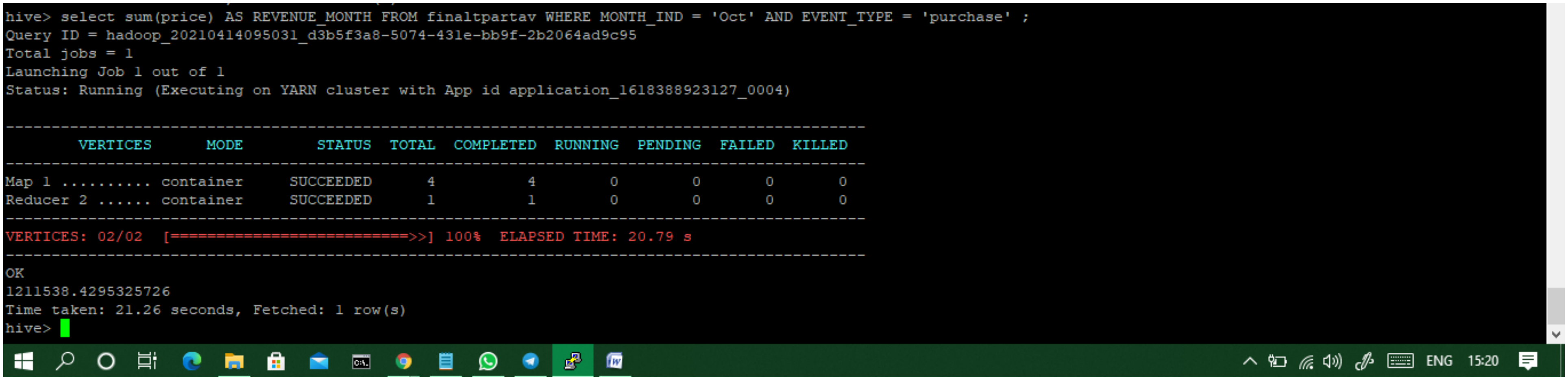
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltparq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094914_4691f205-2287-43e6-b36d-6eef3c39fda3
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     14         14        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 28.24 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 28.83 seconds, Fetched: 1 row(s)
hive>
```

From Parquet format Partitioned table :-



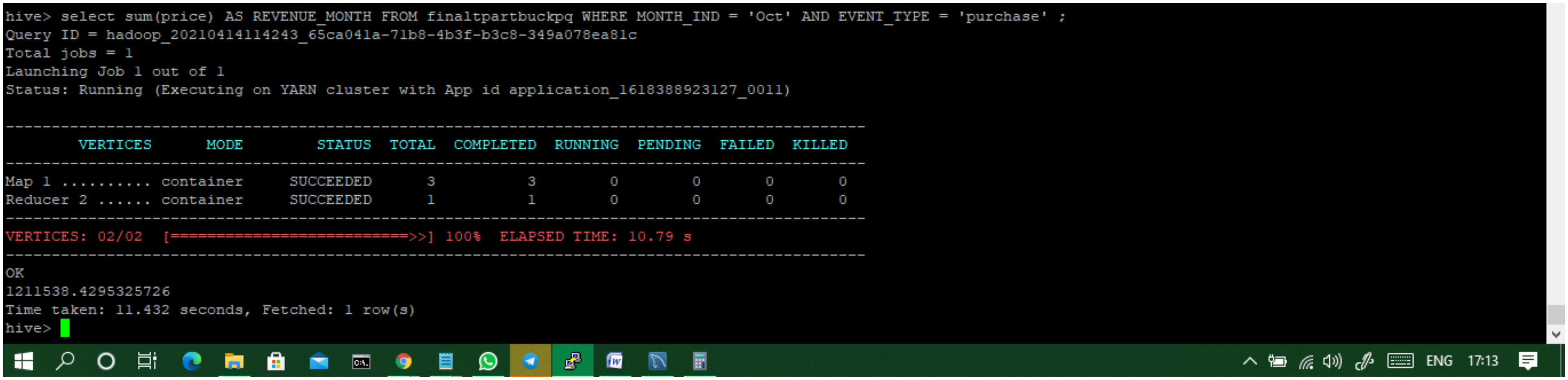
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartpq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414095006_6e2541a5-0a9f-4701-9bab-536774cdaad4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.14 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 8.89 seconds, Fetched: 1 row(s)
hive>
```

From AVRO format Partition table :-



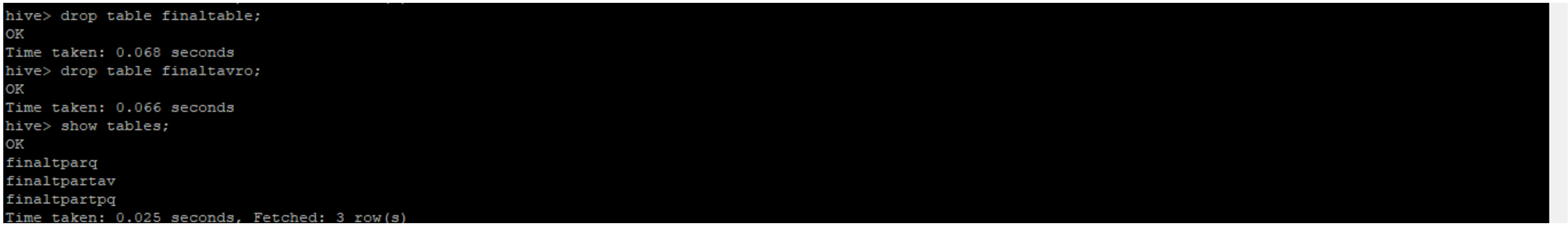
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartav WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414095031_d3b5f3a8-5074-431e-bb9f-2b2064ad9c95
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      4          4        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 20.79 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 21.26 seconds, Fetched: 1 row(s)
hive>
```

From Parquet format Partitioned and bucketed table :-

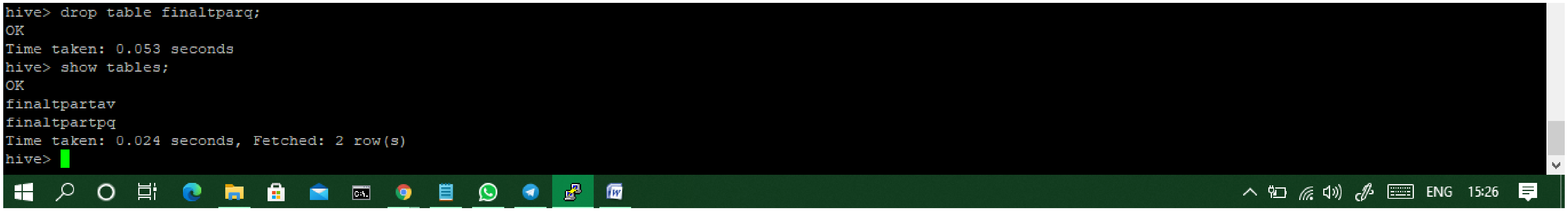


