

HIVE CASE STUDY

Aaron Alex

Problem Statement

With online sales gaining popularity, tech companies are exploring ways to improve their sales by analysing customer behaviour and gaining insights about product trends. Furthermore, the websites make it easier for customers to find the products they require without much scavenging. Needless to say, the role of big data analysts is among the most sought-after job profiles of this decade.

Therefore, as part of this assignment, we will be challenging you, as a big data analyst, to extract data and gather insights from a real-life data set of an e-commerce company.

Implementation

The implementation phase can be divided into the following parts:

EMR

1. Launch an EMR Cluster.
2. Move the data from the S3 Bucket to HDFS.

HIVE

1. Create Structure of Database.
2. Use optimized techniques to run your queries as efficiently as possible.
3. Show the improvement of the performance after using optimization on any single query.
4. Run Hive queries to answer the questions .

CLEANING UP

1. Drop the database.
2. Terminate the cluster.

EMR Stage

1. Launch an EMR Cluster

```
[hadoop@ip-172-31-15-48 ~]$ [REDACTED]
```

login as: hadoop

Authenticating with public key "case_study_key_pair"

_____|_(_|_- / Amazon Linux AMI
____|__|_||

<https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/>

68 package(s) needed for security, out of 107 available

Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE	MMMMMM	MMMMMM	RRRRRRRRRRRRRR
E:::::::::::E	M:::::M	M:::::M	R::::::::::R
EE:::::EEEEEEEEE:::E	M:::::M	M:::::M	R:::::RRRRRR:::R
E:::::E	EEEEEE	M:::::M	RR:::::R R:::::R
E:::::E		M:::::M:::M	R:::::R R:::::R
E:::::E		M:::::M M:::M	R:::::RRRRRR:::R
E:::::::::::E	M:::::M	M:::M:::M	R:::::::::::RR
E:::::::::::E	M:::::M	M:::::M	R:::::RRRRRR:::R
E:::::E		M:::::M	R:::::R R:::::R
E:::::E	EEEEEE	M:::::M	R:::::R R:::::R
EE:::::EEEEEEEEE:::E	M:::::M	MMM	M:::::M R:::::R
E:::::::::::E	M:::::M		M:::::M RR:::::R R:::::R
EEEEEEEEEEEEEEEEEE	MMMMMM	MMMMMM	RRRRRR

2. Move the data from the S3 Bucket to HDFS

2.1 Creating a new directory called ecom_data

```
hadoop fs -mkdir /user/root/tmp  
hadoop fs -mkdir /user/root/tmp/ecom_data
```

```
[hadoop@ip-172-31-15-48 ~]$ hadoop fs -mkdir /user/root/tmp  
[hadoop@ip-172-31-15-48 ~]$ hadoop fs -mkdir /user/root/tmp/ecom_data  
[hadoop@ip-172-31-15-48 ~]$ hadoop fs -ls /user/root/tmp  
Found 1 items  
drwxr-xr-x - hadoop hadoop 0 2021-10-04 07:26 /user/root/tmp/ecom_data  
[hadoop@ip-172-31-15-48 ~]$ █
```

