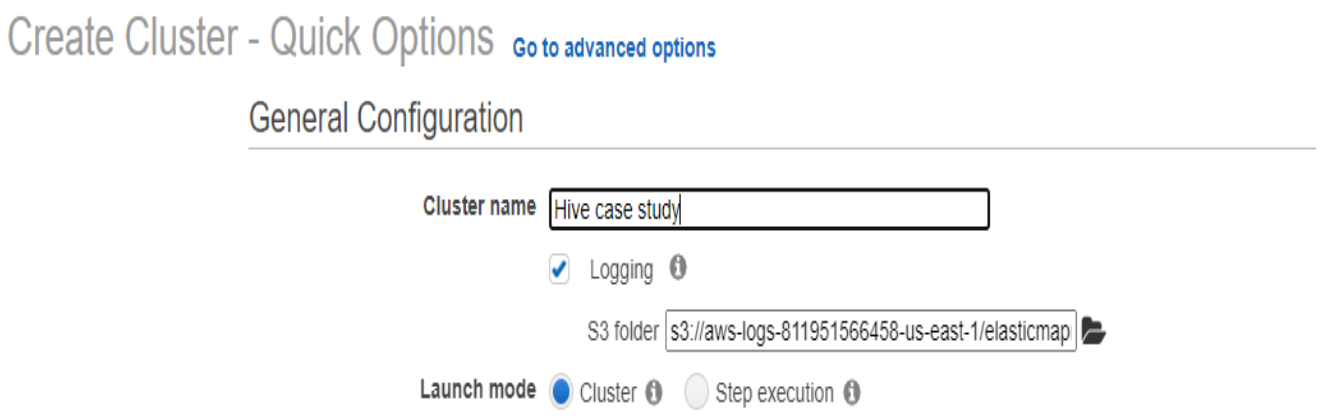
HIVE CASE STUDY ASSIGNMENT

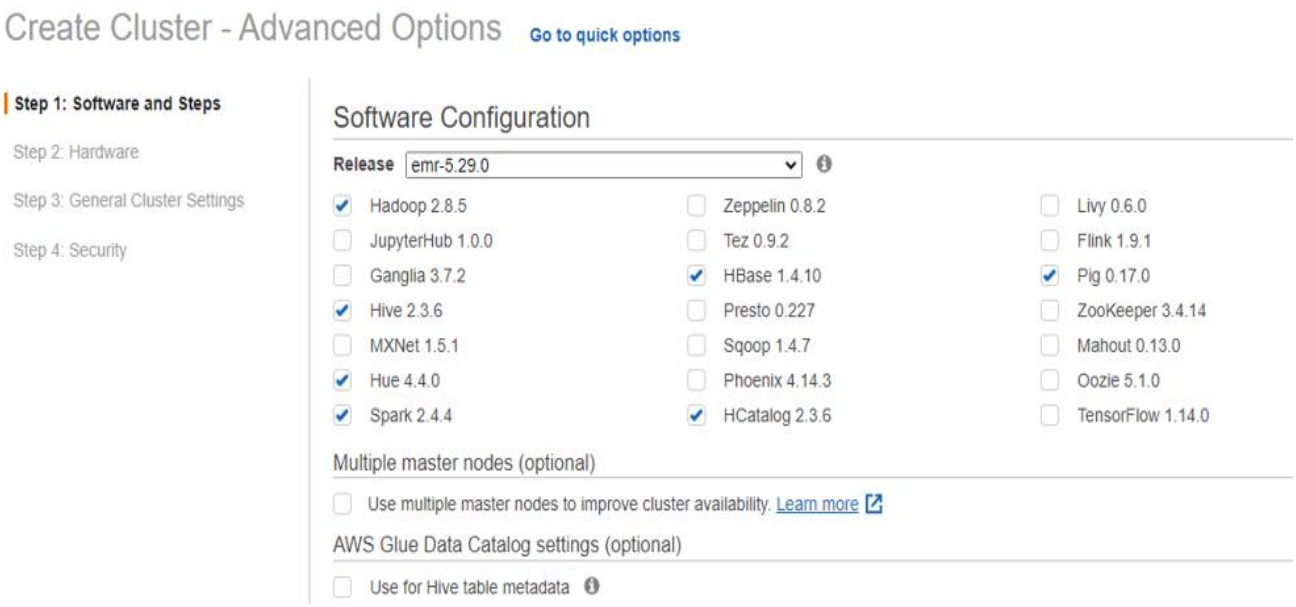
[DS C41 – 2022]

E-COMMERCE SALES DATA ANALYSIS

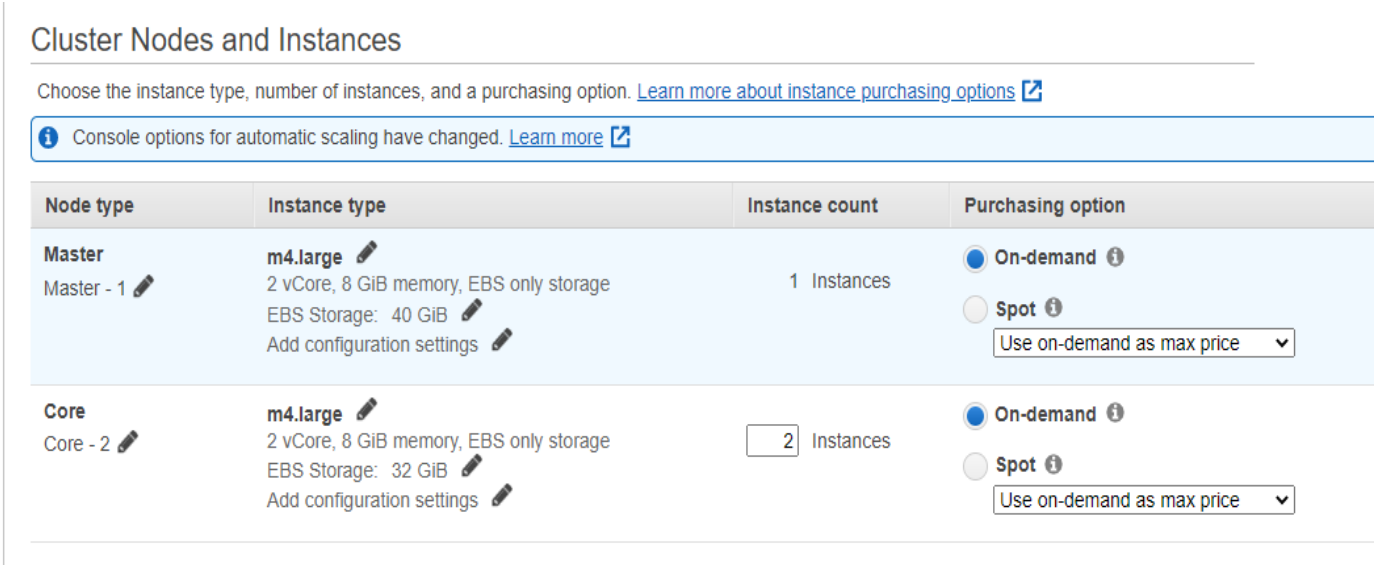
Creating the EMR cluster that utilizes the Hive services and Move the data from the S3 bucket into the HDFS.

