CIS5200 Term Project Tutorial

Authors: Vamsi Sai Krishna Reddy Yarramreddy; Prathima Sarvani Alla; Natakarani,Abhiram; Esquivias, Jeremy O

Instructor: Jongwook Woo

Date: 05/10/2023

Lab Tutorial

Vamsi Sai Krishna Reddy Yarramreddy (vyarram@calstatela.edu)

Prathima Sarvani Alla(palla3@calstatela.edu)

Natakarani,Abhiram (anataka@calstatela.edu)

Esquivias, Jeremy O (jesqui@calstatela.edu)

Ecommerce Behavior data

Objectives

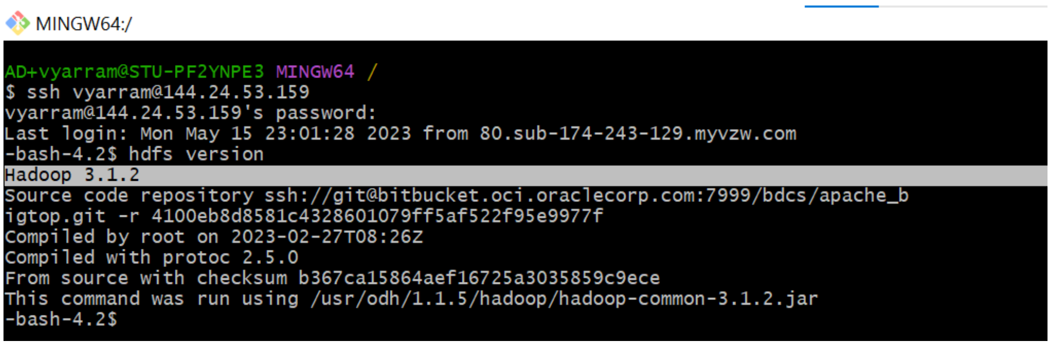
In this hands-on lab, you will learn how to:

1. Download dataset from the source site to the local file directory
2. Upload the dataset from the local file directory to Oracle DBCS using SCP in Git Bash
3. Upload from Oracle DBCS to Hadoop Distributed File System (HDFS)
4. Create table and queries in HiveQL to analyze the dataset
5. Create visualization using Power BI, Excel

Platform Spec

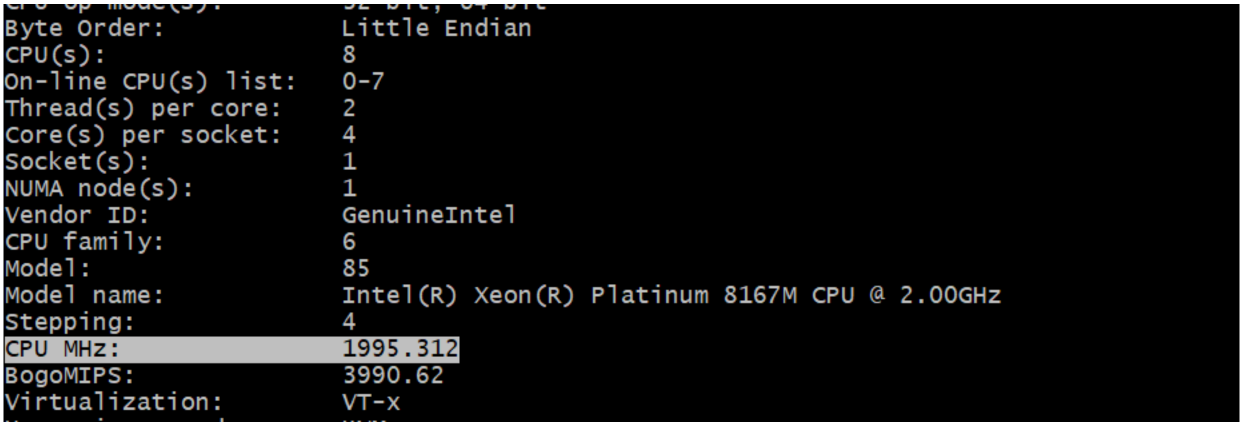
- Cluster version: Hadoop 3.1.2
- CPU Speed: 1995.312 MHz
- # of CPU cores: 8
- # of nodes: 3
- Total Memory Size: 390.7 G

```
-bash-4.2$ hdfs version
```



