

Hive Case Study

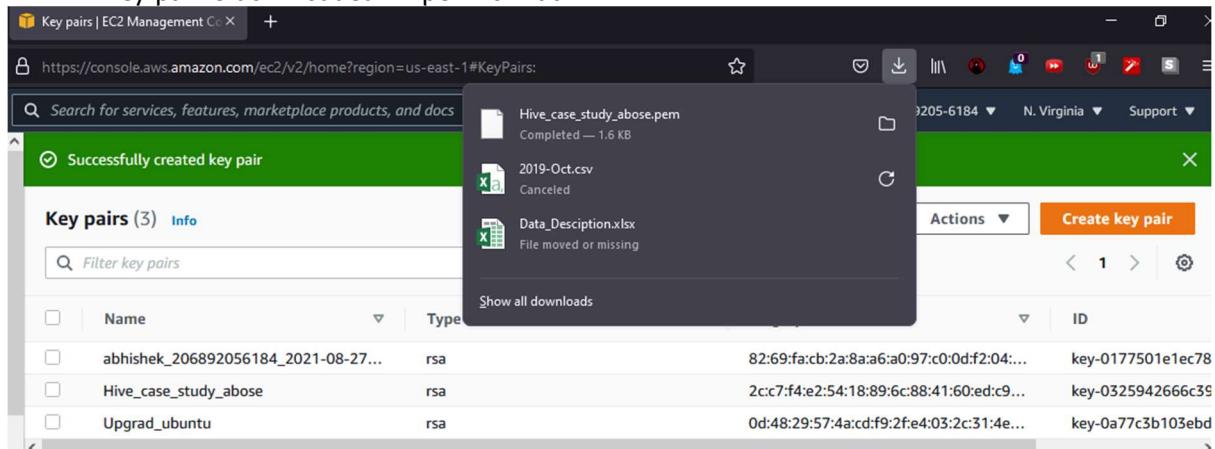
Querying using Hadoop & Hive

Procedure for creation of Key- Pair & EMR Cluster & saving the files on s3

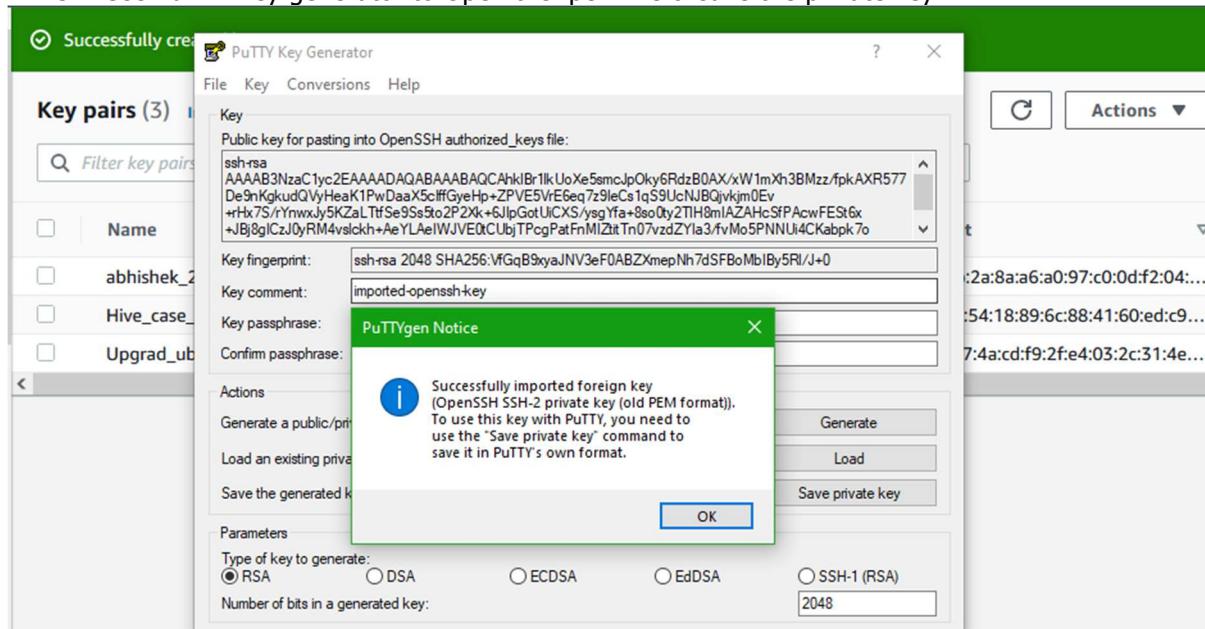
1. Create a key pair from EC2 Dash board >> Key pairs >> Create key pair

The screenshot shows the 'Create key pair' wizard in the AWS Management Console. The 'Name' field is set to 'Hive_case_study_abose'. The 'Key pair type' is selected as 'RSA'. The 'Private key file format' is chosen as '.pem'. Under 'Tags (Optional)', there is a button labeled 'Add tag'. A note at the bottom states: 'The name can include up to 255 ASCII characters. It can't include leading or trailing spaces.'

2. Key pair is downloaded in .pem format



3. Use PuTTY key generator to Open the .pem file & save the private key



| | | | |
|---------------------------|---------------------|-----------------------|--------|
| Data_Description_upgrad | 25-09-2021 12:36 PM | Microsoft Excel W... | 12 KB |
| Steps screenshot | 25-09-2021 12:51 PM | Microsoft Word D... | 212 KB |
| Hive_case_study_key | 25-09-2021 12:50 PM | PutTY Private Key ... | 2 KB |
| Hive_case_study_abose.pem | 25-09-2021 12:44 PM | PEM File | 2 KB |

4. Create the EMR cluster by AWS services >> EMR under Analytics section

Create Cluster - Quick Options [Go to advanced options](#)

General Configuration

Cluster name:

Logging [i](#)

S3 folder: [File](#)

Launch mode: Cluster [i](#) Step execution [i](#)

Software configuration

Release: [i](#)