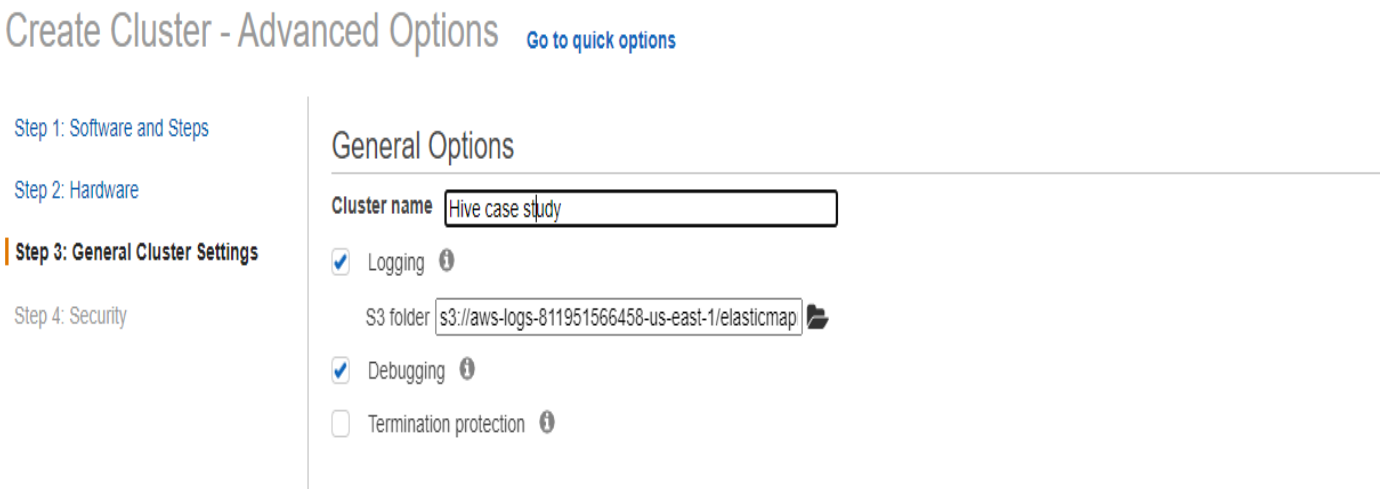


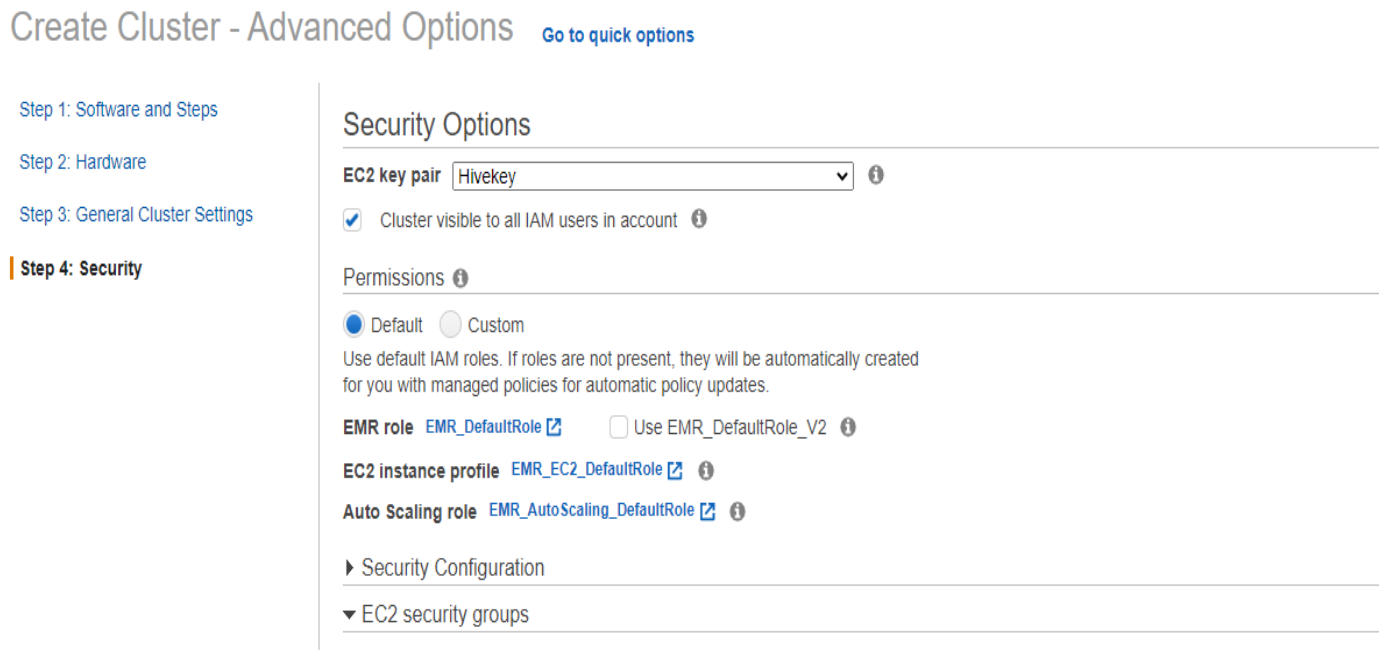
As suggested, we are using the emr-5.29.0 version for case study