```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartbuckpq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414114243_65ca041a-71b8-4b3f-b3c8-349a078ea81c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 10.79 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 11.432 seconds, Fetched: 1 row(s)
hive>
```

Dropping few tables since their output took comparatively more time



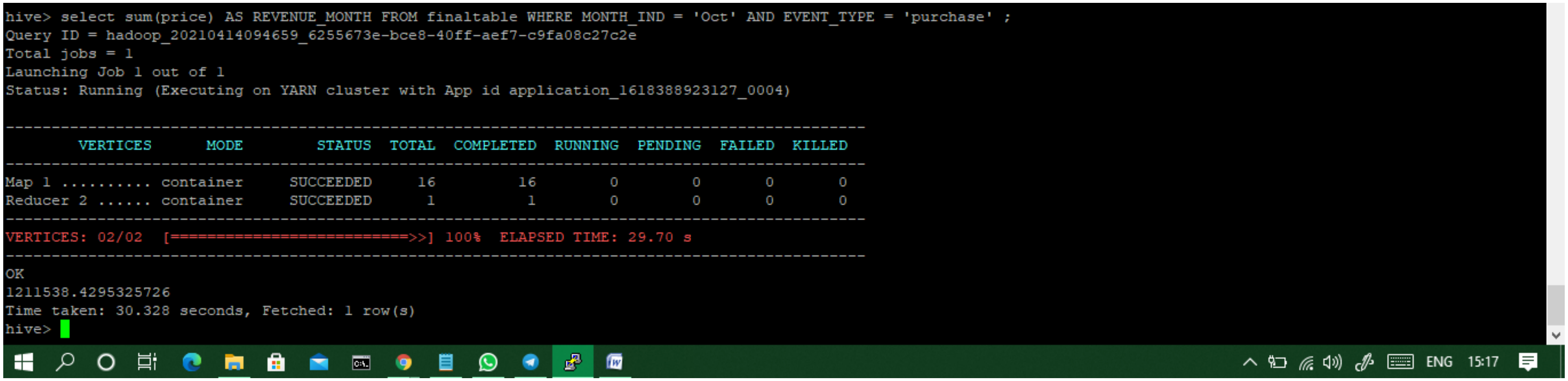
```
hive> drop table finaltable;
OK
Time taken: 0.068 seconds
hive> drop table finaltavro;
OK
Time taken: 0.066 seconds
hive> show tables;
OK
finaltparq
finaltpartav
finaltpartpq
Time taken: 0.025 seconds, Fetched: 3 row(s)
```



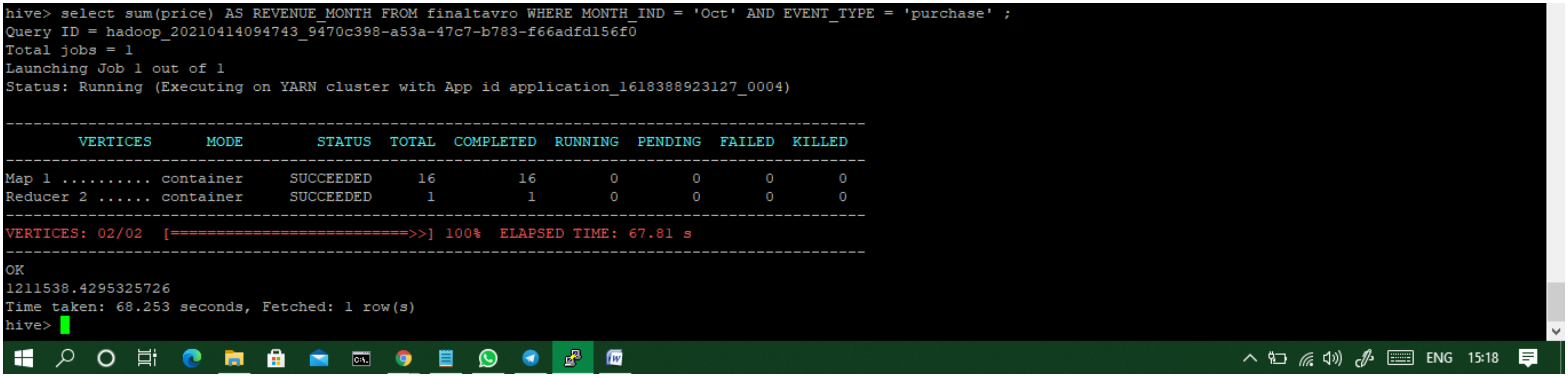
```
hive> drop table finaltparq;
OK
Time taken: 0.053 seconds
hive> show tables;
OK
finaltpartav
finaltpartpq
Time taken: 0.024 seconds, Fetched: 2 row(s)
hive>
```

Analysis using Hive Queries :-

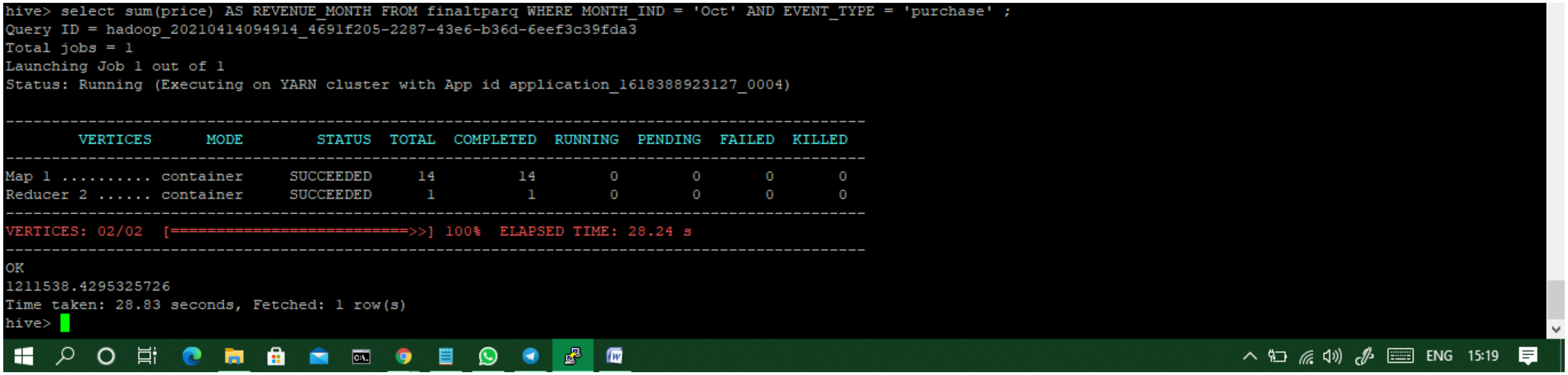
1) Find the total revenue generated due to the purchases made in October.



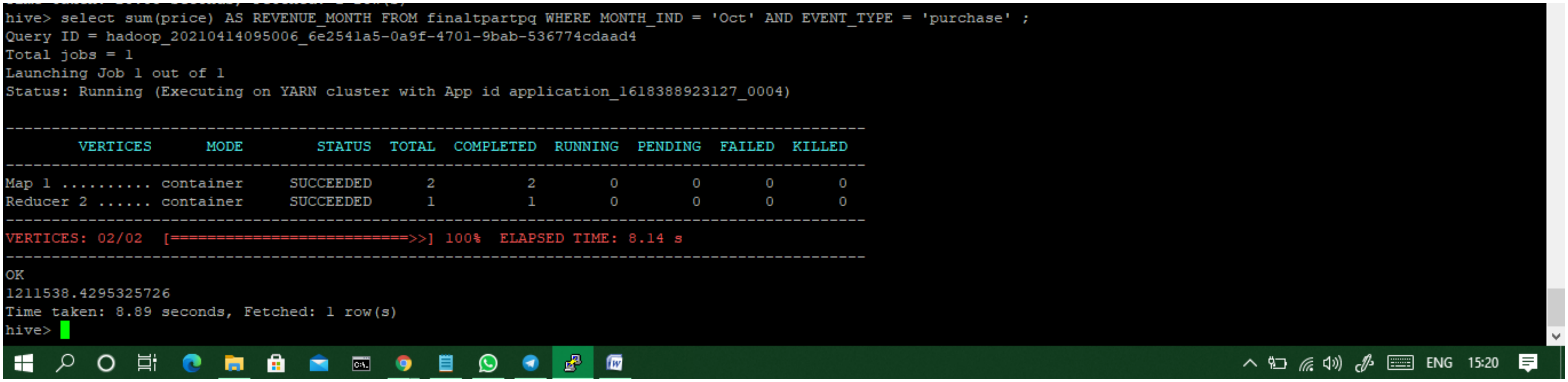
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltable WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094659_6255673e-bce8-40ff-aef7-c9fa08c27c2e
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 29.70 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 30.328 seconds, Fetched: 1 row(s)
hive>
```



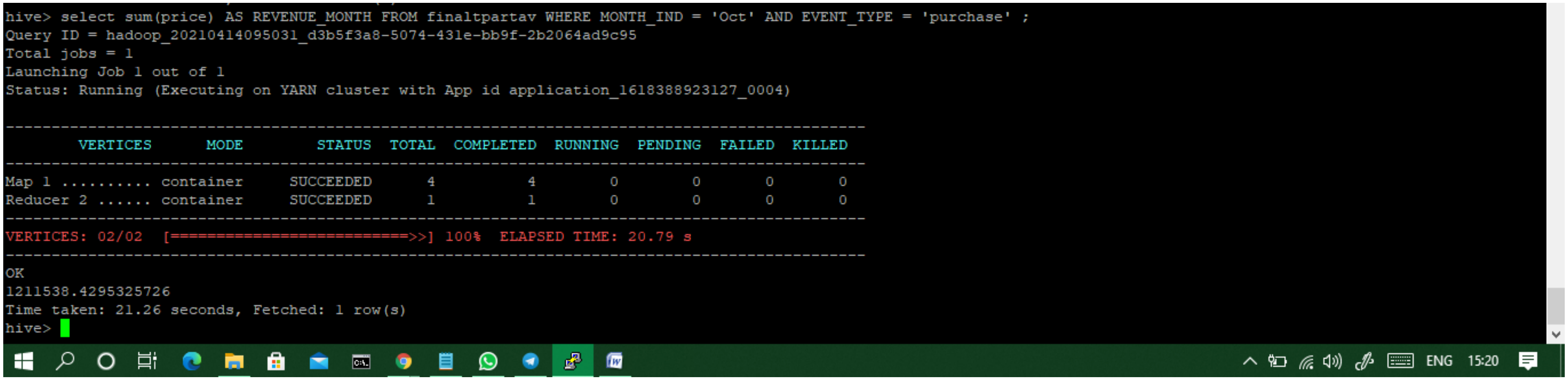
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltavro WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094743_9470c398-a53a-47c7-b783-f66adfd156f0
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 67.81 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 68.253 seconds, Fetched: 1 row(s)
hive>
```



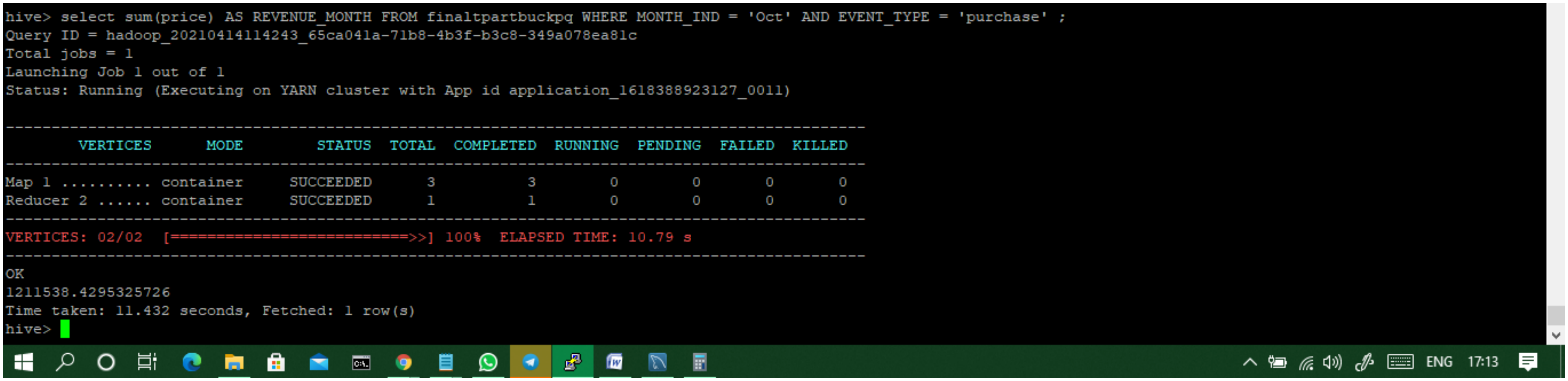
```
hive> select sum(price) AS REVENUE_MONTH FROM finaltparq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414094914_4691f205-2287-43e6-b36d-6eef3c39fda3
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     14         14        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 28.24 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 28.83 seconds, Fetched: 1 row(s)
hive>
```