Applications:

- Core Hadoop: Hadoop 2.8.5 with Ganglia 3.7.2, Hive 2.3.6, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.2
- HBase: HBase 1.4.10 with Ganglia 3.7.2, Hadoop 2.8.5, Hive 2.3.6, Hue 4.4.0, Phoenix 4.14.3, and ZooKeeper 3.4.14
- Presto: Presto 0.227 with Hadoop 2.8.5 HDFS and Hive 2.3.6 Metastore
- Spark: Spark 2.4.4 on Hadoop 2.8.5 YARN with Ganglia 3.7.2 and Zeppelin 0.8.2

5. Select EC2 key pair as the key created earlier
Hardware configuration

Instance type m4.large The selected instance type adds 32 GiB of GP2 EBS storage per instance by default. [Learn more](#)

Number of instances 2 (1 master and 1 core nodes)

Security and access

EC2 key pair Hive_case_study_abose [Learn how to create an EC2 key pair.](#)

Permissions Default Custom Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role EMR_DefaultRole Use EMR_DefaultRole_V2

EC2 instance profile EMR_EC2_DefaultRole [Info](#)

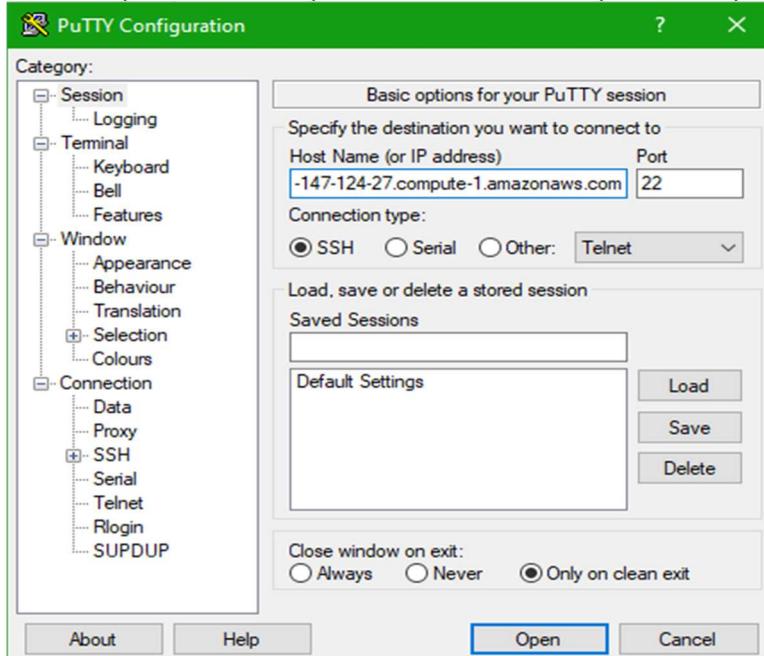
[Cancel](#) [Create cluster](#)

6. Save the data downloaded from the Upgrad Case study module page to s3
hive-case-study-s3 [Info](#)

| Objects | Properties | Permissions | Metrics | Management | Access Points |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------|-------------------------------------|---------------------------------------|
| Objects (2) | | | | | |
| Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more | | | | | |
| <input type="button" value="C"/> | <input type="button" value="Copy S3 URI"/> | <input type="button" value="Copy URL"/> | <input type="button" value="Download"/> | <input type="button" value="Open"/> | <input type="button" value="Delete"/> |
| <input type="button" value="Actions"/> | <input type="button" value="Create folder"/> | <input type="button" value="Upload"/> | | | |
| <input type="text"/> Find objects by prefix | | | | | |
| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
| <input type="checkbox"/> |  2019-Nov.csv | csv | September 25, 2021, 14:38:18 (UTC+05:30) | 520.6 MB | Standard |
| <input type="checkbox"/> |  2019-Oct.csv | csv | September 25, 2021, 14:16:16 (UTC+05:30) | 460.2 MB | Standard |

Querying the data using Hadoop & HQL

1. Open PuTTY and paste the Host name as per Master public DNS of the EMR cluster



2. Select the private key saved in SSH >> Auth. Click on Open login using hadoop

3. Create a directory to store the data files for querying & analysis

```
[hadoop@ip-172-31-11-246:~]
```

```
[hadoop@ip-172-31-11-246 ~]$ hadoop fs -mkdir /user/HiveCaseStudy/
```