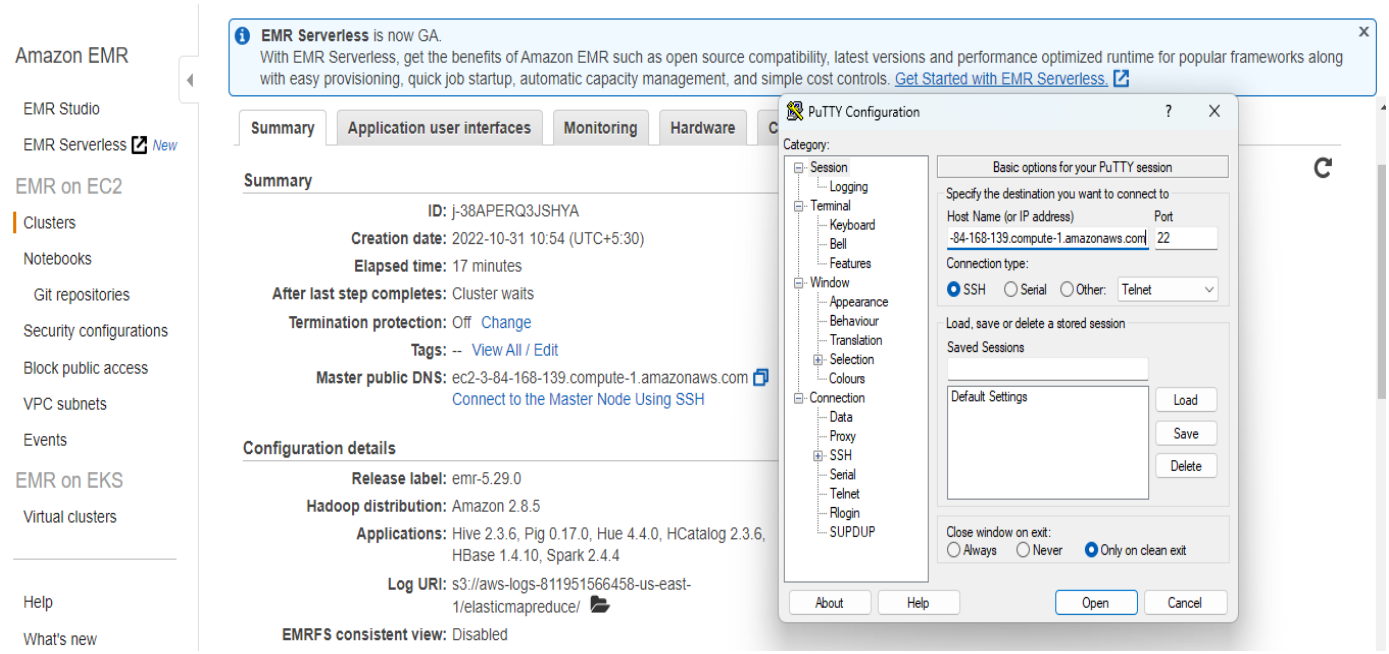
Using a 2-node EMR cluster with both the master and core nodes as M4.large.