```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartpq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414095006_6e2541a5-0a9f-4701-9bab-536774cdaad4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.14 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 8.89 seconds, Fetched: 1 row(s)
hive>
```

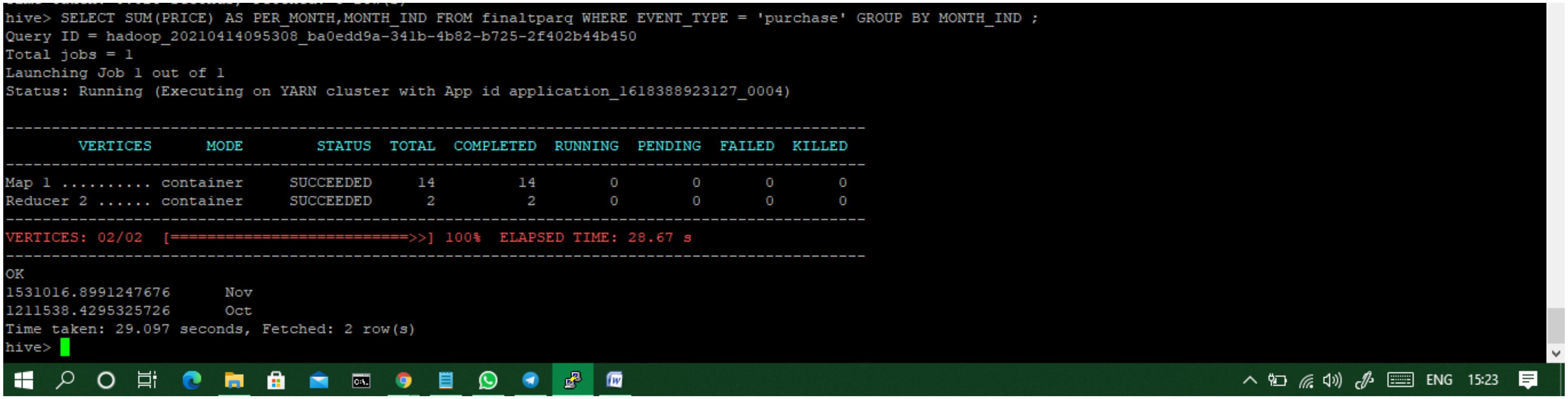


```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartav WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414095031_d3b5f3a8-5074-431e-bb9f-2b2064ad9c95
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      4          4        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 20.79 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 21.26 seconds, Fetched: 1 row(s)
hive>
```

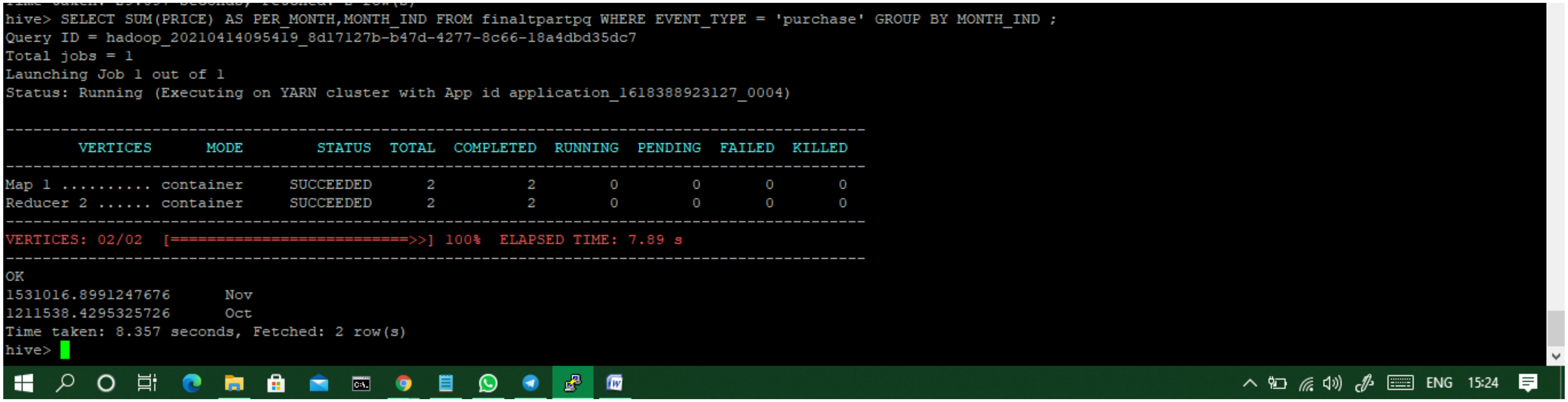


```
hive> select sum(price) AS REVENUE_MONTH FROM finaltpartbuckpq WHERE MONTH_IND = 'Oct' AND EVENT_TYPE = 'purchase' ;
Query ID = hadoop_20210414114243_65ca041a-71b8-4b3f-b3c8-349a078ea81c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 10.79 s
----------------------------------------------------------------------------------------------
OK
1211538.4295325726
Time taken: 11.432 seconds, Fetched: 1 row(s)
hive>
```

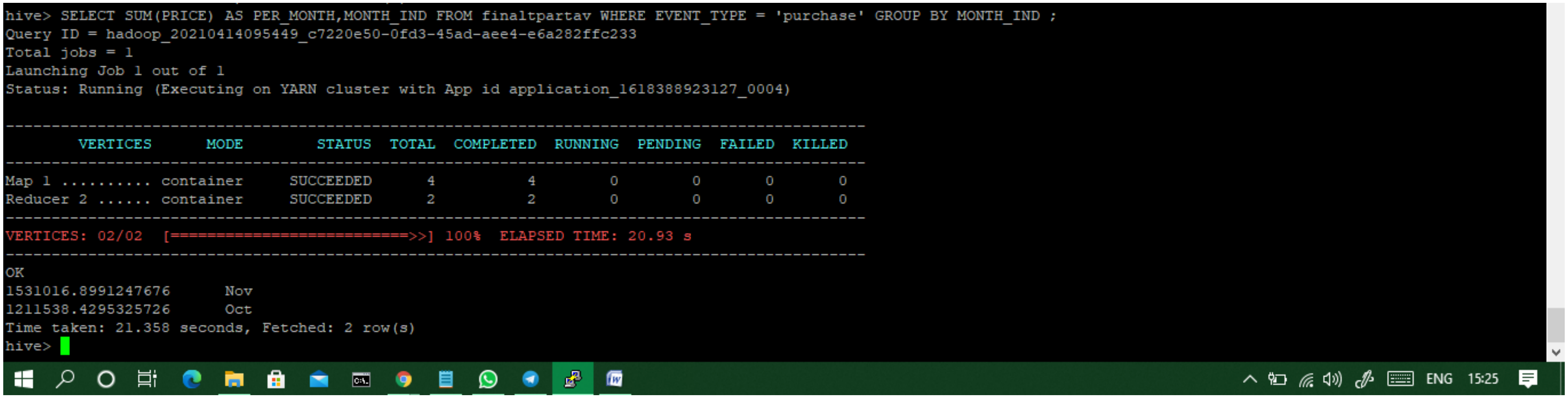
2) Write a query to yield the total sum of purchases per month in a single output.



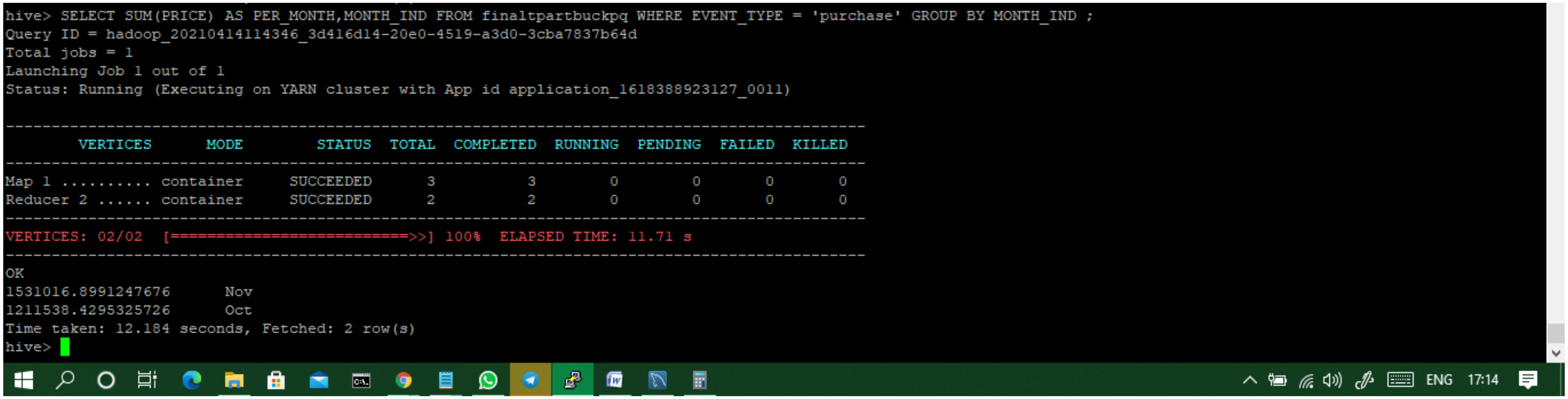
```
hive> SELECT SUM(PRICE) AS PER_MONTH,MONTH_IND FROM finaltparq WHERE EVENT_TYPE = 'purchase' GROUP BY MONTH_IND ;
Query ID = hadoop_20210414095308_ba0edd9a-341b-4b82-b725-2f402b44b450
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     14         14        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 28.67 s
----------------------------------------------------------------------------------------------
OK
1531016.8991247676      Nov
1211538.4295325726      Oct
Time taken: 29.097 seconds, Fetched: 2 row(s)
hive>
```