4. Enter command to load each data file from s3 to HDFS storage

```
[hadoop@ip-172-31-11-246:~]
```

```
[hadoop@ip-172-31-11-246 ~]$ hadoop distcp s3://hive-case-study-s3/2019-Oct.csv /user/HiveCaseStudy/Oct.csv
21/09/25 09:56:08 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=[s3://hive-case-study-s3/2019-Oct.csv], targetPath=/user/HiveCaseStudy/Oct.csv, targetPathExists=false, filtersFile='null'}
21/09/25 09:56:08 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-11-246.ec2.internal/172.31.11.246:8032
21/09/25 09:56:12 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0
21/09/25 09:56:12 INFO tools.SimpleCopyListing: Build file listing completed.
21/09/25 09:56:12 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.ask.io.sort.mb
21/09/25 09:56:12 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
21/09/25 09:56:12 INFO tools.DistCp: Number of paths in the copy list: 1
21/09/25 09:56:12 INFO tools.DistCp: Number of paths in the copy list: 1
21/09/25 09:56:13 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-11-246.ec2.internal/172.31.11.246:8032
21/09/25 09:56:13 INFO mapreduce.JobSubmitter: number of splits:1
21/09/25 09:56:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1632563517868_0001
21/09/25 09:56:14 INFO impl.YarnClientImpl: Submitted application application_1632563517868_0001
21/09/25 09:56:14 INFO mapreduce.Job: The url to track the job: http://ip-172-31-11-246.ec2.internal:20888/proxy/application_1632563517868_0001/
21/09/25 09:56:14 INFO tools.DistCp: DistCp job-id: job_1632563517868_0001
21/09/25 09:56:14 INFO mapreduce.Job: Running job: job_1632563517868_0001
21/09/25 09:56:23 INFO mapreduce.Job: Job job_1632563517868_0001 running in uber mode : false
21/09/25 09:56:23 INFO mapreduce.Job: map 0% reduce 0%
21/09/25 09:56:41 INFO mapreduce.Job: map 100% reduce 0%
21/09/25 09:56:42 INFO mapreduce.Job: Job job_1632563517868_0001 completed successfully
21/09/25 09:56:43 INFO mapreduce.Job: Counters: 38
      File System Counters
          FILE: Number of bytes read=0
          FILE: Number of bytes written=172487
          FILE: Number of read operations=0
          FILE: Number of large read operations=0
          FILE: Number of write operations=0
          HDFS: Number of bytes read=357
          HDFS: Number of bytes written=482542278
          HDFS: Number of read operations=12
          HDFS: Number of large read operations=0
          HDFS: Number of write operations=4
```

```

21/09/25 09:56:14 INFO mapreduce.Job: Running job: job_1632563517868_0001
21/09/25 09:56:23 INFO mapreduce.Job: Job job_1632563517868_0001 running in uber mode : false
21/09/25 09:56:23 INFO mapreduce.Job: map 0% reduce 0%
21/09/25 09:56:41 INFO mapreduce.Job: map 100% reduce 0%
21/09/25 09:56:42 INFO mapreduce.Job: Job job_1632563517868_0001 completed successfully
21/09/25 09:56:43 INFO mapreduce.Job: Counters: 38
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=172487
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=357
        HDFS: Number of bytes written=482542278
        HDFS: Number of read operations=12
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
        S3: Number of bytes read=482542278
        S3: Number of bytes written=0
        S3: Number of read operations=0
        S3: Number of large read operations=0
        S3: Number of write operations=0
    Job Counters
        Launched map tasks=1
        Other local map tasks=1
        Total time spent by all maps in occupied slots (ms)=534784
        Total time spent by all reduces in occupied slots (ms)=0
        Total time spent by all map tasks (ms)=16712
        Total vcore-milliseconds taken by all map tasks=16712
        Total megabyte-milliseconds taken by all map tasks=17113088
    Map-Reduce Framework
        Map input records=1
        Map output records=0
        Input split bytes=136
        Spilled Records=0
        Failed Shuffles=0
        Merged Map outputs=0
        GC time elapsed (ms)=314
        CPU time spent (ms)=19040
        Physical memory (bytes) snapshot=596316160
        Virtual memory (bytes) snapshot=3299373056
        Total committed heap usage (bytes)=511180800

```

```

File Input Format Counters
    Bytes Read=221
File Output Format Counters
    Bytes Written=0
DistCp Counters
    Bytes Copied=482542278
    Bytes Expected=482542278
    Files Copied=1
[hadoop@ip-172-31-11-246 ~]$ 

```

Command to load the 2nd file to HDFS from s3

```

[hadoop@ip-172-31-11-246 ~]$ hadoop distcp s3://hive-case-study-s3/2019-Nov.csv /user/HiveCaseStudy/Nov.csv
21/09/25 10:00:19 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=[s3://hive-case-study-s3/2019-Nov.csv], targetPath=/user/HiveCaseStudy/Nov.csv, targetPathExists=false, filtersFile='null'}
21/09/25 10:00:19 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-11-246.ec2.internal/172.31.1.246:8032
21/09/25 10:00:24 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0
21/09/25 10:00:24 INFO tools.SimpleCopyListing: Build file listing completed.
21/09/25 10:00:24 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
21/09/25 10:00:24 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
21/09/25 10:00:24 INFO tools.DistCp: Number of paths in the copy list: 1
21/09/25 10:00:24 INFO tools.DistCp: Number of paths in the copy list: 1
21/09/25 10:00:24 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-11-246.ec2.internal/172.31.1.246:8032
21/09/25 10:00:24 INFO mapreduce.JobSubmitter: number of splits:1
21/09/25 10:00:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1632563517868_0002
21/09/25 10:00:25 INFO impl.YarnClientImpl: Submitted application application_1632563517868_0002
21/09/25 10:00:25 INFO mapreduce.Job: The url to track the job: http://ip-172-31-11-246.ec2.internal:20888/proxy/application_1632563517868_0002/

```

```

hadoop@ip-172-31-11-246:~$ 
          Total vcore-milliseconds taken by all map tasks=18340
          Total megabyte-milliseconds taken by all map tasks=18780160
      Map-Reduce Framework
          Map input records=1
          Map output records=0
          Input split bytes=136
          Spilled Records=0
          Failed Shuffles=0
          Merged Map outputs=0
          GC time elapsed (ms)=307
          CPU time spent (ms)=20730
          Physical memory (bytes) snapshot=543346688
          Virtual memory (bytes) snapshot=3293978624
          Total committed heap usage (bytes)=444071936
      File Input Format Counters
          Bytes Read=221
      File Output Format Counters
          Bytes Written=0
      DistCp Counters
          Bytes Copied=545839412
          Bytes Expected=545839412
          Files Copied=1
[hadoop@ip-172-31-11-246 ~]$ 

```

5. Checking the data saved in directory under HiveCaseStudy & starting Hive

```

[hadoop@ip-172-31-11-246 ~]$ hadoop fs -ls /user/HiveCaseStudy/
Found 2 items
-rw-r--r--  1 hadoop hadoop  545839412 2021-09-25 10:00 /user/HiveCaseStudy/Nov.csv
-rw-r--r--  1 hadoop hadoop  482542278 2021-09-25 09:56 /user/HiveCaseStudy/Oct.csv
[hadoop@ip-172-31-11-246 ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> 

```

6. Creating an external table as per the schema provided in the Data description provided in case study

```

hive> create external table if not exists Shopping(event_time timestamp, event_type string, product_id string, category_code string, brand string, price float, user_id bigint, user_session string) row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde' stored as textfile location '/user/HiveCaseStudy/' tblproperties("skip.header.line.count"="1");
OK
Time taken: 1.04 seconds
hive> 