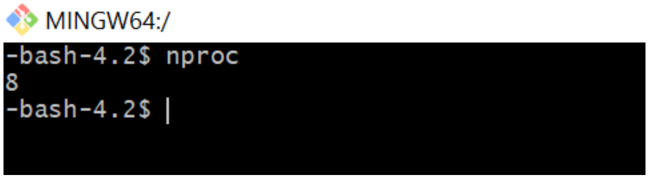
```
AD+vyarram@STU-PF2YNPE3 MINGW64 /
$ ssh vyarram@144.24.53.159
vyarram@144.24.53.159's password:
Last login: Mon May 15 23:01:28 2023 from 80.sub-174-243-129.myvzw.com
-bash-4.2$ hdfs version
Hadoop 3.1.2
Source code repository ssh://git@bitbucket.oci.oraclecorp.com:7999/bdcs/apache_b
igtop.git -r 4100eb8d8581c4328601079ff5af522f95e9977f
Compiled by root on 2023-02-27T08:26Z
Compiled with protoc 2.5.0
From source with checksum b367ca15864aef16725a3035859c9ece
This command was run using /usr/odh/1.1.5/hadoop/hadoop-common-3.1.2.jar
-bash-4.2$
```

```
-bash-4.2$ lscpu
```



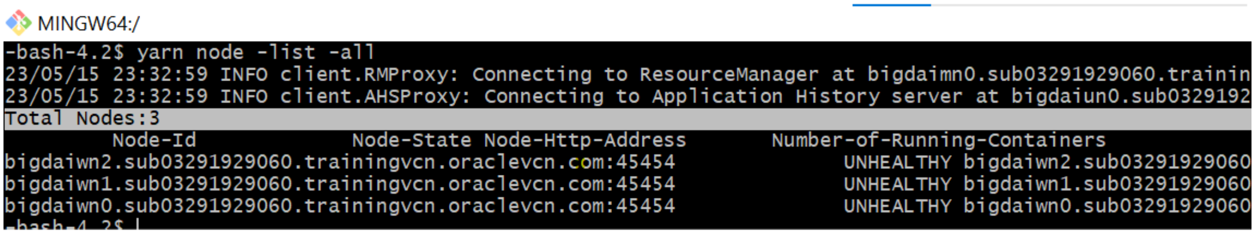
```
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                8
On-line CPU(s) list:   0-7
Thread(s) per core:    2
Core(s) per socket:    4
Socket(s):             1
NUMA node(s):          1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 85
Model name:            Intel(R) Xeon(R) Platinum 8167M CPU @ 2.00GHz
Stepping:              4
CPU MHz:               1995.312
BogoMIPS:              3990.62
Virtualization:        VT-x
```

```
-bash-4.2$ nproc
```



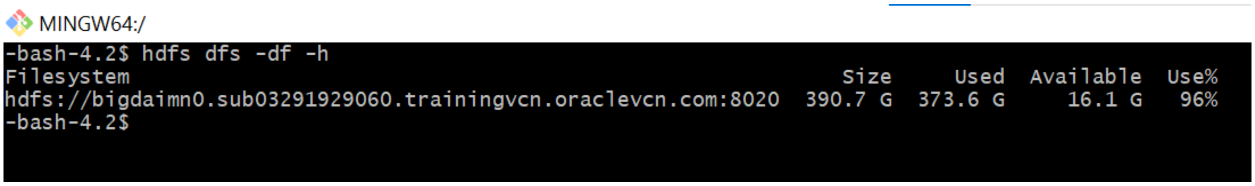
```
-bash-4.2$ nproc
8
-bash-4.2$ |
```

```
-bash-4.2$ yarn node -list -all
```



```
-bash-4.2$ yarn node -list -all
23/05/15 23:32:59 INFO client.RMProxy: Connecting to ResourceManager at bigdaimn0.sub03291929060.trainin
23/05/15 23:32:59 INFO client.AHSProxy: Connecting to Application History server at bigdaiun0.sub0329192
Total Nodes:3
         Node-Id             Node-State Node-Http-Address       Number-of-Running-Containers
bigdaiwn2.sub03291929060.trainingvcn.oraclevcn.com:45454              UNHEALTHY bigdaiwn2.sub03291929060
bigdaiwn1.sub03291929060.trainingvcn.oraclevcn.com:45454              UNHEALTHY bigdaiwn1.sub03291929060
bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com:45454              UNHEALTHY bigdaiwn0.sub03291929060
-bash-4.2$
```

```
-bash-4.2$ hdfs dfs -df -h
```



```
-bash-4.2$ hdfs dfs -df -h
Filesystem                                                   Size     Used  Available  Use%
hdfs://bigdaimn0.sub03291929060.trainingvcn.oraclevcn.com:8020  390.7 G  373.6 G     16.1 G   96%
-bash-4.2$
```