```
hive> SELECT SUM(PRICE) AS PER_MONTH,MONTH_IND FROM finaltpartpq WHERE EVENT_TYPE = 'purchase' GROUP BY MONTH_IND ;
Query ID = hadoop_20210414095419_8d17127b-b47d-4277-8c66-18a4dbd35dc7
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 7.89 s
----------------------------------------------------------------------------------------------
OK
1531016.8991247676      Nov
1211538.4295325726      Oct
Time taken: 8.357 seconds, Fetched: 2 row(s)
hive>
```

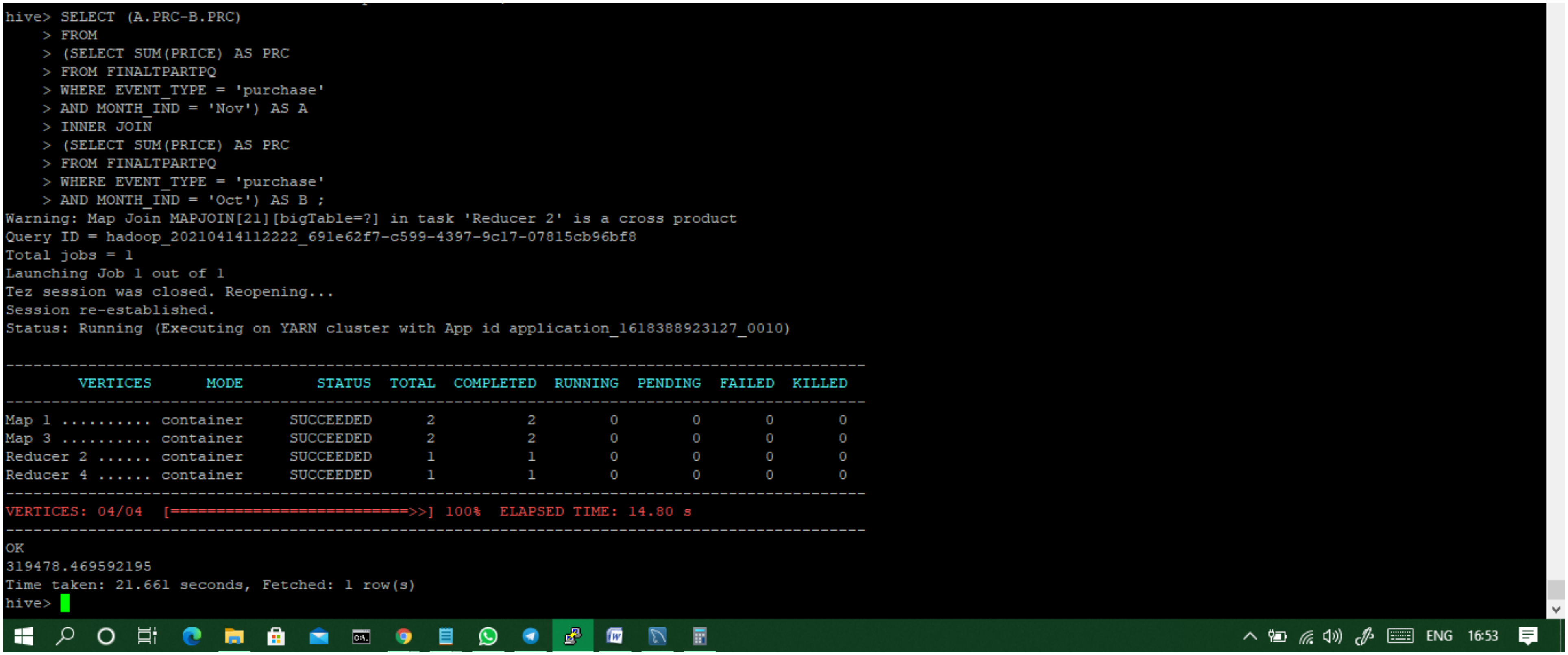


```
hive> SELECT SUM(PRICE) AS PER_MONTH,MONTH_IND FROM finaltpartav WHERE EVENT_TYPE = 'purchase' GROUP BY MONTH_IND ;
Query ID = hadoop_20210414095449_c7220e50-0fd3-45ad-aee4-e6a282ffc233
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0004)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      4          4        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 20.93 s
----------------------------------------------------------------------------------------------
OK
1531016.8991247676      Nov
1211538.4295325726      Oct
Time taken: 21.358 seconds, Fetched: 2 row(s)
hive>
```

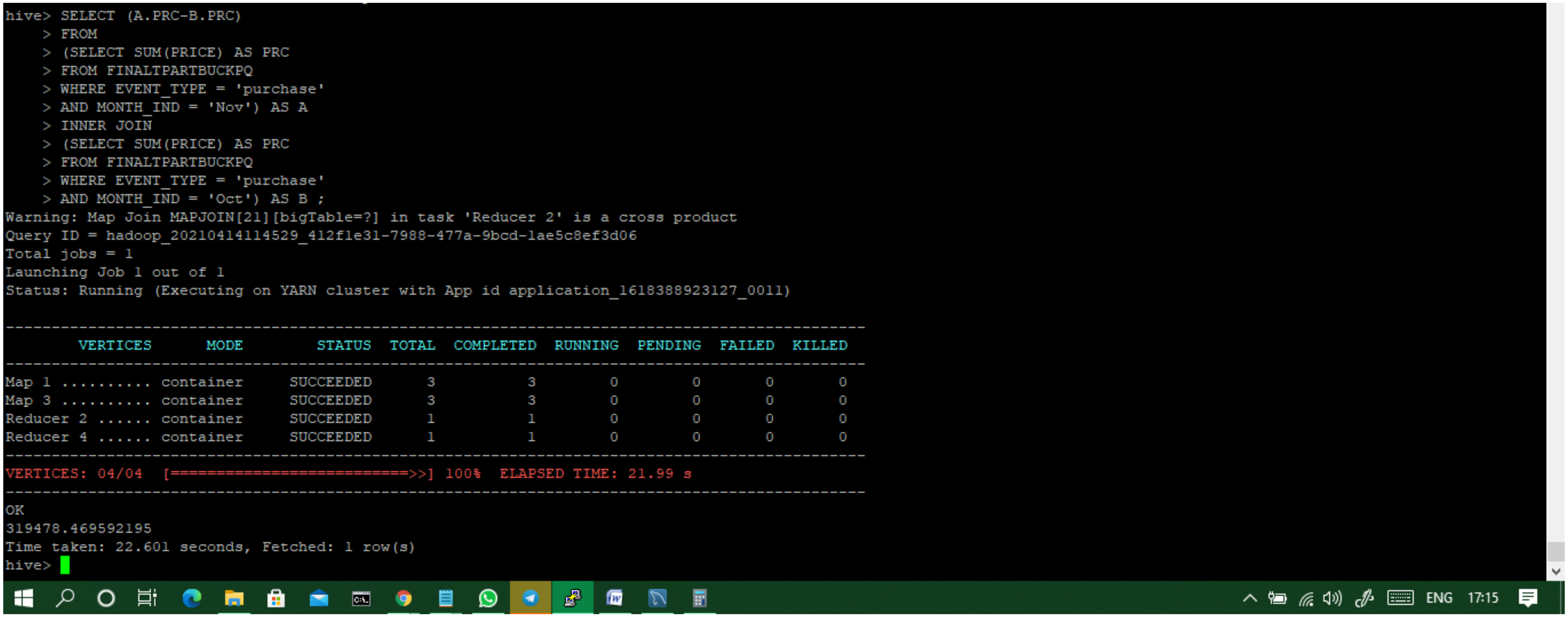


```
hive> SELECT SUM(PRICE) AS PER_MONTH,MONTH_IND FROM finaltpartbuckpq WHERE EVENT_TYPE = 'purchase' GROUP BY MONTH_IND ;
Query ID = hadoop_20210414114346_3d416d14-20e0-4519-a3d0-3cba7837b64d
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 11.71 s
----------------------------------------------------------------------------------------------
OK
1531016.8991247676      Nov
1211538.4295325726      Oct
Time taken: 12.184 seconds, Fetched: 2 row(s)
hive>
```

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



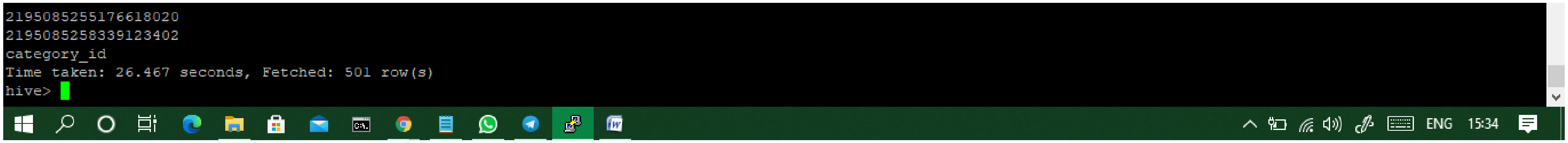
```
hive> SELECT (A.PRC-B.PRC)
    > FROM
    > (SELECT SUM(PRICE) AS PRC
    > FROM FINALTPARTPQ
    > WHERE EVENT_TYPE = 'purchase'
    > AND MONTH_IND = 'Nov') AS A
    > INNER JOIN
    > (SELECT SUM(PRICE) AS PRC
    > FROM FINALTPARTPQ
    > WHERE EVENT_TYPE = 'purchase'
    > AND MONTH_IND = 'Oct') AS B ;
Warning: Map Join MAPJOIN[21][bigTable=?] in task 'Reducer 2' is a cross product
Query ID = hadoop_20210414112222_691e62f7-c599-4397-9c17-07815cb96bf8
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1618388923127_0010)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Map 3 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
Reducer 4 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 04/04  [==========================>>] 100%  ELAPSED TIME: 14.80 s
----------------------------------------------------------------------------------------------
OK
319478.469592195
Time taken: 21.661 seconds, Fetched: 1 row(s)
hive>
```