```

7. Enabling the heading in the output & verifying the schema is defined correctly

```

hive> set hive.cli.print.header=True;
hive> select * from Shopping limit 5;
OK
shopping.event_time      shopping.event_type      shopping.product_id      shopping
.category_id      shopping.category_code      shopping.brand      shopping.price      shopping
.user_id      shopping.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-alcc-af0575a34ffb
2019-11-01 00:00:10 UTC view      5837166 1783999064103190764      pnb      2
2.22      556138645      57ed222e-a54a-4907-9944-5a875c2d7f4f
2019-11-01 00:00:11 UTC cart      5876812 1487580010100293687      jessnail
3.16      564506666      186c1951-8052-4b37-adce-dd9644bld5f7
2019-11-01 00:00:24 UTC remove_from_cart      5826182 1487580007483048900      3
.33      553329724      2067216c-31b5-455d-alcc-af0575a34ffb
Time taken: 0.293 seconds, Fetched: 5 row(s)
hive> 

```

Questions & command screenshots

Q1 Find the total revenue generated due to purchases made in October.

```
hive> select sum(price) as Oct_total_revenue from Shopping where date_format(event_time, 'MM')=10 and event_type='purchase';
Query ID = hadoop_20210925104224_2b760b7a-e827-4575-834e-c26bladdaa3e
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_1632563517868_0007)

Map 1: -- Reducer 2: 0/1
Map 1: 0/2 Reducer 2: 0/1
Map 1: 0/2 Reducer 2: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/1
```

```
Map 1: 0(+2)/2 Reducer 2: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/1
Map 1: 2/2 Reducer 2: 0(+1)/1
Map 1: 2/2 Reducer 2: 1/1
OK
oct_total_revenue
1211538.4299997438
Time taken: 124.736 seconds, Fetched: 1 row(s)
hive>
```

Answer 1= 1,211,538

Q2 Write a query to yield the total sum of purchases per month in a single output.

```
hive> select date_format(event_time, 'MM') as months, count(event_type) as Sum_of_purchases from Shopping where event_type='purchase' group by date_format(event_time, 'MM');
Query ID = hadoop_20210925104827_af0f69da-6fac-4b5c-8bd3-a8c940b42aaa
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1632563517868_0007)

Map 1: 0/2 Reducer 2: 0/3
Map 1: 0/2 Reducer 2: 0/3
Map 1: 0(+2)/2 Reducer 2: 0/3
```

```
Map 1: 0(+2)/2 Reducer 2: 0/3
Map 1: 1(+1)/2 Reducer 2: 0(+1)/3
Map 1: 1(+1)/2 Reducer 2: 0(+1)/3
Map 1: 1(+1)/2 Reducer 2: 0(+2)/3
Map 1: 1(+1)/2 Reducer 2: 0(+2)/3
Map 1: 2/2 Reducer 2: 0(+3)/3
Map 1: 2/2 Reducer 2: 2(+1)/3
Map 1: 2/2 Reducer 2: 3/3
OK
months    sum_of_purchases
10        245624
11        322417
Time taken: 59.883 seconds, Fetched: 2 row(s)
hive> 
```

Answer 2 =

| Month | Sum of Purchases |
|--------------|-------------------------|
| Oct | 245,624 |
| Nov | 322,417 |

Q3 Write a query to find the change in revenue generated due to purchases from October to November.

```
hive> with monthly_revenue as (select sum(case when date_format(event_time, 'MM')=10 then price else 0 end) as Oct_revenue, sum(case when date_format(event_time, 'MM')=11 then price else 0 end) as Nov_revenue from Shopping where event_type = 'purchase' and date_format(event_time, 'MM') in('10', '11')) select Nov_revenue, Oct_revenue, (Nov_revenue - Oct_revenue) as Change_in_revenue from monthly_revenue;
Query ID = hadoop_20210925105722_b5261ff4-5579-4508-9a3f-e67ff58eb887
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_1632563517868_0008)

Map 1: 0/2      Reducer 2: 0/1
Map 1: 0/2      Reducer 2: 0/1
Map 1: 0(+2)/2  Reducer 2: 0/1
Map 1: 1(+1)/2  Reducer 2: 0(+1)/1
Map 1: 1(+1)/2  Reducer 2: 0(+1)/1
Map 1: 2/2       Reducer 2: 0(+1)/1
Map 1: 2/2       Reducer 2: 1/1
OK
+-----+-----+-----+
| nov_revenue | oct_revenue | change_in_revenue |
+-----+-----+-----+
| 1531016.900000122 | 1211538.4299997438 | 319478.4700003781 |
+-----+-----+-----+
Time taken: 73.828 seconds, Fetched: 1 row(s)
hive>
```

Answer 3

| Nov Revenue | Oct Revenue | Change in revenue |
|--------------------|--------------------|--------------------------|
| 1,531,017 | 1,211,538 | 319,478 |

Q4 Find distinct categories of products. Categories with null category code can be ignored.

```
hive> select distinct split(category_code,'\\.')[0] as Category from Shopping where split(category_code,'\\.')[0] <> '' ;
Query ID = hadoop_20210925110718_084df04a-elf0-4518-a9be-626510af18f4
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_1632563517868_0009)

Map 1: 0/2      Reducer 2: 0/5
Map 1: 0/2      Reducer 2: 0/5
Map 1: 0(+2)/2  Reducer 2: 0/5

Map 1: 1(+1)/2  Reducer 2: 0(+1)/5
Map 1: 1(+1)/2  Reducer 2: 0(+1)/5
Map 1: 1(+1)/2  Reducer 2: 0(+2)/5
Map 1: 2/2       Reducer 2: 0(+3)/5
Map 1: 2/2       Reducer 2: 1(+2)/5
Map 1: 2/2       Reducer 2: 1(+3)/5
Map 1: 2/2       Reducer 2: 2(+2)/5
Map 1: 2/2       Reducer 2: 3(+2)/5
Map 1: 2/2       Reducer 2: 4(+1)/5
Map 1: 2/2       Reducer 2: 4(+0)/5
Map 1: 2/2       Reducer 2: 5/5

OK
category
furniture
appliances
accessories
apparel
sport
stationery
Time taken: 66.814 seconds, Fetched: 6 row(s)
hive>
```