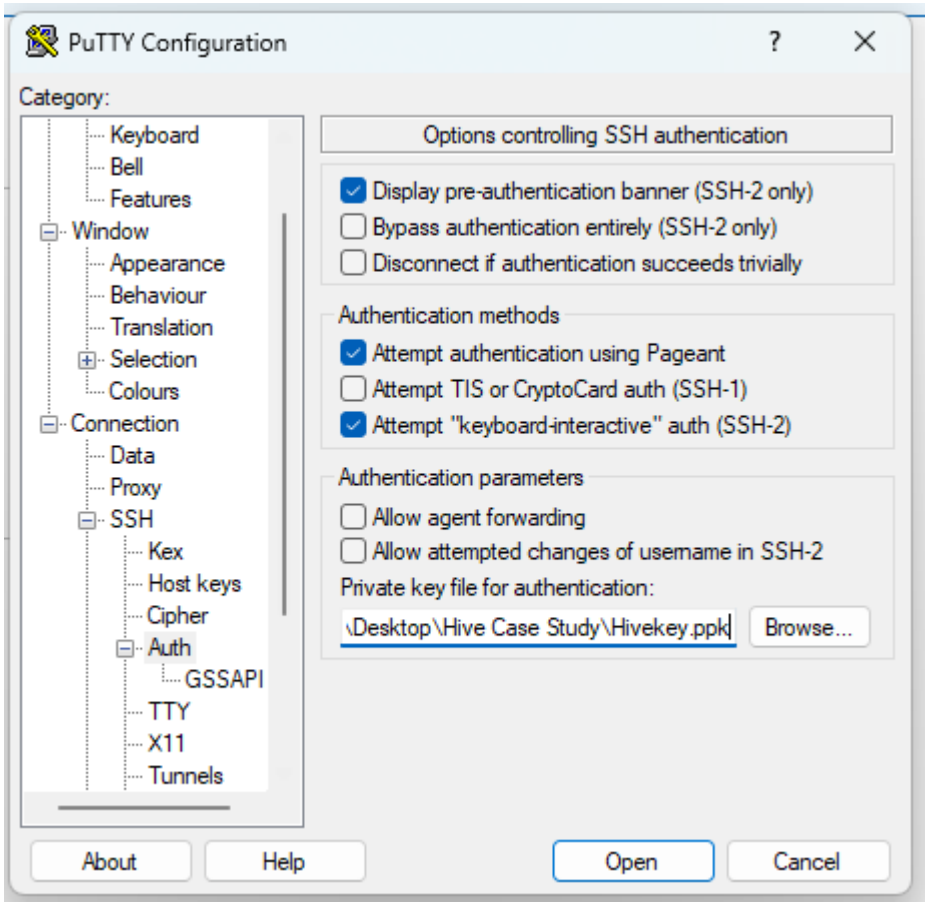




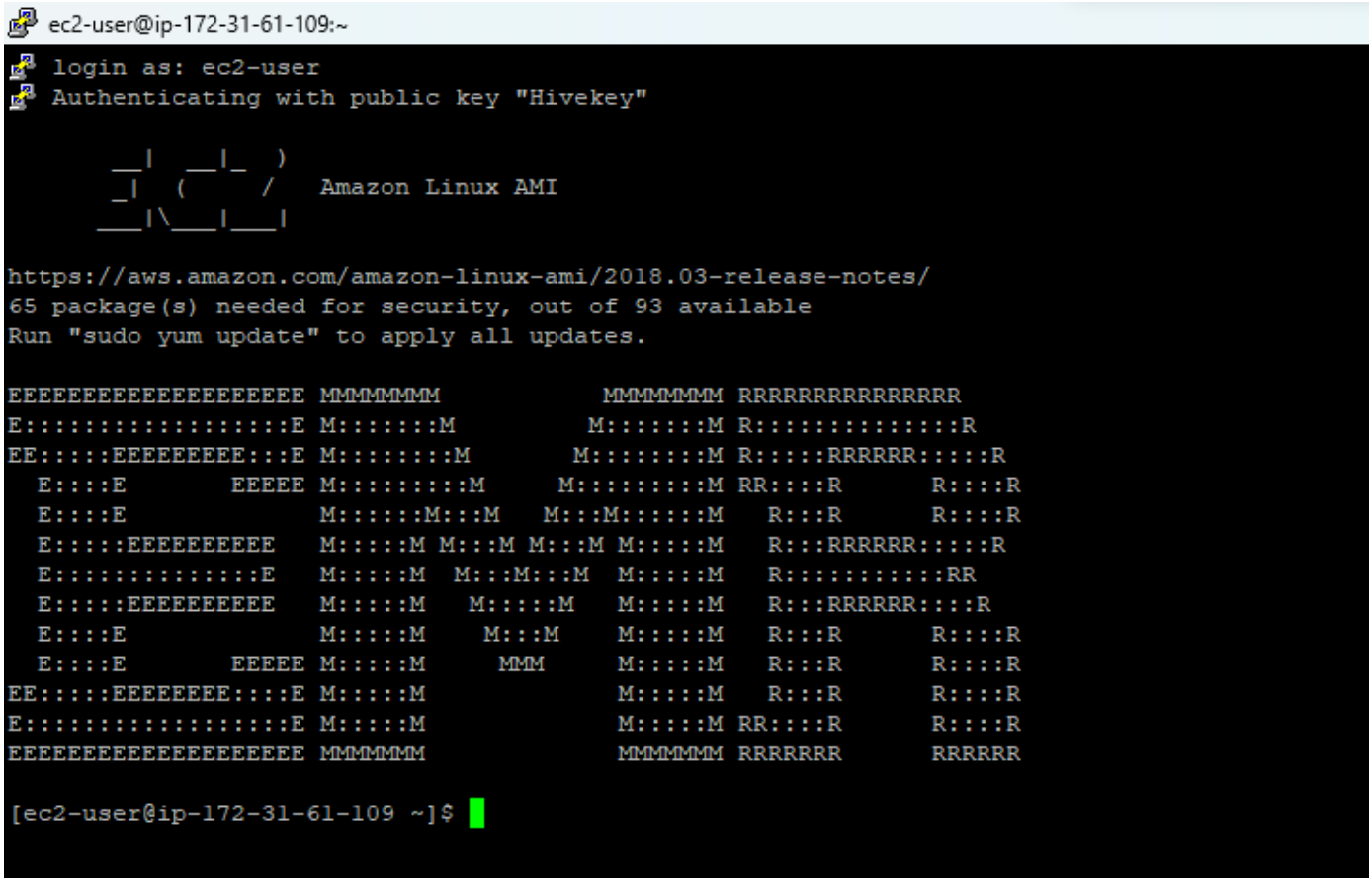


Connecting to putty

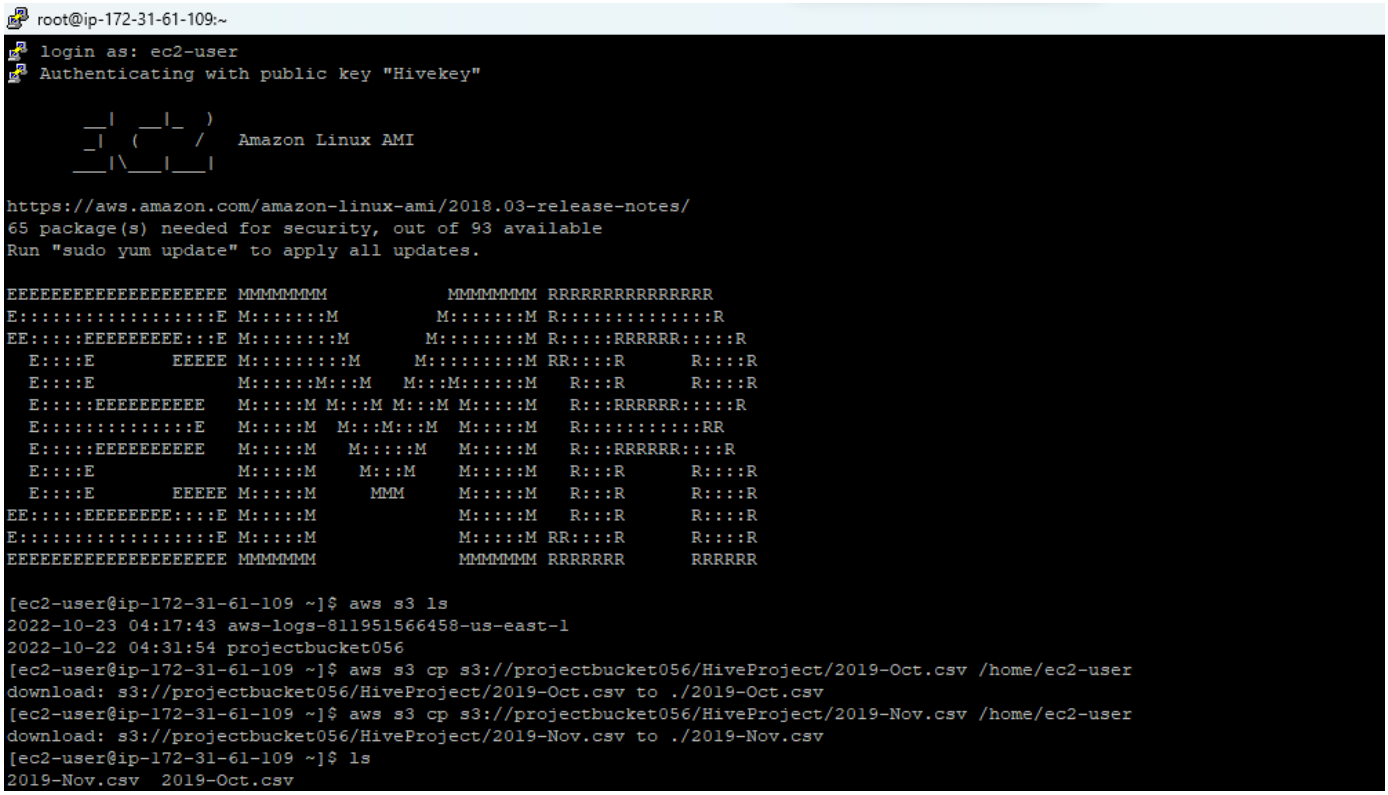