```
hive> SELECT (A.PRC-B.PRC)
    > FROM
    > (SELECT SUM(PRICE) AS PRC
    > FROM FINALTPARTBUCKPQ
    > WHERE EVENT_TYPE = 'purchase'
    > AND MONTH_IND = 'Nov') AS A
    > INNER JOIN
    > (SELECT SUM(PRICE) AS PRC
    > FROM FINALTPARTBUCKPQ
    > WHERE EVENT_TYPE = 'purchase'
    > AND MONTH_IND = 'Oct') AS B ;
Warning: Map Join MAPJOIN[21][bigTable=?] in task 'Reducer 2' is a cross product
Query ID = hadoop_20210414114529_412f1e31-7988-477a-9bcd-1ae5c8ef3d06
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Map 3 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
Reducer 4 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 04/04  [==========================>>] 100%  ELAPSED TIME: 21.99 s
----------------------------------------------------------------------------------------------
OK
319478.469592195
Time taken: 22.601 seconds, Fetched: 1 row(s)
hive>
```

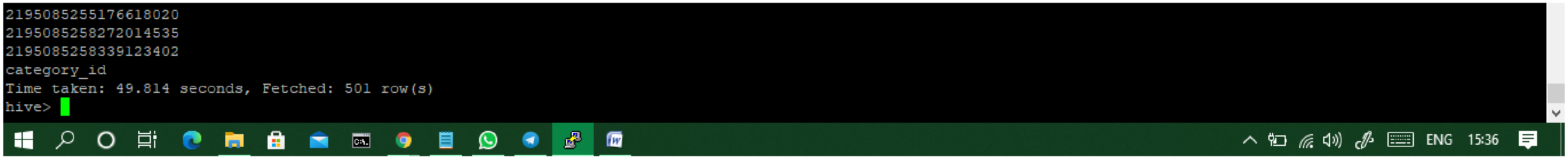
4) Find distinct categories of products.

SELECT DISTINCT(CATEGORY_ID) FROM finaltpartpq ;

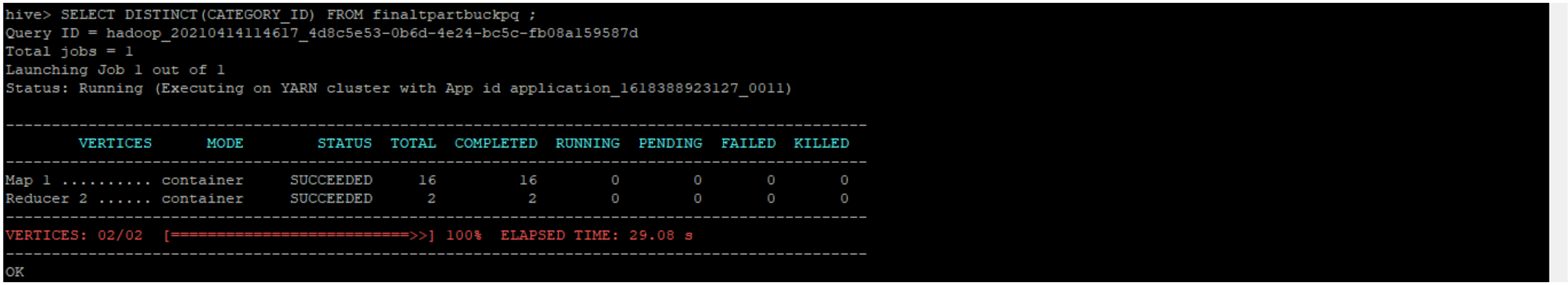


```
2195085255176618020
2195085258339123402
category_id
Time taken: 26.467 seconds, Fetched: 501 row(s)
hive>
```

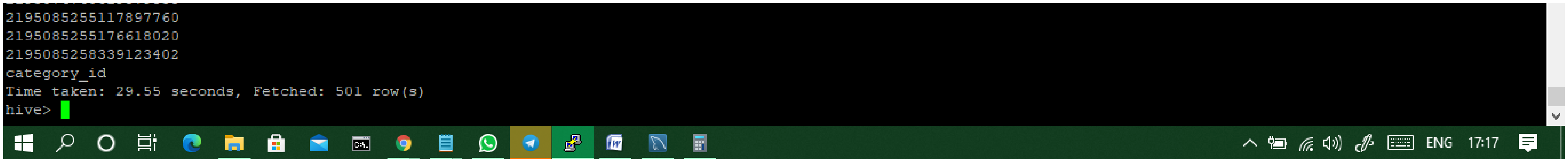
SELECT DISTINCT(CATEGORY_ID) FROM finaltpartav ;



```
2195085255176618020
2195085258272014535
2195085258339123402
category_id
Time taken: 49.814 seconds, Fetched: 501 row(s)
hive> 
```

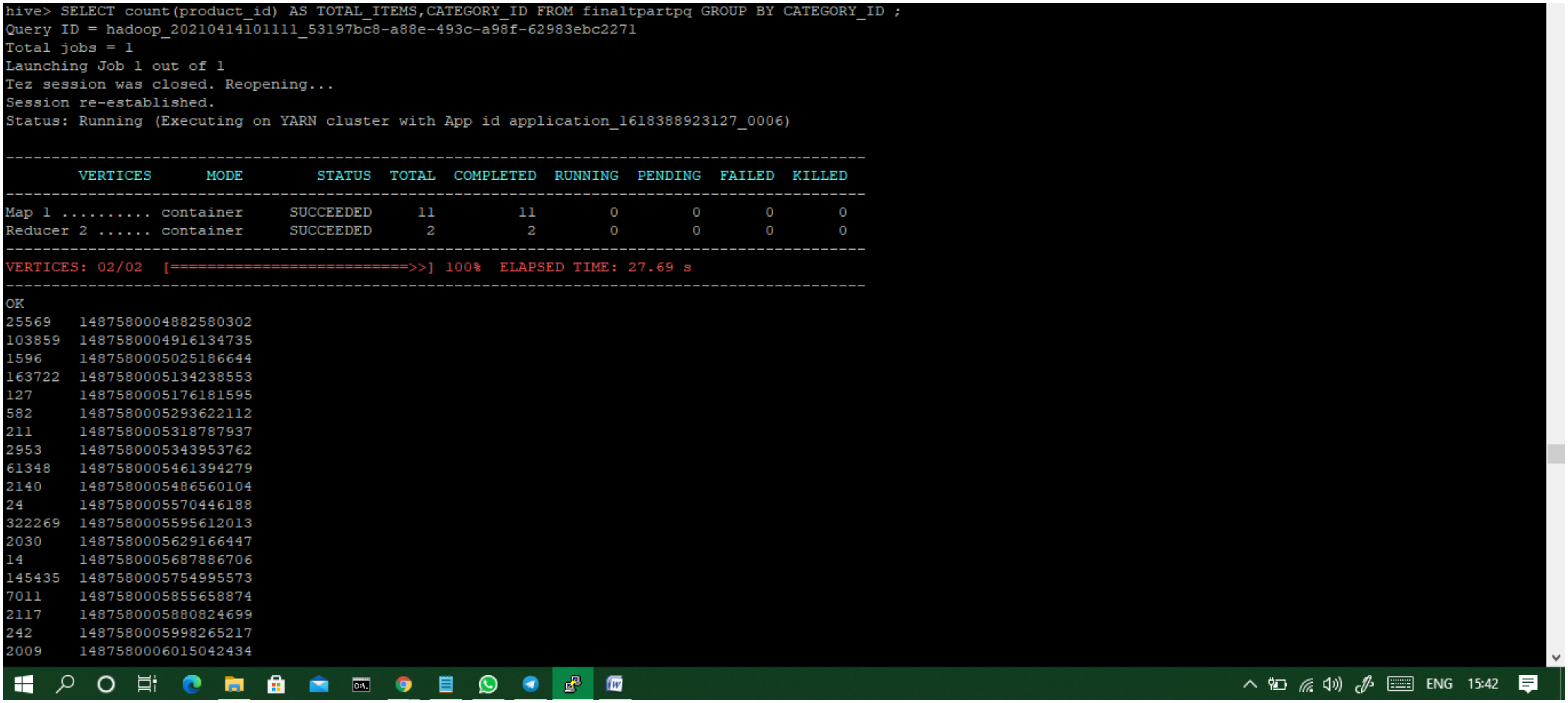


```
hive> SELECT DISTINCT(CATEGORY_ID) FROM finaltpartbuckpq ;
Query ID = hadoop_20210414114617_4d8c5e53-0b6d-4e24-bc5c-fb08a159587d
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 29.08 s
----------------------------------------------------------------------------------------------
OK
```

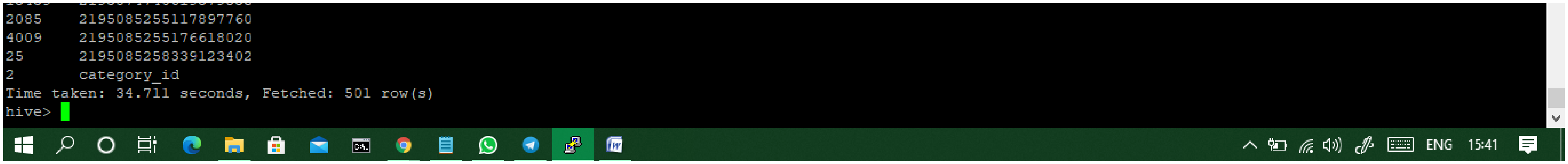