Answer 4

| Categories |
|-------------|
| Furniture |
| Appliances |
| Accessories |
| Apparel |
| Sport |
| Stationery |

Q5 Find the total number of products available under each category.

```
hive> select split(category_code,'\\.') [0] as category, count(product_id) as no_of_products from Shopping where split(category_code,'\\.') [0] <> '' group by split(category_code,'\\.') [0] order by no_of_products desc;
Query ID = hadoop_20210925111510_48cac138-4ada-4f30-8f35-0ad30d334ed0
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_1632563517868_0010)

Map 1: -- Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 0/2 Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 0/2 Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/5 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 3(+2)/5 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 4(+1)/5 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 5/5 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 5/5 Reducer 3: 1/1
OK
category      no_of_products
8594895
appliances    61736
stationery    26722
furniture     23604
apparel       18232
accessories   12929
sport          2
Time taken: 72.268 seconds, Fetched: 7 row(s)
hive>
```

Answer 5

| Category | No of products |
|-------------|----------------|
| appliances | 61,736 |
| stationery | 26,722 |
| furniture | 23,604 |
| apparel | 18,232 |
| accessories | 12,929 |
| sport | 2 |

Q6 Which brand had the maximum sales in October and November combined?

```
hive> with max_sales_brand as (select brand, sum(case when date_format(event_time, 'MM')=10 then price else 0 end) as Oct_sales, sum(case when date_format(event_time, 'MM')=11 then price else 0 end) as Nov_sales from shopping where (event_type='purchase' and date_format(event_time, 'MM') in ('10','11')) and brand <> '') group by brand) select brand, Nov_sales + Oct_sales as Total_sales from max_sales_brand order by Total_sales desc limit 1;
Query ID = hadoop_20210925112751_cfffb0057-4c9b-4e22-b5fb-1758c469831e
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_1632563517868_0011)

Map 1: -- Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 0(+2)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 1(+1)/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 1/1
OK
brand total_sales
1094188.2999999556
Time taken: 75.382 seconds, Fetched: 1 row(s)
hive>
```

**(I don't know why there seems to be blank values in brand column)*

```
brand total_sales
1094188.2999999556
Time taken: 75.382 seconds, Fetched: 1 row(s)
hive> with max_sales_brand as (select brand, sum(case when date_format(event_time, 'MM')=10 then price else 0 end) as Oct_sales, sum(case when date_format(event_time, 'MM')=11 then price else 0 end) as Nov_sales from shopping where (event_type='purchase' and date_format(event_time, 'MM') in ('10','11')) and brand <> '') group by brand) select brand, Nov_sales + Oct_sales as Total_sales from max_sales_brand order by Total_sales desc limit 2;
Query ID = hadoop_20210925113345_73c6ff31-6741-4645-8981-31046c2ea7d9
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1632563517868_0011)

Map 1: 0/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0(+1)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 0(+2)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 1(+1)/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 1/1
OK
brand total_sales
1094188.2999999556
runail 148297.9400000003
Time taken: 69.037 seconds, Fetched: 2 row(s)
hive>
```

Answer 6

Runail | 148,298

Q7 Which brands increased their sales from October to November?

```
hive> with monthly_revenue as (select brand, sum(case when date_format(event_time, 'MM')=10 then price else 0 end) as Oct_revenue, sum(case when date_format(event_time, 'MM')=11 then price else 0 end) as Nov_revenue from Shopping where event_type='purchase' and date_format(event_time, 'MM') in ('10','11') group by brand) select brand, Oct_revenue, Nov_revenue, Nov_revenue - Oct_revenue as Change_in_sales from monthly_revenue where (Nov_revenue - Oct_revenue)>0 order by Change_in_sales;
Query ID = hadoop_20210925114545_5f9ab8a7-6b78-43c0-bfe9-036c10004289
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_1632563517868_0012)

Map 1: -/- Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0(+1)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 0(+2)/2 Reducer 3: 0/1
Map 1: 2/2 Reducer 2: 1(+1)/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 0(+1)/1
Map 1: 2/2 Reducer 2: 2/2 Reducer 3: 1/1
OK
brand oct_revenue nov_revenue change_in_sales
ovale 2.54 3.1 0.56
cosima 20.23 20.92999999999993 0.6999999999999922
grace 100.92000000000002 102.61000000000001 1.6899999999999977
helloganic 0.0 3.1 3.1
skinity 8.88 12.440000000000001 3.5600000000000005
bodyton 1376.339999999974 1380.639999999992 4.3000000000017735
moyou 5.71 10.280000000000001 4.5700000000000001
neoleor 43.41 51.7 8.290000000000006
soleo 204.200000000003 212.529999999998 8.32999999999501
strong 29196.6299999994 38671.269999999924 9474.639999999985
ingarden 23161.390000000138 33566.21000000009 10404.8199999999
49
lianail 5892.839999999975 16394.240000000245 10501.40000000027
uno 35302.02999999977 51039.749999998035 15737.719999998262
grattol 35445.5400000011 71472.71000000068 36027.169999999576
474679.0599999623 619509.2399999934 144830.18000003108
Time taken: 75.579 seconds, Fetched: 161 row(s)
hive> [hadoop@ip-172-31-11-246 ~]$
```