2.2 Adding files to the target directory

```
hadoop distcp s3a://e-commerce-events-ml/2019-Oct.csv /user/root/tmp/ecom_data  
hadoop distcp s3a://e-commerce-events-ml/2019-Nov.csv /user/root/tmp/ecom_data
```

```
[hadoop@ip-172-31-15-48 ~]$ hadoop distcp s3a://e-commerce-events-ml/2019-Oct.csv /user/root/tmp/ecom_data  
21/10/04 07:28:18 INFO tools.DistCp: Input Options: DistCpOptions{atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false, skipCRC=false, blocking=true, numListstatusThreads=0, maxMaps=20, mapBandwidth=100, sslConfigurationFile='null', copyStrategy='uniformsize', preserveStatus=[], preserveRawXattrs=false, atomicWorkPath=null, logPath=null, sourceFileListing=null, sourcePaths=[s3a://e-commerce-events-ml/2019-Oct.csv], targetPath=/user/root/tmp/ecom_data, targetPathExists=true, filtersFile='null'}  
21/10/04 07:28:18 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-15-48.ec2.internal/172.31.15.48:8032  
21/10/04 07:28:20 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0  
21/10/04 07:28:20 INFO tools.SimpleCopyListing: Build file listing completed.  
21/10/04 07:28:20 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb  
21/10/04 07:28:20 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor  
21/10/04 07:28:21 INFO tools.DistCp: Number of paths in the copy list: 1  
21/10/04 07:28:21 INFO tools.DistCp: Number of paths in the copy list: 1  
21/10/04 07:28:21 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-15-48.ec2.internal/172.31.15.48:8032  
21/10/04 07:28:21 INFO mapreduce.JobSubmitter: number of splits:1  
21/10/04 07:28:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1633331362226_0004  
21/10/04 07:28:21 INFO impl.YarnClientImpl: Submitted application application_1633331362226_0004  
21/10/04 07:28:21 INFO mapreduce.Job: The url to track the job: http://ip-172-31-15-48.ec2.internal:20888/proxy/application_1633331362226_0004/  
21/10/04 07:28:21 INFO tools.DistCp: DistCp job-id: job_1633331362226_0004  
21/10/04 07:28:21 INFO mapreduce.Job: Running job: job_1633331362226_0004  
21/10/04 07:28:29 INFO mapreduce.Job: Job job_1633331362226_0004 running in uber mode : false  
21/10/04 07:28:29 INFO mapreduce.Job: map 0% reduce 0%  
21/10/04 07:28:47 INFO mapreduce.Job: map 100% reduce 0%  
21/10/04 07:28:48 INFO mapreduce.Job: Job job_1633331362226_0004 completed successfully
```

```
[hadoop@ip-172-31-15-48 ~]$ hadoop distcp s3a://e-commerce-events-ml/2019-Nov.csv /user/root/tmp/ecom_data  
21/10/04 07:30:11 INFO tools.DistCp: Input Options: DistCpOptions{atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false, skipCRC=false, blocking=true, numListstatusThreads=0, maxMaps=20, mapBandwidth=100, sslConfigurationFile='null', copyStrategy='uniformsize', preserveStatus=[], preserveRawXattrs=false, atomicWorkPath=null, logPath=null, sourceFileListing=null, sourcePaths=[s3a://e-commerce-events-ml/2019-Nov.csv], targetPath=/user/root/tmp/ecom_data, targetPathExists=true, filtersFile='null'}  
21/10/04 07:30:11 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-15-48.ec2.internal/172.31.15.48:8032  
21/10/04 07:30:14 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0  
21/10/04 07:30:14 INFO tools.SimpleCopyListing: Build file listing completed.  
21/10/04 07:30:14 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb  
21/10/04 07:30:14 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor  
21/10/04 07:30:14 INFO tools.DistCp: Number of paths in the copy list: 1  
21/10/04 07:30:14 INFO tools.DistCp: Number of paths in the copy list: 1  
21/10/04 07:30:14 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-15-48.ec2.internal/172.31.15.48:8032  
21/10/04 07:30:14 INFO mapreduce.JobSubmitter: number of splits:1  
21/10/04 07:30:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1633331362226_0005  
21/10/04 07:30:15 INFO impl.YarnClientImpl: Submitted application application_1633331362226_0005  
21/10/04 07:30:15 INFO mapreduce.Job: The url to track the job: http://ip-172-31-15-48.ec2.internal:20888/proxy/application_1633331362226_0005/  
21/10/04 07:30:15 INFO tools.DistCp: DistCp job-id: job_1633331362226_0005  
21/10/04 07:30:15 INFO mapreduce.Job: Running job: job_1633331362226_0005  
21/10/04 07:30:23 INFO mapreduce.Job: Job job_1633331362226_0005 running in uber mode : false  
21/10/04 07:30:23 INFO mapreduce.Job: map 0% reduce 0%  
21/10/04 07:30:42 INFO mapreduce.Job: map 100% reduce 0%  
21/10/04 07:30:44 INFO mapreduce.Job: Job job_1633331362226_0005 completed successfully
```