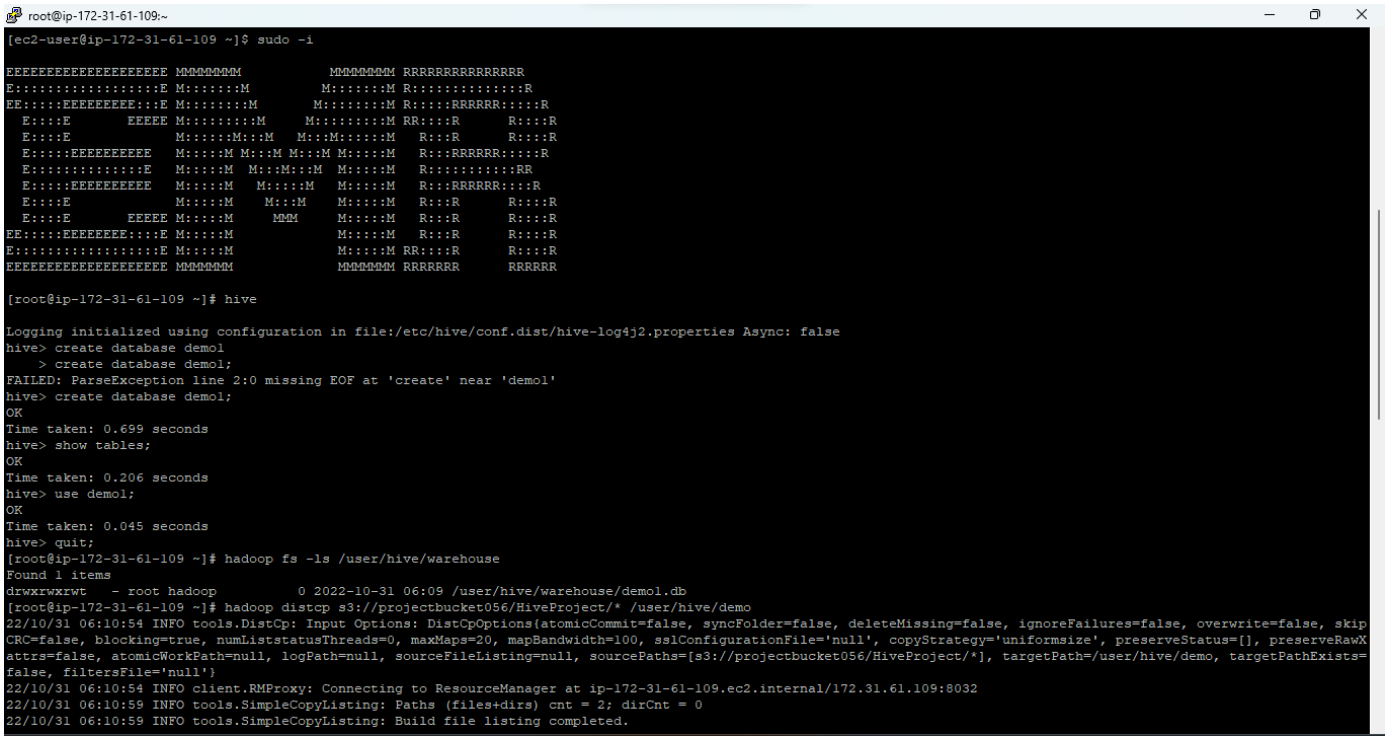


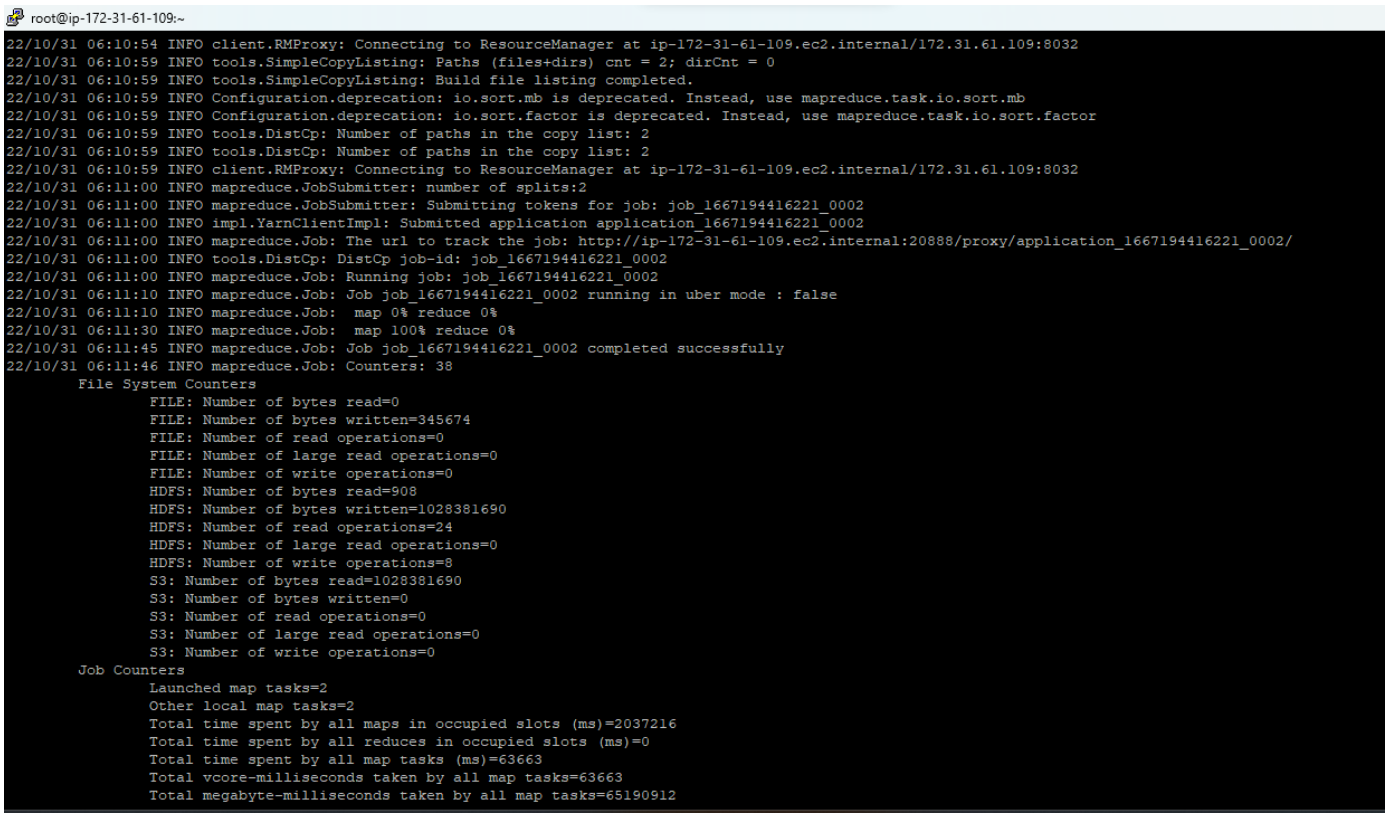
Launching an EMR Cluster that utilizes the Hive services.