Answer 7

| Brand | Oct Revenue | Nov Revenue | Change In Sales |
|------------|-------------|-------------|-----------------|
| ovale | 2.54 | 3.1 | 0.56 |
| cosima | 20.23 | 20.93 | 0.7 |
| grace | 100.92 | 102.61 | 1.69 |
| helloganic | 0 | 3.1 | 3.1 |
| skinity | 8.88 | 12.44 | 3.56 |
| bodyton | 1376.34 | 1380.64 | 4.3 |
| moyou | 5.71 | 10.28 | 4.57 |

| | | | |
|---------------|---------|---------|-------|
| neoleor | 43.41 | 51.7 | 8.29 |
| soleo | 204.2 | 212.53 | 8.33 |
| jaguar | 1102.11 | 1110.65 | 8.54 |
| tertio | 236.16 | 245.8 | 9.64 |
| fly | 17.14 | 27.17 | 10.03 |
| rasyan | 18.8 | 28.94 | 10.14 |
| deoproce | 316.84 | 329.17 | 12.33 |
| barbie | 0 | 12.39 | 12.39 |
| supertan | 50.37 | 66.51 | 16.14 |
| treaclemoon | 163.37 | 181.49 | 18.12 |
| kamill | 63.01 | 81.49 | 18.48 |
| juno | 0 | 21.08 | 21.08 |
| veraclara | 50.11 | 71.21 | 21.1 |
| glysolid | 69.73 | 91.59 | 21.86 |
| godefroy | 401.22 | 425.12 | 23.9 |
| binacil | 0 | 24.26 | 24.26 |
| blixz | 38.95 | 63.4 | 24.45 |
| profepil | 93.36 | 118.02 | 24.66 |
| estelare | 444.81 | 471.87 | 27.06 |
| orly | 902.38 | 931.09 | 28.71 |
| biore | 60.65 | 90.31 | 29.66 |
| beautyblender | 78.74 | 109.41 | 30.67 |
| vilenta | 197.6 | 231.21 | 33.61 |
| mavala | 409.04 | 446.32 | 37.28 |
| likato | 296.06 | 340.97 | 44.91 |
| ladykin | 125.65 | 170.57 | 44.92 |
| foamie | 35.04 | 80.49 | 45.45 |
| elskin | 251.09 | 307.65 | 56.56 |
| balbcare | 155.33 | 212.38 | 57.05 |
| koelcia | 55.5 | 112.75 | 57.25 |
| profhenna | 679.23 | 736.85 | 57.62 |
| kares | 0 | 59.45 | 59.45 |
| marutaka-foot | 49.22 | 109.33 | 60.11 |
| dewal | 0 | 61.29 | 61.29 |
| inm | 288.02 | 351.21 | 63.19 |
| laboratorium | 246.5 | 312.52 | 66.02 |
| cutrin | 299.37 | 367.62 | 68.25 |
| egomania | 77.47 | 146.04 | 68.57 |
| konad | 739.83 | 810.67 | 70.84 |
| nirvel | 163.04 | 234.33 | 71.29 |
| koelf | 422.73 | 507.29 | 84.56 |
| plazan | 101.37 | 194.01 | 92.64 |
| aura | 83.95 | 177.51 | 93.56 |
| kerasys | 430.91 | 525.2 | 94.29 |

| | | | |
|--------------|----------|----------|--------|
| enjoy | 41.35 | 136.57 | 95.22 |
| depilflax | 2707.07 | 2803.78 | 96.71 |
| eos | 54.34 | 152.61 | 98.27 |
| carmex | 145.08 | 243.36 | 98.28 |
| batiste | 772.4 | 874.17 | 101.77 |
| osmo | 645.58 | 762.31 | 116.73 |
| dizao | 819.13 | 945.51 | 126.38 |
| igrobeauty | 513.66 | 645.07 | 131.41 |
| finish | 98.38 | 230.38 | 132 |
| nefertiti | 233.52 | 366.64 | 133.12 |
| elizavecca | 70.53 | 204.3 | 133.77 |
| miskin | 158.04 | 293.07 | 135.03 |
| latinoil | 249.52 | 384.59 | 135.07 |
| farmona | 1692.46 | 1843.43 | 150.97 |
| cristalinas | 427.63 | 584.95 | 157.32 |
| chi | 358.94 | 538.61 | 179.67 |
| matreshka | 0 | 182.67 | 182.67 |
| freshbubble | 318.7 | 502.34 | 183.64 |
| mane | 66.79 | 260.26 | 193.47 |
| keen | 236.35 | 435.62 | 199.27 |
| ecocraft | 41.16 | 241.95 | 200.79 |
| fedua | 52.38 | 263.81 | 211.43 |
| provoc | 827.99 | 1063.82 | 235.83 |
| skinlite | 651.94 | 890.45 | 238.51 |
| entity | 479.71 | 719.26 | 239.55 |
| trind | 298.07 | 542.96 | 244.89 |
| protokeratin | 201.25 | 456.79 | 255.54 |
| beauugreen | 511.51 | 768.35 | 256.84 |
| bluesky | 10307.24 | 10565.53 | 258.29 |
| candy | 534.96 | 799.38 | 264.42 |
| insight | 1443.7 | 1721.96 | 278.26 |
| kocostar | 310.85 | 594.93 | 284.08 |
| happyfons | 801.92 | 1091.59 | 289.67 |
| kims | 330.04 | 632.04 | 302 |
| shary | 871.96 | 1176.49 | 304.53 |
| nitrile | 847.28 | 1162.68 | 315.4 |
| lowence | 242.84 | 567.75 | 324.91 |
| jas | 3318.96 | 3657.43 | 338.47 |
| ellips | 245.85 | 606.04 | 360.19 |
| lador | 2083.61 | 2471.53 | 387.92 |
| naomi | 0 | 389 | 389 |
| kiss | 421.55 | 817.33 | 395.78 |
| yu-r | 271.41 | 673.71 | 402.3 |
| sophin | 1067.86 | 1515.52 | 447.66 |