```
2195085255117897760
2195085255176618020
2195085258339123402
category_id
Time taken: 29.55 seconds, Fetched: 501 row(s)
hive> 
```

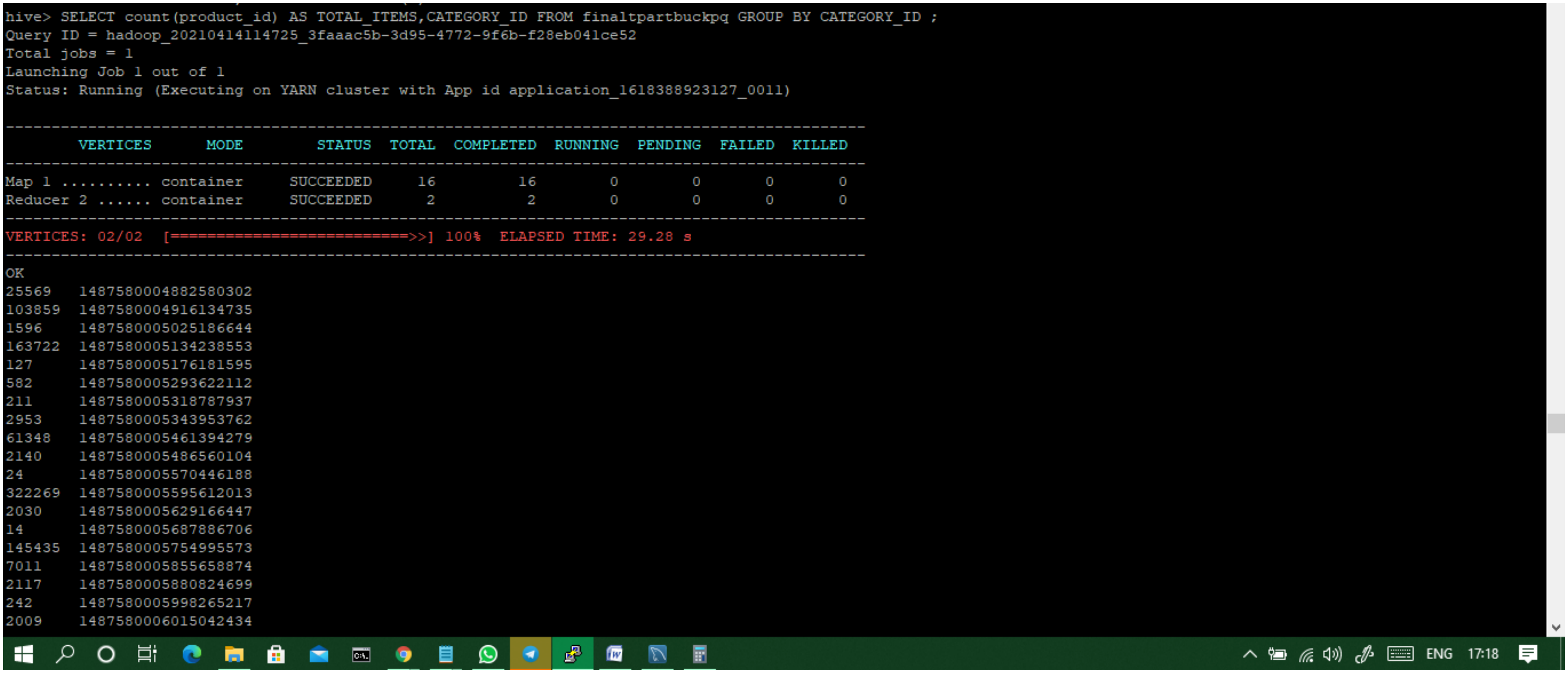
5) Find the total number of products available under each category.



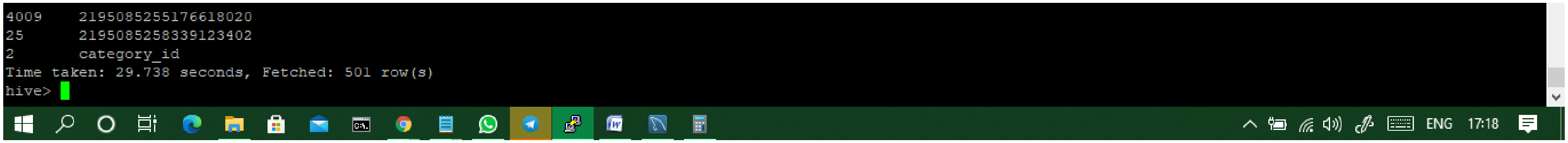
```
hive> SELECT count(product_id) AS TOTAL_ITEMS,CATEGORY_ID FROM finaltpartpq GROUP BY CATEGORY_ID ;
Query ID = hadoop_20210414101111_53197bc8-a88e-493c-a98f-62983ebc2271
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1618388923127_0006)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     11         11        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 27.69 s
----------------------------------------------------------------------------------------------
OK
25569   1487580004882580302
103859  1487580004916134735
1596    1487580005025186644
163722  1487580005134238553
127     1487580005176181595
582     1487580005293622112
211     1487580005318787937
2953    1487580005343953762
61348   1487580005461394279
2140    1487580005486560104
24      1487580005570446188
322269  1487580005595612013
2030    1487580005629166447
14      1487580005687886706
145435  1487580005754995573
7011    1487580005855658874
2117    1487580005880824699
242     1487580005998265217
2009    1487580006015042434
```



```
2085    2195085255117897760
4009    2195085255176618020
25      2195085258339123402
2       category_id
Time taken: 34.711 seconds, Fetched: 501 row(s)
hive> 
```

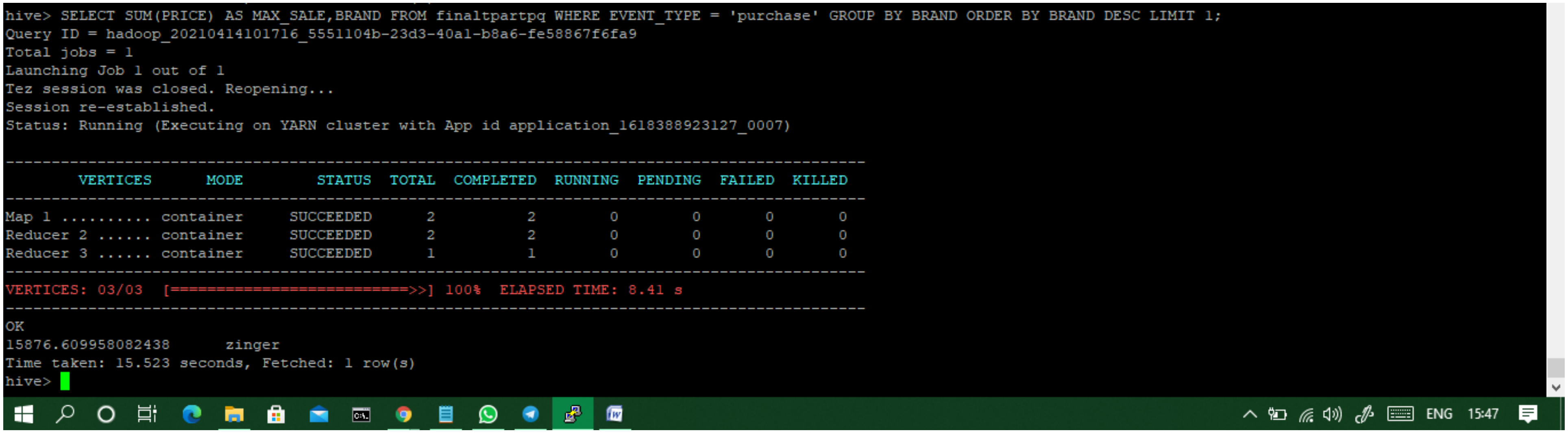


```
hive> SELECT count(product_id) AS TOTAL_ITEMS,CATEGORY_ID FROM finaltpartbuckpq GROUP BY CATEGORY_ID ;
Query ID = hadoop_20210414114725_3faaac5b-3d95-4772-9f6b-f28eb041ce52
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     16         16        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 29.28 s
----------------------------------------------------------------------------------------------
OK
25569   1487580004882580302
103859  1487580004916134735
1596    1487580005025186644
163722  1487580005134238553
127     1487580005176181595
582     1487580005293622112
211     1487580005318787937
2953    1487580005343953762
61348   1487580005461394279
2140    1487580005486560104
24      1487580005570446188
322269  1487580005595612013
2030    1487580005629166447
14      1487580005687886706
145435  1487580005754995573
7011    1487580005855658874
2117    1487580005880824699
242     1487580005998265217
2009    1487580006015042434
```

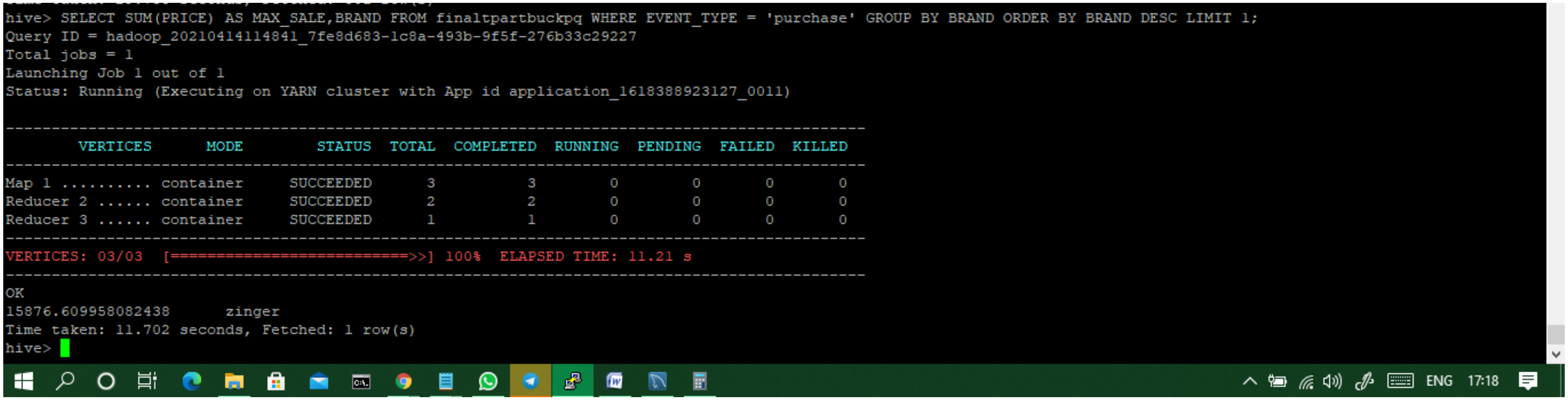