2.3 Verification

```
hadoop fs -ls /user/root/tmp/ecom_data
```

```
[hadoop@ip-172-31-15-48 ~]$ hadoop fs -ls /user/root/tmp/ecom_data
Found 2 items
-rw-r--r-- 1 hadoop hadoop 545839412 2021-10-04 07:30 /user/root/tmp/ecom_data/2019-Nov.csv
-rw-r--r-- 1 hadoop hadoop 482542278 2021-10-04 07:28 /user/root/tmp/ecom_data/2019-Oct.csv
```

HIVE Stage

1. Launch HIVE

hive

2. Create Database Structure

2.1 Create database

```
create database case_study;
```

```
hive> create database case_study;
OK
Time taken: 0.826 seconds
```

2.2 Create table - retail_clickstream

```
create table if not exists retail_clickstream(
    event_time timestamp,
    event_type string,
    product_id string,
    category_id string,
    category_code string,
    brand string,
    price decimal (10,3),
    user_id bigint,
    user_session string
)
row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
with serdeproperties (
    "separatorChar"=",",
    "quoteChar"="\",
    "escapeChar"="\\"
)
stored as textfile
location '/user/root/tmp/ecom_data'
tblproperties ("skip.header.line.count"="1");
```

```
hive> create table if not exists retail_clickstream(
  >     event_time timestamp,
  >     event_type string,
  >     product_id string,
  >     category_id string,
  >     category_code string,
  >     brand string,
  >     price decimal (10,3),
  >     user_id bigint,
  >     user_session string
  > )
  > row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
  > with serdeproperties (
  >     "separatorChar"=",",
  >     "quoteChar"="\",
  >     "escapeChar"="\\"
  > )
  > stored as textfile
  > location '/user/root/tmp/ecom_data'
  > tblproperties ("skip.header.line.count"="1");
```

OK

Time taken: 0.425 seconds

```
hive> █
```

2.3 Verifying retail_clickstream table schema

```
describe retail_clickstream;
```

```
hive> describe retail_clickstream;
OK
event_time          string
event_type          string
product_id          string
category_id         string
category_code       string
brand               string
price               string
user_id              string
user_session         string
Time taken: 0.196 seconds, Fetched: 9 row(s)
hive> █
```

2.4 Creating table with target datatypes - rc

```
hive> create table if not exists rc(
  >   event_date date,
  >   event_type string,
  >   product_id string,
  >   category_id string,
  >   category_code string,
  >   brand string,
  >   price decimal (10,3),
  >   user_id bigint,
  >   user_session string
  > )
  > row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile;
OK
Time taken: 0.498 seconds
```

2.5 Verifying rc table schema

```
hive> describe rc;
OK
event_date          date
event_type          string
product_id          string
category_id         string
category_code       string
brand               string
price               decimal(10,3)
user_id              bigint
user_session        string
Time taken: 0.044 seconds, Fetched: 9 row(s)
hive> █
```

2.6 Loading data into rc

```
hive> > insert overwrite table rc
> select to_date(event_time) as event_date,
> event_type string,
> product_id string,
> category_id string,
> category_code string,
> brand string,
> cast(price as decimal(10,3)) as price,
> cast(user_id as bigint) as user_id,
> user_session string
>
> from retail_clickstream;
Query ID = hadoop_20211004153500_9e682904-cdc5-4301-a7b6-8e7f7bb3af9f
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0010)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	2	2	0	0	0	0

VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 95.08 s

Loading data to table default.rc
OK
Time taken: 98.994 seconds

2.7 Commands to allow dynamic partitioning

```
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode = nonstrict;
hive> █
```

2.8 To create partitioned table - prc