| | | | |
|-------------|----------|----------|---------|
| farmavita | 837.37 | 1291.97 | 454.6 |
| bioaqua | 942.89 | 1398.12 | 455.23 |
| greymy | 29.21 | 489.49 | 460.28 |
| gehwol | 1089.07 | 1557.68 | 468.61 |
| matrix | 3243.25 | 3726.74 | 483.49 |
| limoni | 1308.9 | 1796.6 | 487.7 |
| s.care | 412.68 | 913.07 | 500.39 |
| coifin | 903 | 1428.49 | 525.49 |
| uskusi | 5142.27 | 5690.31 | 548.04 |
| airnails | 5118.9 | 5691.52 | 572.62 |
| brownxenna | 14331.37 | 14916.73 | 585.36 |
| kinetics | 6334.25 | 6945.26 | 611.01 |
| kosmekka | 1181.44 | 1813.37 | 631.93 |
| kaaral | 4412.43 | 5086.07 | 673.64 |
| refectocil | 2716.18 | 3475.58 | 759.4 |
| rosi | 3077.04 | 3841.56 | 764.52 |
| solomeya | 1899.7 | 2685.8 | 786.1 |
| misssha | 1293.83 | 2150.28 | 856.45 |
| levissime | 2227.5 | 3085.31 | 857.81 |
| art-visage | 2092.71 | 2997.8 | 905.09 |
| ecolab | 262.85 | 1214.3 | 951.45 |
| nagaraku | 4369.74 | 5327.68 | 957.94 |
| sanoto | 157.14 | 1209.68 | 1052.54 |
| markell | 1768.75 | 2834.43 | 1065.68 |
| metzger | 5373.45 | 6457.16 | 1083.71 |
| de.lux | 1659.7 | 2775.51 | 1115.81 |
| swarovski | 1887.93 | 3043.16 | 1155.23 |
| beauty-free | 554.17 | 1782.86 | 1228.69 |
| zeitun | 708.66 | 2009.63 | 1300.97 |
| joico | 705.52 | 2015.1 | 1309.58 |
| severina | 4775.88 | 6120.48 | 1344.6 |
| irisk | 45591.96 | 46946.04 | 1354.08 |
| oniq | 8425.41 | 9841.65 | 1416.24 |
| levrana | 2243.56 | 3664.1 | 1420.54 |
| roublöff | 3491.36 | 4913.77 | 1422.41 |
| smart | 4457.26 | 5902.14 | 1444.88 |
| shik | 3341.2 | 4839.72 | 1498.52 |
| domix | 10472.05 | 12009.17 | 1537.12 |
| artex | 2730.64 | 4327.25 | 1596.61 |
| beautix | 10493.95 | 12222.95 | 1729 |
| milv | 3904.94 | 5642.01 | 1737.07 |
| masura | 31266.08 | 33058.47 | 1792.39 |
| f.o.x | 6624.23 | 8577.28 | 1953.05 |
| kapous | 11927.16 | 14093.08 | 2165.92 |

| | | | |
|------------|----------|----------|----------|
| concept | 11032.14 | 13380.4 | 2348.26 |
| estel | 21756.75 | 24142.67 | 2385.92 |
| kaypro | 881.34 | 3268.7 | 2387.36 |
| benovy | 409.62 | 3259.97 | 2850.35 |
| italwax | 21940.24 | 24799.37 | 2859.13 |
| yoko | 8756.91 | 11707.88 | 2950.97 |
| haruyama | 9390.69 | 12352.91 | 2962.22 |
| marathon | 7280.75 | 10273.1 | 2992.35 |
| lovely | 8704.38 | 11939.06 | 3234.68 |
| bpw.style | 11572.15 | 14837.44 | 3265.29 |
| staleks | 8519.73 | 11875.61 | 3355.88 |
| freedecor | 3421.78 | 7671.8 | 4250.02 |
| runail | 71539.28 | 76758.66 | 5219.38 |
| polarus | 6013.72 | 11371.93 | 5358.21 |
| cosmoprofi | 8322.81 | 14536.99 | 6214.18 |
| jessnail | 26287.84 | 33345.23 | 7057.39 |
| strong | 29196.63 | 38671.27 | 9474.64 |
| ingarden | 23161.39 | 33566.21 | 10404.82 |
| lianail | 5892.84 | 16394.24 | 10501.4 |
| uno | 35302.03 | 51039.75 | 15737.72 |
| grattol | 35445.54 | 71472.71 | 36027.17 |

Q8 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 Total_exp from Shopping where event_type='purchase' group by user_id order by Total_exp desc limit 10;
Query ID = hadoop_20210925121527_3fc2c18f-7c72-4e82-8032-7e0334f542f2
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1632563517868_0015)

Map 1: 0/2      Reducer 2: 0/3  Reducer 3: 0/1
Map 1: 0/2      Reducer 2: 0/3  Reducer 3: 0/1
Map 1: 0(+1)/2 Reducer 2: 0/3  Reducer 3: 0/1
Map 1: 0(+2)/2 Reducer 2: 0/3  Reducer 3: 0/1

Map 1: 0(+2)/2 Reducer 2: 0/3  Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/3  Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+1)/3  Reducer 3: 0/1
Map 1: 1(+1)/2 Reducer 2: 0(+2)/3  Reducer 3: 0/1
Map 1: 2/2      Reducer 2: 0(+3)/3  Reducer 3: 0/1
Map 1: 2/2      Reducer 2: 1(+2)/3  Reducer 3: 0(+1)/1
Map 1: 2/2      Reducer 2: 2(+1)/3  Reducer 3: 0(+1)/1
Map 1: 2/2      Reducer 2: 3/3   Reducer 3: 0(+1)/1
Map 1: 2/2      Reducer 2: 3/3   Reducer 3: 1/1
OK
user_id total_exp
557790271      2715.869999999991
150318419       1645.97
562167663       1352.8500000000004
531900924       1329.4500000000003
557850743       1295.4800000000002
522130011       1185.389999999994
561592095       1109.699999999996
431950134       1097.589999999995
566576008       1056.3600000000017
521347209       1040.909999999999
Time taken: 60.344 seconds, Fetched: 10 row(s)
hive>
```

Answer 8

| User_id | Total Expense |
|-----------|---------------|
| 557790271 | 2,716 |
| 150318419 | 1,646 |
| 562167663 | 1,353 |
| 531900924 | 1,329 |
| 557850743 | 1,295 |
| 522130011 | 1,185 |
| 561592095 | 1,110 |
| 431950134 | 1,098 |
| 566576008 | 1,056 |
| 521347209 | 1,041 |

Using optimization techniques & verifying improvement of performance

1. Allowing dynamic partition to be allowed instead of default & using non strict mode so all partitions can be determined dynamically

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

2. Creating table with partition on event_types and bucket on price columns