```
4009    2195085255176618020
25      2195085258339123402
2       category_id
Time taken: 29.738 seconds, Fetched: 501 row(s)
hive> 
```

6) Which brand had the maximum sales in October and November combined?

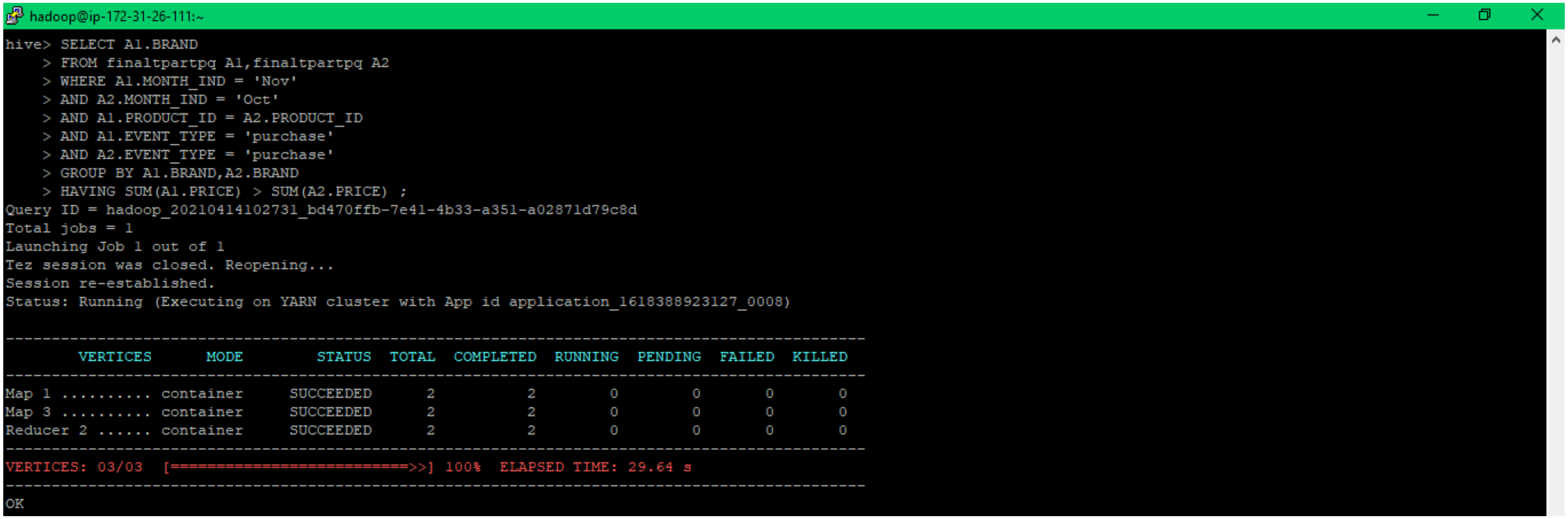


```
hive> SELECT SUM(PRICE) AS MAX_SALE,BRAND FROM finaltpartpq WHERE EVENT_TYPE = 'purchase' GROUP BY BRAND ORDER BY BRAND DESC LIMIT 1;
Query ID = hadoop_20210414101716_5551104b-23d3-40a1-b8a6-fe58867f6fa9
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1618388923127_0007)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 8.41 s
----------------------------------------------------------------------------------------------
OK
15876.609958082438      zinger
Time taken: 15.523 seconds, Fetched: 1 row(s)
hive>
```

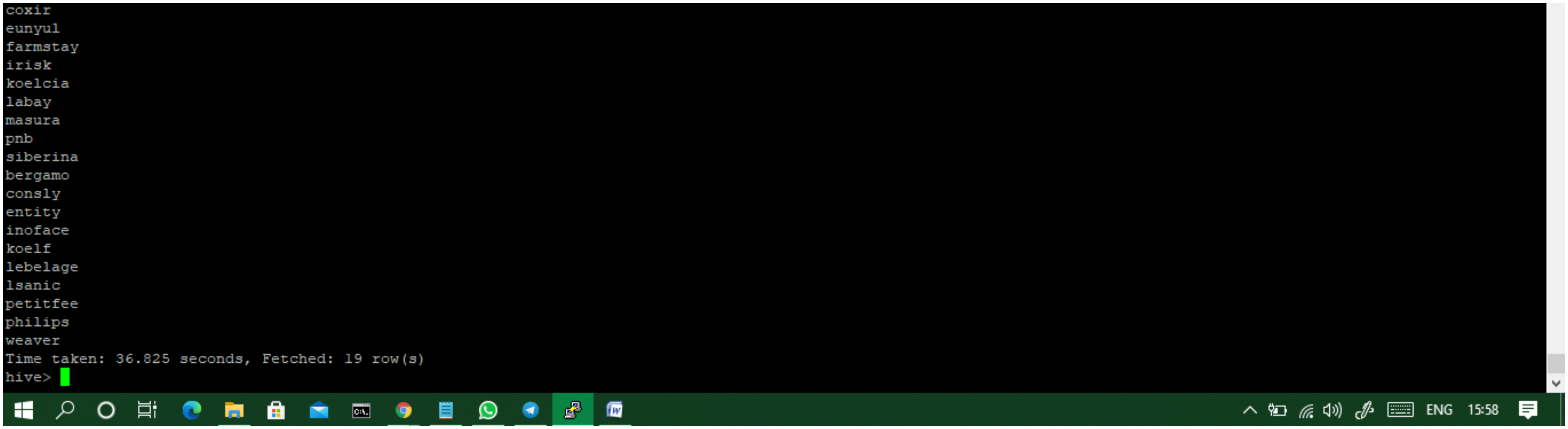


```
hive> SELECT SUM(PRICE) AS MAX_SALE,BRAND FROM finaltpartbuckpq WHERE EVENT_TYPE = 'purchase' GROUP BY BRAND ORDER BY BRAND DESC LIMIT 1;
Query ID = hadoop_20210414114841_7fe8d683-1c8a-493b-9f5f-276b33c29227
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 11.21 s
----------------------------------------------------------------------------------------------
OK
15876.609958082438      zinger
Time taken: 11.702 seconds, Fetched: 1 row(s)
hive>
```

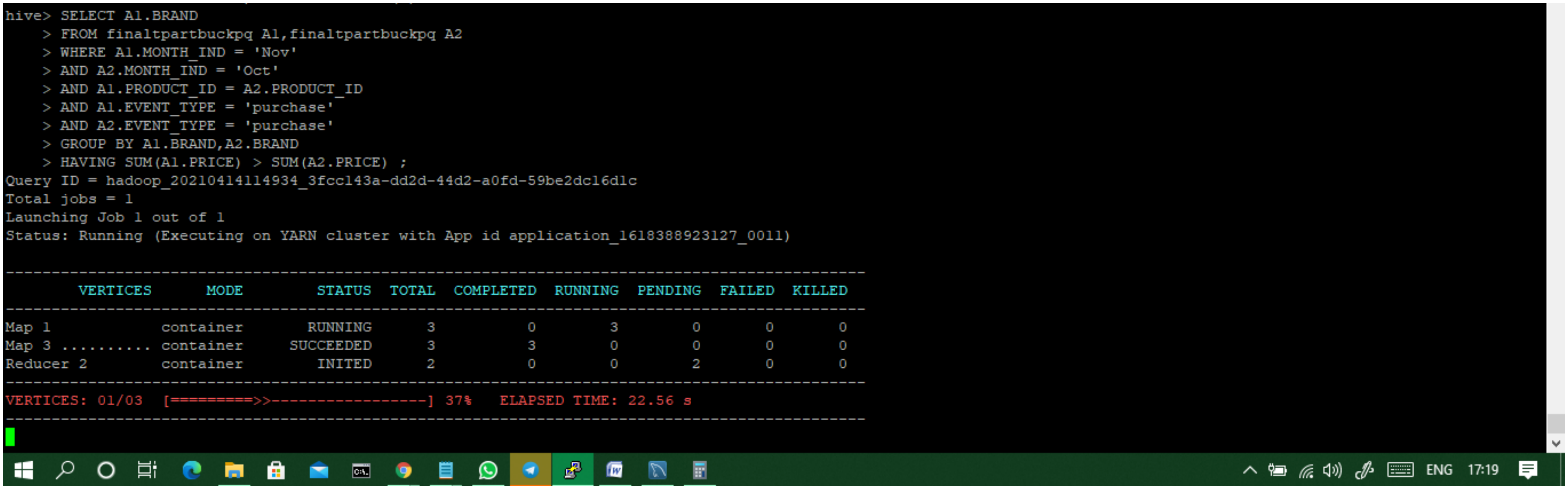
7) Which brands increased their sales from October to November?



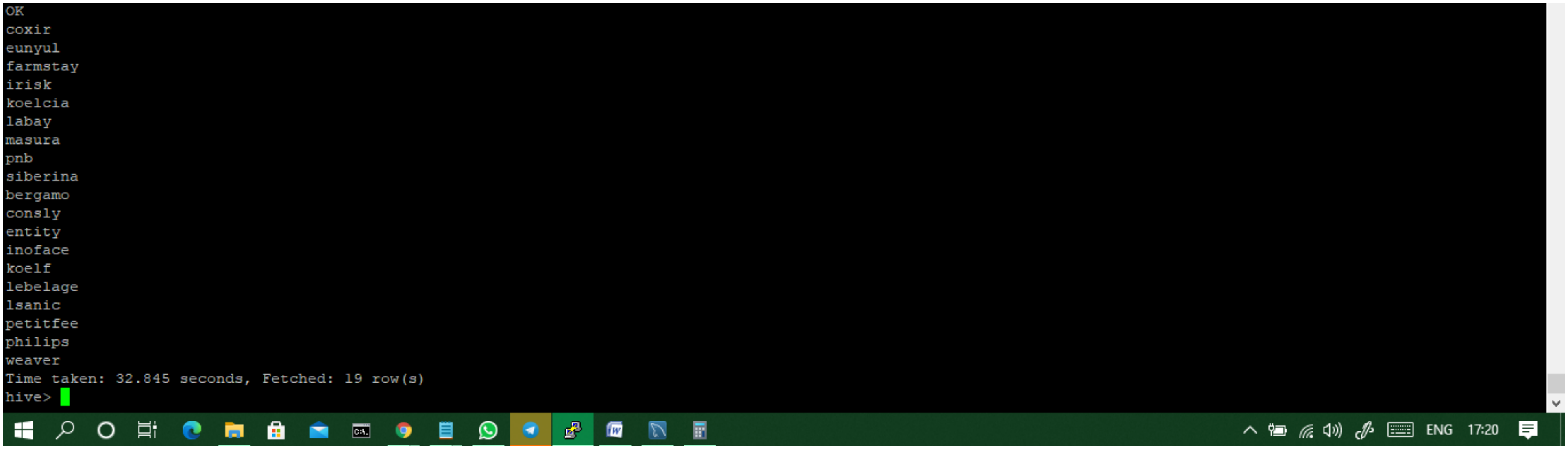
```
hive> SELECT A1.BRAND
    > FROM finaltpartpq A1,finaltpartpq A2
    > WHERE A1.MONTH_IND = 'Nov'
    > AND A2.MONTH_IND = 'Oct'
    > AND A1.PRODUCT_ID = A2.PRODUCT_ID
    > AND A1.EVENT_TYPE = 'purchase'
    > AND A2.EVENT_TYPE = 'purchase'
    > GROUP BY A1.BRAND,A2.BRAND
    > HAVING SUM(A1.PRICE) > SUM(A2.PRICE) ;
Query ID = hadoop_20210414102731_bd470ffb-7e41-4b33-a351-a02871d79c8d
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1618388923127_0008)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Map 3 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 29.64 s
----------------------------------------------------------------------------------------------
OK
```