```
hive> create table if not exists prc(
>     event_type string,
>     product_id string,
>     category_id string,
>     category_code string,
>     brand string,
>     price decimal (10,3),
>     user_id bigint,
>     user_session string
> )
> partitioned by (event_date date)
> clustered by (user_id) into 20 buckets
> row format delimited fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
OK
Time taken: 0.088 seconds
hive> █
```

2.9 Verfying prc table schema

```
hive> describe prc;
OK
event_type          string
product_id          string
category_id         string
category_code       string
brand               string
price               decimal(10,3)
user_id              bigint
user_session        string
event_date          date

# Partition Information
# col_name           data_type      comment
event_date          date
Time taken: 0.077 seconds, Fetched: 14 row(s)
hive> █
```

2.10 Loading data into prc

```
hive> insert into table prc
    >   partition(event_date)
    >     select event_type, product_id, category_id, category_code, brand, price, user_id, user_session, event_date from rc;
Query ID = hadoop_20211004162656_260336df-617c-4621-9d7c-3bc1fc16ba9e
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0013)

-----

| VERTICES        | MODE      | STATUS    | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|-----------------|-----------|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1 .....     | container | SUCCEEDED | 7     | 7         | 0       | 0       | 0      | 0      |
| Reducer 2 ..... | container | SUCCEEDED | 4     | 4         | 0       | 0       | 0      | 0      |


-----  
VERTICES: 02/02  [=====>>] 100%  ELAPSED TIME: 96.04 s  
-----  
Loading data to table default.prc partition (event_date=null)  
  
Loaded : 61/61 partitions.  
          Time taken to load dynamic partitions: 3.951 seconds  
          Time taken for adding to write entity : 0.03 seconds  
OK  
Time taken: 102.178 seconds  
hive> █
```

3. Questions

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

Ans: 1211538.430

Time taken to run the query on the partitioned table was 22.897 seconds

```
hive> select sum(price) as revenue
  > from prc
  > where event_type = 'purchase' and (event_date >= '2019-10-01' and event_date <= '2019-10-31');
Query ID = hadoop_20211004165520_44d6f148-a4b4-4ecf-adb6-2395d2a538ea
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0014)

-----  

      VERTICES    MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED  

-----  

Map 1 ..... container  SUCCEEDED      6          6          0          0          0          0
Reducer 2 ..... container  SUCCEEDED      1          1          0          0          0          0
-----  

VERTICES: 02/02  [=====>>] 100%  ELAPSED TIME: 22.26 s
-----  

OK
1211538.430
Time taken: 22.897 seconds, Fetched: 1 row(s)
hive> █
```

Time taken to run the query on the non-partitioned table was **27.681 seconds**

```
hive> select sum(price) as revenue
  > from rc
  > where event_type = 'purchase' and (event_date >= '2019-10-01' and event_date <= '2019-10-31');
Query ID = hadoop_20211004165754_20216ba6-b865-434d-b809-378bd1eda59d
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0014)

-----

| VERTICES        | MODE      | STATUS    | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|-----------------|-----------|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1 .....     | container | SUCCEEDED | 7     | 7         | 0       | 0       | 0      | 0      |
| Reducer 2 ..... | container | SUCCEEDED | 1     | 1         | 0       | 0       | 0      | 0      |


-----  
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 27.09 s  
-----  
OK  
1211538.430  
Time taken: 27.681 seconds, Fetched: 1 row(s)  
hive> █
```

The query took approximately **20% longer** to be executed on the non-partitioned table as compared to the partitioned table. This shows the improvement in performance after using appropriate optimizations on the table.

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

Ans:

10 (October) - 245624

11 (November) - 322417

```
hive> select month(event_date) as month, count(event_type) as count
  >   from prc
  >   where event_type = 'purchase'
  >   group by month(event_date);
Query ID = hadoop_20211004170444_2fa9f994-7512-44c7-8276-9ca090a270e8
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0015)

-----  

      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: 27.54 s  

-----  

OK
10      245624
11      322417
Time taken: 28.412 seconds, Fetched: 2 row(s)
hive> █
```

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

Ans: 319478.470