```
hive> create table if not exists dyc_par_buc_shopping(event_time timestamp, product_id string, category_id string, category_code string, brand string, price float, user_id bigint, user_session string) partitioned by (event_type string) clustered by (price) into 7 buckets row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde' stored as textfile;
OK
Time taken: 0.257 seconds
```

3. Adding data into the partitioned & bucketed table from already available Shopping table

```
hive> insert into table dyc_par_buc_shopping partition (event_type) select event_time, product_id, category_id, category_code, brand, price, user_id, user_session, event_type from Shopping;
Query ID = hadoop_20210925123013_1629c243-0e84-4daf-bafa-88543f1bc5cf
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_1632563517868_0016)
```

```
Map 1: 2/2      Reducer 2: 3(+2)/5
Map 1: 2/2      Reducer 2: 4(+1)/5
Map 1: 2/2      Reducer 2: 5/5
Loading data to table default.dyc_par_buc_shopping partition (event_type=null)

Time taken to load dynamic partitions: 0.743 seconds
Time taken for adding to write entity : 0.003 seconds
OK
event_time      product_id      category_id      category_code      brand      price      user_id      user_session      event_type
Time taken: 167.189 seconds
hive>
```

Before checking partitioned and bucketed table first we need to exit from Hive environment then we can check whether the partitioned and bucketed table are present in hive warehouse

```
hive> exit;
[hadoop@ip-172-31-11-246 ~]$ hadoop fs -ls /user/hive/warehouse/dyc_par_buc_shopping
Found 4 items
drwxrwxrwt  - hadoop hadoop          0 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=cart
drwxrwxrwt  - hadoop hadoop          0 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase
drwxrwxrwt  - hadoop hadoop          0 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=remove_from_cart
drwxrwxrwt  - hadoop hadoop          0 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=view
[hadoop@ip-172-31-11-246 ~]$
```

Checking partition by event type = purchase

```
[hadoop@ip-172-31-11-246 ~]$ hadoop fs -ls /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase
Found 7 items
-rwxrwxrwt  1 hadoop hadoop  13052654 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000000_0
-rwxrwxrwt  1 hadoop hadoop  9399111 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000001_0
-rwxrwxrwt  1 hadoop hadoop  12636711 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000002_0
-rwxrwxrwt  1 hadoop hadoop  10650131 2021-09-25 12:31 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000003_0
-rwxrwxrwt  1 hadoop hadoop  7226455 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000004_0
-rwxrwxrwt  1 hadoop hadoop  10737803 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000005_0
-rwxrwxrwt  1 hadoop hadoop  7825305 2021-09-25 12:32 /user/hive/warehouse/dyc_par_buc_shopping/event_type=purchase/000006_0
[hadoop@ip-172-31-11-246 ~]$
```

Verifying the performance improvement by executing the Q8 command on partitioned and bucketed table

```
hive> select user_id, sum(price) as Total_exp from dyc_par_buc_shopping where event_type='purchase' group by user_id order by Total_exp desc limit 10;
Query ID = hadoop_20210925124552_4af6ec3-7ed3-4ca7-8219-421a59732f39
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1632563517868_0017)

Map 1: 0/3      Reducer 2: 0/1  Reducer 3: 0/1
Map 1: 0/3      Reducer 2: 0/1  Reducer 3: 0/1
Map 1: 0/3      Reducer 2: 0/1  Reducer 3: 0/1
Map 1: 0(+1)/3  Reducer 2: 0/1  Reducer 3: 0/1
Map 1: 0(+2)/3  Reducer 2: 0/1  Reducer 3: 0/1
Map 1: 0(+3)/3  Reducer 2: 0/1  Reducer 3: 0/1
```

```

Map 1: 3/3      Reducer 2: 1/1  Reducer 3: 0 (+1)/1
Map 1: 3/3      Reducer 2: 1/1  Reducer 3: 1/1
OK
557790271      2715.869999999996
150318419      1645.97
562167663      1352.8500000000001
531900924      1329.4500000000003
557850743      1295.4800000000005
522130011      1185.389999999999
561592095      1109.7
431950134      1097.5900000000001
566576008      1056.3600000000006
521347209      1040.9100000000003
Time taken: 25.235 seconds, Fetched: 10 row(s)

```

The time taken has decreased to 25.235 seconds from 60.344 seconds thereby confirming the reduction in execution time by partitioning

Now we will terminate the EMR cluster

Cluster: Hive_case_study **Terminating** Terminated by user request

[Summary](#) [Application user interfaces](#) [Monitoring](#) [Hardware](#) [Configurations](#) [Events](#) [Steps](#) [Bootstrap actions](#)

Summary

ID: j-28VKLR2PIOCMD
Creation date: 2021-09-25 15:14 (UTC+5:30)
Elapsed time: 3 hours, 15 minutes
After last step completes: Cluster waits
Termination protection: Off
Tags: --
Master public DNS: ec2-54-147-124-27.compute-1.amazonaws.com [Connect to the Master Node Using SSH](#)

Configuration details

Release label: emr-5.29.0
Hadoop distribution: Amazon 2.8.5
Applications: Ganglia 3.7.2, Hive 2.3.6, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2
Log URI: s3://aws-logs-206892056184-us-east-1/elasticmapreduce/ [Logs](#)
EMRFS consistent view: Disabled