Creating the database and launching Hive queries on your EMR cluster




root@ip-172-31-61-109:~
[ec2-user@ip-172-31-61-109 ~]$ sudo -i
[root@ip-172-31-61-109 ~]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database demo1
> create database demo1;
FAILED: ParseException line 2:0 missing EOF at 'create' near 'demo1'
hive> create database demo1;
OK
Time taken: 0.699 seconds
hive> show tables;
OK
Time taken: 0.206 seconds
hive> use demo1;
OK
Time taken: 0.045 seconds
hive> quit;
[root@ip-172-31-61-109 ~]# hadoop fs -ls /user/hive/warehouse
Found 1 items
drwxrwxrwt - root hadoop 0 2022-10-31 06:09 /user/hive/warehouse/demo1.db
[root@ip-172-31-61-109 ~]# hadoop distcp s3://projectbucket056/HiveProject/* /user/hive/demo
22/10/31 06:10:54 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://projectbucket056/HiveProject/*], targetPath=/user/hive/demo, targetPathExists=false, filtersFile='null'}
22/10/31 06:10:54 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-61-109.ec2.internal/172.31.61.109:8032
22/10/31 06:10:59 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 2; dirCnt = 0
22/10/31 06:10:59 INFO tools.SimpleCopyListing: Build file listing completed.


root@ip-172-31-61-109:~
22/10/31 06:10:54 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-61-109.ec2.internal/172.31.61.109:8032
22/10/31 06:10:59 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 2; dirCnt = 0
22/10/31 06:10:59 INFO tools.SimpleCopyListing: Build file listing completed.
22/10/31 06:10:59 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
22/10/31 06:10:59 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
22/10/31 06:10:59 INFO tools.DistCp: Number of paths in the copy list: 2
22/10/31 06:10:59 INFO tools.DistCp: Number of paths in the copy list: 2
22/10/31 06:10:59 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-61-109.ec2.internal/172.31.61.109:8032
22/10/31 06:11:00 INFO mapreduce.JobSubmitter: number of splits:2
22/10/31 06:11:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667194416221_0002
22/10/31 06:11:00 INFO impl.YarnClientImpl: Submitted application application_1667194416221_0002
22/10/31 06:11:00 INFO mapreduce.Job: The url to track the job: http://ip-172-31-61-109.ec2.internal:20888/proxy/application_1667194416221_0002/
22/10/31 06:11:00 INFO tools.DistCp: DistCp job-id: job_1667194416221_0002
22/10/31 06:11:00 INFO mapreduce.Job: Running job: job_1667194416221_0002
22/10/31 06:11:10 INFO mapreduce.Job: Job job_1667194416221_0002 running in uber mode : false
22/10/31 06:11:10 INFO mapreduce.Job: map 0% reduce 0%
22/10/31 06:11:30 INFO mapreduce.Job: map 100% reduce 0%
22/10/31 06:11:45 INFO mapreduce.Job: Job job_1667194416221_0002 completed successfully
22/10/31 06:11:46 INFO mapreduce.Job: Counters: 38
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=345674
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=908
HDFS: Number of bytes written=1028381690
HDFS: Number of read operations=24
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
S3: Number of bytes read=1028381690
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=2
Other local map tasks=2
Total time spent by all maps in occupied slots (ms)=2037216
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=63663
Total vcore-milliseconds taken by all map tasks=63663
Total megabyte-milliseconds taken by all map tasks=65190912

```
root@ip-172-31-61-109:~
                FILE: Number of bytes written=345674
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=908
                HDFS: Number of bytes written=1028381690
                HDFS: Number of read operations=24
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=8
                S3: Number of bytes read=1028381690
                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=2
                Other local map tasks=2
                Total time spent by all maps in occupied slots (ms)=2037216
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=63663
                Total vcore-milliseconds taken by all map tasks=63663
                Total megabyte-milliseconds taken by all map tasks=65190912
        Map-Reduce Framework
                Map input records=2
                Map output records=0
                Input split bytes=270
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=1211
                CPU time spent (ms)=42620
                Physical memory (bytes) snapshot=1130541056
                Virtual memory (bytes) snapshot=6622527488
                Total committed heap usage (bytes)=933756928
        File Input Format Counters
                Bytes Read=638
        File Output Format Counters
                Bytes Written=0
        DistCp Counters
                Bytes Copied=1028381690
                Bytes Expected=1028381690
                Files Copied=2
[root@ip-172-31-61-109 ~]#
```

Creating the structure of database

```
[root@ip-172-31-61-109 ~]# hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> use demo1;
OK
Time taken: 0.813 seconds
```

Creating an External Table e_commerce_event

```
hive> create external table if not exists e_commerce_event(event_time timestamp ,event_type string ,product_id string ,category_id string ,category_code string ,brand s
tring ,price float ,user_id bigint ,user_session string) row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'stored as textfile location '/user/hive/demo1/'tbl
properties ("skip.header.line.count"="1");
OK
Time taken: 0.118 seconds
hive>
```

```
hive> select * from e_commerce_event limit 5;
OK
2019-11-01 00:00:02 UTC view    5802432 1487580009286598681                 0.32    562076640       09fafd6c-6c99-46b1-834f-33527f4de241
2019-11-01 00:00:09 UTC cart    5844397 1487580006317032337                 2.38    553329724       2067216c-31b5-455d-a1cc-af0575a34ffb
2019-11-01 00:00:10 UTC view    5837166 1783999064103190764         pnb     22.22   556138645       57ed222e-a54a-4907-9944-5a875c2d7f4f
2019-11-01 00:00:11 UTC cart    5876812 1487580010100293687         jessnail        3.16    564506666       186c1951-8052-4b37-adce-dd9644b1d5f7
2019-11-01 00:00:24 UTC remove_from_cart        5826182 1487580007483048900                 3.33    553329724       2067216c-31b5-455d-a1cc-af0575a34ffb
Time taken: 4.093 seconds, Fetched: 5 row(s)
```

```
hive> desc e_commerce_event;
OK
event_time              string                  from deserializer
event_type              string                  from deserializer
product_id              string                  from deserializer
category_id             string                  from deserializer
category_code           string                  from deserializer
brand                   string                  from deserializer
price                   string                  from deserializer
user_id                 string                  from deserializer
user_session            string                  from deserializer
Time taken: 0.07 seconds, Fetched: 9 row(s)
hive>
```