```
hive> select revenue - lead(revenue, 1,0) over (order by revenue DESC) as difference
  >   from (
  >     select month(event_date) as month, sum(price) as revenue
  >     from prc
  >     where event_type = 'purchase'
  >     group by month(event_date)
  >     order by revenue DESC
  >   )a
  > limit 1;
```

Query ID = hadoop_20211004171822_2f1ddbf0-5833-4d66-b3f7-95a8e3df3902

Total jobs = 1

Launching Job 1 out of 1

Status: Running (Executing on YARN cluster with App id application_1633356599843_0016)

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	6	6	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: 28.36 s

OK

319478.470

Time taken: 29.481 seconds, Fetched: 1 row(s)

```
hive> █
```

3.4 Find distinct categories of products. Categories with null category code can be ignored.

```
hive> select distinct(category_code)
  >   from prc;
Query ID = hadoop_20211004173240_49d35774-8edd-43f0-b11f-6020f3b5151b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0016)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	6	6	0	0	0	0
Reducer 2	container	SUCCEEDED	4	4	0	0	0	0

```
VERTICES: 02/02  [=====>>] 100%  ELAPSED TIME: 27.15 s
```

```
OK
```

```
accessories.bag
appliances.environment.vacuum
appliances.personal.hair_cutter
sport.diving
```

```
apparel.glove
furniture.bathroom.bath
furniture.living_room.cabinet
stationery.cartrige
accessories.cosmetic_bag
appliances.environment.air_conditioner
furniture.living_room.chair
Time taken: 27.718 seconds, Fetched: 12 row(s)
```

```
hive> █
```

3.5 Find the total number of products available under each category.

```
hive> select category_code, count(product_id) as prod_count
>   from prc
>   group by category_code
>   order by prod_count DESC;
Query ID = hadoop_20211004173752_7b297b83-2a80-4265-b262-82838be4c314
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0016)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	6	6	0	0	0	0
Reducer 2	container	SUCCEEDED	4	4	0	0	0	0
Reducer 3	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 03/03  [=====>>] 100% ELAPSED TIME: 27.76 s
```

```
OK
```

```
8594895
appliances.environment.vacuum      59761
stationery.cartrige                26722
apparel.glove                      18232
furniture.living_room.cabinet     13439
accessories.bag                     11681
furniture.bathroom.bath            9857
appliances.personal.hair_cutter    1643
accessories.cosmetic_bag           1248
appliances.environment.air_conditioner 332
furniture.living_room.chair        308
sport.diving                       2
Time taken: 28.316 seconds, Fetched: 12 row(s)
hive>
```

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

Ans: runail

```
hive> select brand, count(event_type) as sales
  >   from prc
  >   where event_type ='purchase'
  >   group by brand
  >   order by sales DESC
  >   limit 5;
Query ID = hadoop_20211004175043_27870868-e873-4763-8d95-f9e30b246675
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_1633356599843_0017)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	6	6	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: 27.42 s

```
OK
    238636
runail 47935
irisk 32013
masura 22837
bpw.style 22492
Time taken: 35.714 seconds, Fetched: 5 row(s)
hive> █
```

3.7 Which brands increased their sales from October to November?

Ans:

3.8 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.

```
hive> select user_id, sum(price) as expenditure
  >   from prc
  >   where event_type = 'purchase'
  >   group by user_id
  >   order by expenditure DESC
  >   limit 10;
Query ID = hadoop_20211004175444_e336c6bd-da4d-4454-85a6-c750659214f9
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1633356599843_0017)

-----

| VERTICES        | MODE      | STATUS    | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|-----------------|-----------|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1 .....     | container | SUCCEEDED | 6     | 6         | 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: 31.17 s  
-----  
OK  
557790271      2715.870  
150318419      1645.970  
562167663      1352.850  
531900924      1329.450  
557850743      1295.480  
522130011      1185.390  
561592095      1109.700  
431950134      1097.590  
566576008      1056.360  
521347209      1040.910  
Time taken: 31.834 seconds, Fetched: 10 row(s)  
hive>
```