```
coxir
eunyul
farmstay
irisk
koelcia
labay
masura
pnb
siberina
bergamo
consly
entity
inoface
koelf
lebelage
lsanic
petitfee
philips
weaver
Time taken: 36.825 seconds, Fetched: 19 row(s)
hive>
```

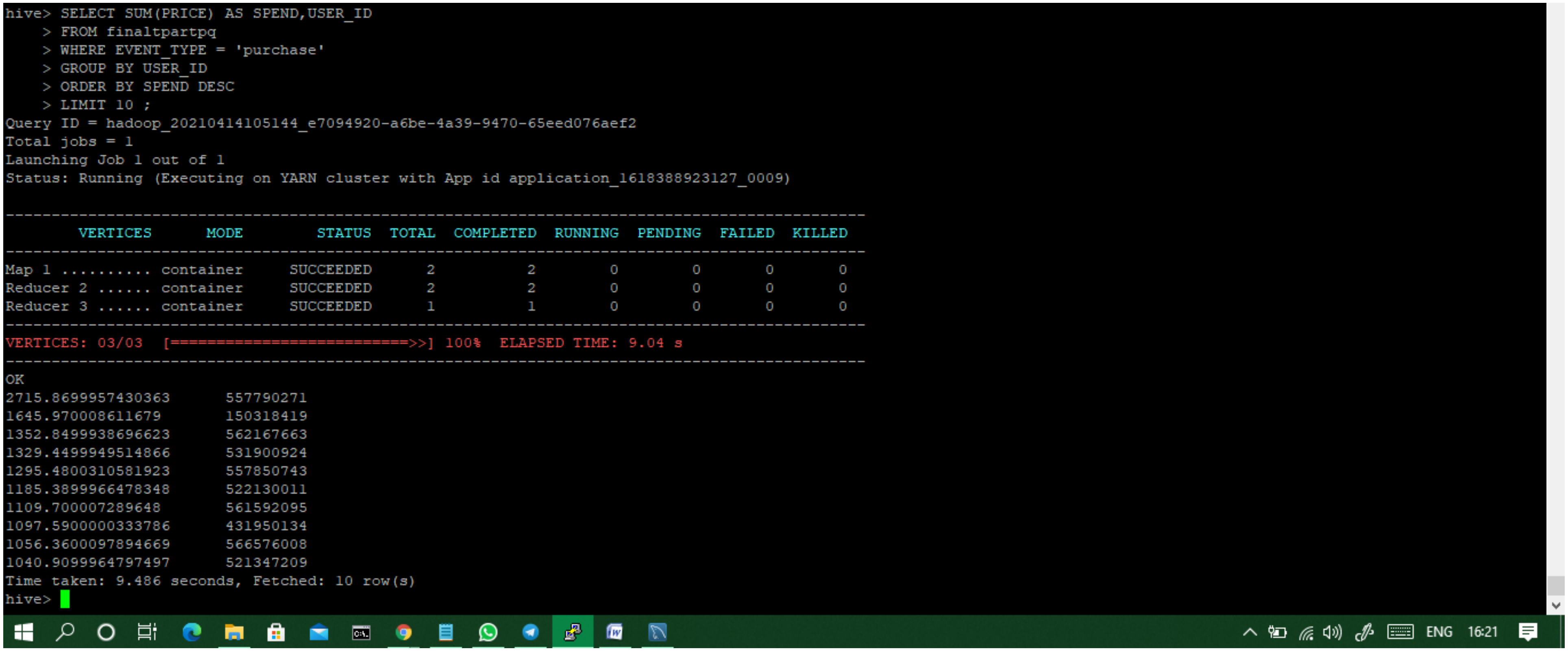


```
hive> SELECT A1.BRAND
    > FROM finaltpartbuckpq A1,finaltpartbuckpq A2
    > WHERE A1.MONTH_IND = 'Nov'
    > AND A2.MONTH_IND = 'Oct'
    > AND A1.PRODUCT_ID = A2.PRODUCT_ID
    > AND A1.EVENT_TYPE = 'purchase'
    > AND A2.EVENT_TYPE = 'purchase'
    > GROUP BY A1.BRAND,A2.BRAND
    > HAVING SUM(A1.PRICE) > SUM(A2.PRICE) ;
Query ID = hadoop_20210414114934_3fcc143a-dd2d-44d2-a0fd-59be2dc16d1c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1            container       RUNNING      3          0        3        0       0       0
Map 3 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2        container        INITED      2          0        0        2       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/03  [=========>>-----------------] 37%   ELAPSED TIME: 22.56 s
----------------------------------------------------------------------------------------------
```

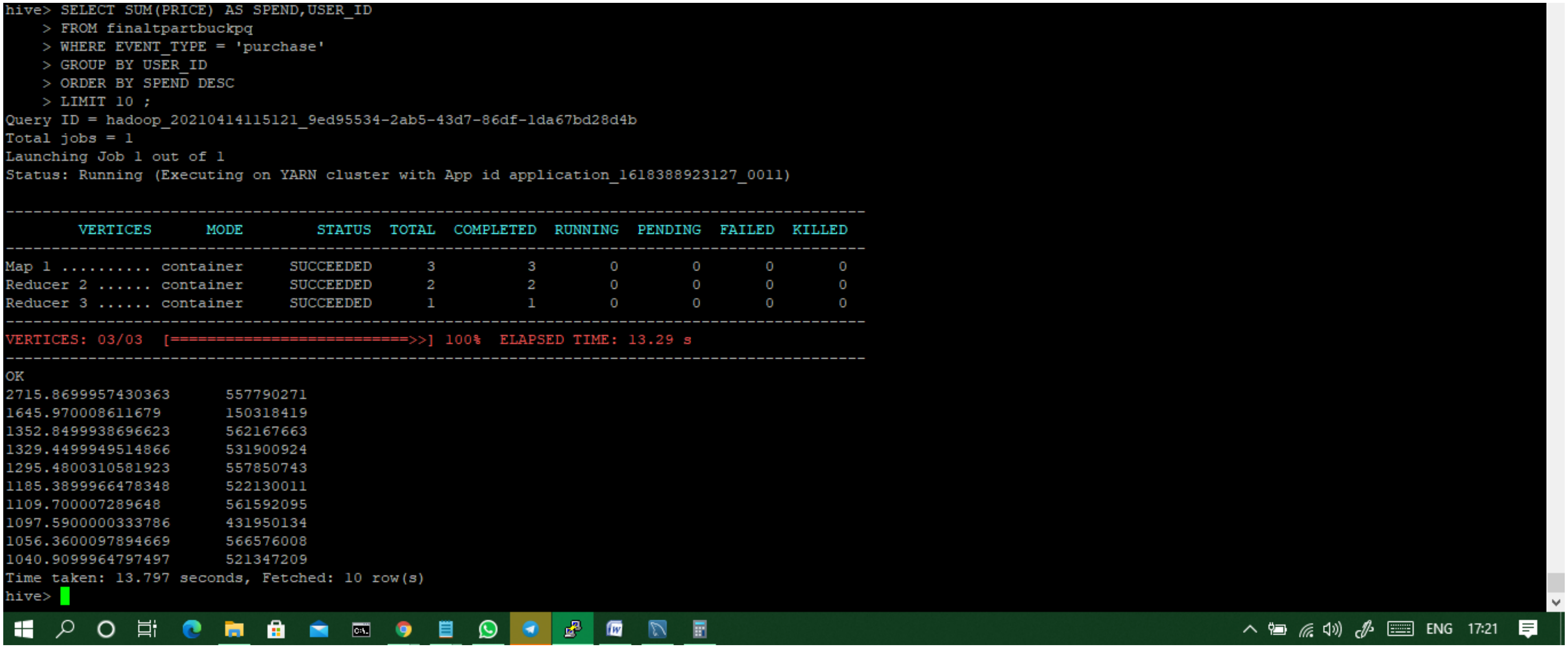


```
OK
coxir
eunyul
farmstay
irisk
koelcia
labay
masura
pnb
siberina
bergamo
consly
entity
inoface
koelf
lebelage
lsanic
petitfee
philips
weaver
Time taken: 32.845 seconds, Fetched: 19 row(s)
hive>
```

8) 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.



```
hive> SELECT SUM(PRICE) AS SPEND,USER_ID
    > FROM finaltpartpq
    > WHERE EVENT_TYPE = 'purchase'
    > GROUP BY USER_ID
    > ORDER BY SPEND DESC
    > LIMIT 10 ;
Query ID = hadoop_20210414105144_e7094920-a6be-4a39-9470-65eed076aef2
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0009)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 9.04 s
----------------------------------------------------------------------------------------------
OK
2715.8699957430363      557790271
1645.970008611679       150318419
1352.8499938696623      562167663
1329.4499949514866      531900924
1295.4800310581923      557850743
1185.3899966478348      522130011
1109.700007289648       561592095
1097.5900000333786      431950134
1056.3600097894669      566576008
1040.9099964797497      521347209
Time taken: 9.486 seconds, Fetched: 10 row(s)
hive>
```



```
hive> SELECT SUM(PRICE) AS SPEND,USER_ID
    > FROM finaltpartbuckpq
    > WHERE EVENT_TYPE = 'purchase'
    > GROUP BY USER_ID
    > ORDER BY SPEND DESC
    > LIMIT 10 ;
Query ID = hadoop_20210414115121_9ed95534-2ab5-43d7-86df-1da67bd28d4b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1618388923127_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 13.29 s
----------------------------------------------------------------------------------------------
OK
2715.8699957430363      557790271
1645.970008611679       150318419
1352.8499938696623      562167663
1329.4499949514866      531900924
1295.4800310581923      557850743
1185.3899966478348      522130011
1109.700007289648       561592095
1097.5900000333786      431950134
1056.3600097894669      566576008
1040.9099964797497      521347209
Time taken: 13.797 seconds, Fetched: 10 row(s)
hive>
```