Loading the data into the new table e_comm from the old e_commerce_event table

```
hive> set hive.cli.print.header = true;
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> create External table if not exists e_comm(event_time timestamp ,product_id string ,category_id string ,category_code string ,brand string ,price float ,user_id b
igint ,user_session string)PARTITIONED BY (event_type string) CLUSTERED BY (user_id) into 5 buckets ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' STORED
 AS TEXTFILE ;
OK
Time taken: 0.362 seconds
```

```
hive> desc e_comm;
OK
col_name            data_type           comment
event_time                  string                      from deserializer
product_id                  string                      from deserializer
category_id                 string                      from deserializer
category_code               string                      from deserializer
brand                       string                      from deserializer
price                       string                      from deserializer
user_id                     string                      from deserializer
user_session                string                      from deserializer
event_type                  string

# Partition Information
# col_name                  data_type                   comment

event_type                  string
Time taken: 0.105 seconds, Fetched: 14 row(s)
hive>
```

```
hive> insert into table e_comm partition (event_type) select event_time, product_id ,category_id ,category_code ,brand ,price ,user_id ,user_session ,event_type from e_
commerce_event;
Query ID = root_20221031070924_e64e7aa3-a04f-47e5-98c0-0f843dd3cb31
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_1667194416221_0005)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      5          5        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 99.15 s
--------------------------------------------------------------------------------
Loading data to table demo1.e_comm partition (event_type=null)

Loaded : 4/4 partitions.
        Time taken to load dynamic partitions: 0.57 seconds
        Time taken for adding to write entity : 0.008 seconds
OK
event_time      product_id      category_id     category_code   brand   price   user_id user_session    event_type
Time taken: 111.667 seconds
hive>
```

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

```
hive> select sum(price) from e_commerce_event WHERE Month(event_time)=10 AND event_type='purchase';
Query ID = root_20221031074924_326a5a9d-250a-417f-a2b8-671d04dc4351
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_1667194416221_0008)

--------------------------------------------------------------------------------
        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: 39.36 s
--------------------------------------------------------------------------------
OK
_c0
1211538.4299997438
Time taken: 48.531 seconds, Fetched: 1 row(s)
```

Before partitioning time :- **48.531 seconds**

The total sales in the month of October is **1211538.42999**.

```
hive> select sum(price) from e_comm WHERE Month(event_time)=10 AND event_type='purchase';
Query ID = root_20221031075043_d37ede10-390a-46e1-91c3-089d61d67a80
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0008)

--------------------------------------------------------------------------------
        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: 17.52 s
--------------------------------------------------------------------------------
OK
_c0
1211538.429999907
Time taken: 18.41 seconds, Fetched: 1 row(s)
```

After partitioning time :-**18.41 seconds**

The total sales in the month of October is **1211538.42999**.

```
hive> select sum(price) from e_comm WHERE Month(event_time)=10 AND event_type='purchase';
Query ID = root_20221031080236_259f42d1-72df-49ef-bc5d-fe3bb1b01d4d
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0009)

--------------------------------------------------------------------------------
        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: 15.69 s
--------------------------------------------------------------------------------
OK
_c0
1211538.429999907
Time taken: 16.38 seconds, Fetched: 1 row(s)
```

After **Dynamic partitioning** time :-**16.38 seconds.**

The total sales in the month of October is **1211538.42999**.

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

```
hive> SELECT Month(event_time) as Month, sum(price) as sum, COUNT(event_type)as cnt FROM e_comm WHERE event_type = 'purchase' GROUP BY Month(event_time);
Query ID = root_20221031072358_125acb86-3ebd-4fbb-97a1-7908d1cef97a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0006)

--------------------------------------------------------------------------------
        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: 17.28 s
--------------------------------------------------------------------------------
OK
month   sum     cnt
10      1211538.429999907       245624
11      1531016.899999943       322417
Time taken: 18.101 seconds, Fetched: 2 row(s)
hive>
```

In the month of **October** the total purchases are **245624** and sales is **1211538.42999**

In the month of **November** the total purchases are **322417** and sales is **1531016.89999**

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

```
hive> WITH CTE1 as (SELECT sum(case when Month(event_time)='10' then price else 0 end) as Oct,sum(case when Month(event_time)='11' then price else 0 end) as Nov from e_
comm WHERE event_type ='purchase' and Month(event_time) in (10,11)) SELECT Oct,Nov,(Nov-Oct) as diff from CTE1;
Query ID = root_20221031074016_478a92fd-3ebf-4019-a3d3-e450dc688971
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_1667194416221_0007)

--------------------------------------------------------------------------------
        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: 18.85 s
--------------------------------------------------------------------------------
OK
oct     nov     diff
1211538.429999907       1531016.899999943       319478.47000003606
Time taken: 28.158 seconds, Fetched: 1 row(s)
hive>
```

We can see that the difference in the revenue is **319478.47000.**

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

```
hive> SELECT category_code FROM e_comm WHERE (category_code is not null) GROUP BY category_code;
Query ID = root_20221031083824_ad4d664d-72b6-4f29-9f27-21b7595f0762
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0011)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      6          6        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      5          5        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 42.11 s
----------------------------------------------------------------------------------------------
OK
category_code

accessories.cosmetic_bag
stationery.cartrige
accessories.bag
appliances.environment.vacuum
furniture.living_room.chair
sport.diving
appliances.personal.hair_cutter
appliances.environment.air_conditioner
apparel.glove
furniture.bathroom.bath
furniture.living_room.cabinet
Time taken: 42.699 seconds, Fetched: 12 row(s)
```

The distinct categories are Furniture, Appliances, Accessories, Apparel, sport, and stationery.

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

```
hive> Select category_code as category,count(product_id) as products from e_comm where category_code is not null group by category_code;
Query ID = root_20221031085403_c9a18063-bd25-46f9-8d0d-382465799515
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0012)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      6          6        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      5          5        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 37.66 s
----------------------------------------------------------------------------------------------
OK
category        products
        8594895
accessories.cosmetic_bag        1248
stationery.cartrige     26722
accessories.bag 11681
appliances.environment.vacuum   59761
furniture.living_room.chair     308
sport.diving    2
appliances.personal.hair_cutter 1643
appliances.environment.air_conditioner  332
apparel.glove   18232
furniture.bathroom.bath 9857
furniture.living_room.cabinet   13439
Time taken: 38.387 seconds, Fetched: 12 row(s)
```

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

```
hive> SELECT brand, sum(price) AS totalsales FROM e_comm WHERE brand <>'' AND event_type = 'purchase' GROUP BY brand ORDER BY totalsales desc limit 1;
Query ID = root_20221031090527_527197c7-05f1-4786-89d3-79f17ca34f2f
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1667194416221_0012)

--------------------------------------------------------------------------------
        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
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 15.26 s
--------------------------------------------------------------------------------
OK
brand   totalsales
runail  148297.93999999992
Time taken: 16.056 seconds, Fetched: 1 row(s)
```

Runail is the brand with the maximum sales for October and November combined with total **148297.9399**

Q7. Which brands increased their sales from October to November?

```
hive> WITH CTE2 as (SELECT brand,sum(case when Month(event_time)='10' then price else 0 end) as Oct,sum(case when Month(event_time)='11' then price else 0 end) as     N
ov from e_comm WHERE event_type ='purchase' GROUP BY brand) SELECT brand ,Oct,Nov,(Nov-Oct) as diff from CTE2 WHERE (Nov-Oct)>0 ORDER BY diff;
Query ID = root_20221031095156_5cd1ba3f-6a67-4b97-844d-0b2520f5a892
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_1667207948182_0007)

--------------------------------------------------------------------------------
        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
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 19.63 s
--------------------------------------------------------------------------------
OK
brand   oct     nov     diff
ovale   5.08    6.2     1.12
cosima  40.459999999999994      41.85999999999999       1.3999999999999986
grace   201.83999999999995      205.21999999999997      3.380000000000024
helloganic      0.0     6.2     6.2
skinity 17.76   24.880000000000003      7.120000000000001
bodyton 2752.6800000000003      2761.280000000001       8.600000000000819
moyou   11.42   20.560000000000002      9.140000000000002
neoleor 86.82   103.4   16.580000000000013
soleo   408.3999999999992       425.05999999999926      16.660000000000082
jaguar  2204.2200000000003      2221.3  17.079999999999927
tertio  472.3200000000001       491.6   19.279999999999916
fly     34.28   54.34   20.060000000000002
rasyan  37.599999999999994      57.87999999999998       20.279999999999987
deoproce        633.68  658.34  24.660000000000082
barbie  0.0     24.78   24.78
supertan        100.74  133.02000000000004      32.280000000000044
treaclemoon     326.74  362.9799999999999       36.239999999999895
kamill  126.02000000000001      162.98  36.95999999999998
juno    0.0     42.16   42.16
veraclara       100.21999999999998      142.42  42.2
glysolid        139.45999999999995      183.17999999999998      43.72000000000003
godefroy        802.44  850.24  47.799999999999955
binacil 0.0     48.52   48.52
```

```
de.lux  3319.3999999999787      5551.019999999948       2231.619999999969
swarovski       3775.8599999999824      6086.319999999985       2310.4600000000028
beauty-free     1108.3400000000001      3565.720000000001       2457.380000000001
zeitun  1417.3200000000006      4019.260000000002       2601.9400000000014
joico   1411.04 4030.2000000000007      2619.1600000000008
severina        9551.759999999955       12240.959999999943      2689.199999999988
irisk   91183.92000000093       93892.0800000005        2708.159999999567
oniq    16850.820000000007      19683.30000000002       2832.480000000014
levrana 4487.119999999999       7328.200000000008       2841.080000000009
roubloff        6982.720000000003       9827.540000000006       2844.8200000000033
smart   8914.51999999999        11804.279999999988      2889.7599999999984
shik    6682.400000000004       9679.440000000008       2997.0400000000036
domix   20944.09999999993       24018.339999999895      3074.239999999965
artex   5461.2800000000025      8654.499999999993       3193.2199999999903
beautix 20987.899999999976      24445.899999999994      3458.000000000018
milv    7809.8799999999555      11284.020000000055      3474.1400000000995
masura  62532.15999999802       66116.93999999891       3584.780000000894
f.o.x   13248.459999999992      17154.559999999958      3906.099999999966
kapous  23854.32000000002       28186.160000000047      4331.840000000026
concept 22064.279999999802      26760.799999999745      4696.519999999942
estel   43513.5000000001        48285.34000000022       4771.84000000012
kaypro  1762.6799999999998      6537.4  4774.719999999999
benovy  819.24  6519.939999999997        5700.699999999997
italwax 43880.47999999981       49598.73999999989       5718.260000000082
yoko    17513.81999999997       23415.759999999915      5901.939999999944
haruyama        18781.380000000056      24705.820000000313      5924.440000000257
marathon        14561.500000000005      20546.199999999997      5984.699999999992
lovely  17408.759999999947      23878.119999999923      6469.359999999975
bpw.style       23144.300000000752      29674.880000001613      6530.58000000086
staleks 17039.46000000002       23751.22000000003       6711.760000000009
freedecor       6843.5599999999995      15343.600000000013      8500.040000000014
runail  143078.56000000093      153517.31999999797      10438.75999999704
polarus 12027.440000000002      22743.86000000001       10716.42000000001
cosmoprofi      16645.620000000035      29073.98000000017       12428.360000000135
jessnail        52575.68000000047       66690.45999999983       14114.779999999359
strong  58393.26000000001       77342.53999999992       18949.27999999991
ingarden        46322.78000000045       67132.41999999931       20809.639999998864
lianail 11785.679999999958      32788.480000000214      21002.800000000258
uno     70604.06        102079.49999999997      31475.439999999973
grattol 70891.08000000182       142945.41999999993      72054.3399999981
        949358.1200000312       1239018.4799999818      289660.3599999506
Time taken: 29.45 seconds, Fetched: 161 row(s)
hive>
```

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 Totalpurchases from ecom_sales where event_type='purchase'group by user_id order by Totalpurchases DESC limit 10;
Query ID = hadoop_20221027080502_2fee6b31-6a80-4cc7-9e1d-7943aeea4560
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_1666856056197_0003)

--------------------------------------------------------------------------------
        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
Reducer 3 ...... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 16.72 s
--------------------------------------------------------------------------------
OK
user_id totalpurchases
557790271       2715.869999999991
150318419       1645.9700000000005
562167663       1352.85
531900924       1329.45
557850743       1295.4800000000005
522130011       1185.3899999999999
561592095       1109.7000000000005
431950134       1097.5899999999997
566576008       1056.3600000000004
521347209       1040.9099999999999
Time taken: 27.353 seconds, Fetched: 10 row(s)
hive>
```

Cleaning up

Dropping database demo1

```
hive> show databases;
OK
default
demo
demo1
Time taken: 0.247 seconds, Fetched: 3 row(s)
hive> drop database demo1 cascade;
OK
Time taken: 0.706 seconds
hive> show databases;
OK
default
demo
Time taken: 0.015 seconds, Fetched: 2 row(s)
hive>
```

Terminating the cluster

			Name	ID	Status	Creation
☐	▶		Hive case study	j-3U2NSYRPTN4YH	Terminated User request	2022-10
☐	▶	●	Hive case study	j-38APERQ3JSHYA	Terminated with errors Instance failure	2022-10