Dataset detail:

Dataset name: 2019-Oct
Dataset Source URL: https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?resource=download
Data set Size: 5.67 GB

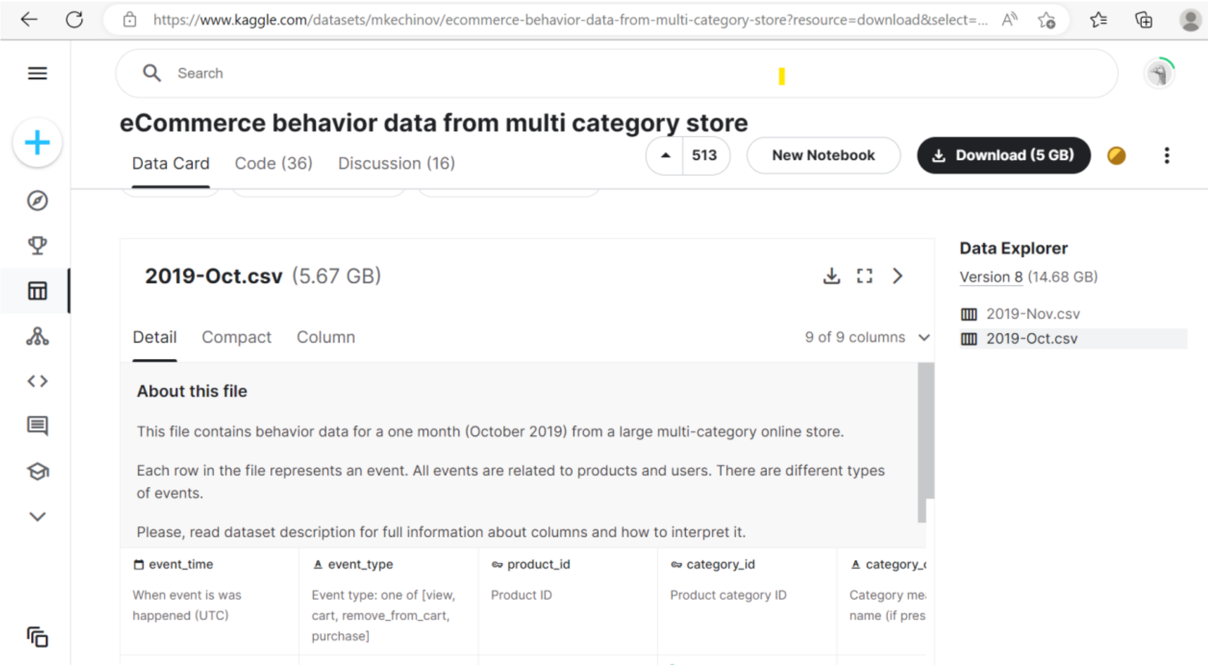
Step 1: Download data file from source website

The dataset we are using for this project is from the website, Kaggle. The first step is to download the dataset from Kaggle's website to your local computer directory.

1. Click on the link below to go to Kaggle to view the dataset.
https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?resource=download

2. The page below will appear. Scroll down to find the red circle area, click on the arrow to download the dataset to your local computer directory. The file will be downloaded as a zip file. Locate the file in your local computer directory, it is usually in C:\Users\vyarram\Downloads.

Note: Should have a Kaggle account to download Data Set


https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store?resource=download&select=...
Search

eCommerce behavior data from multi category store

2019-Oct.csv (5.67 GB)

About this file

This file contains behavior data for a one month (October 2019) from a large multi-category online store.

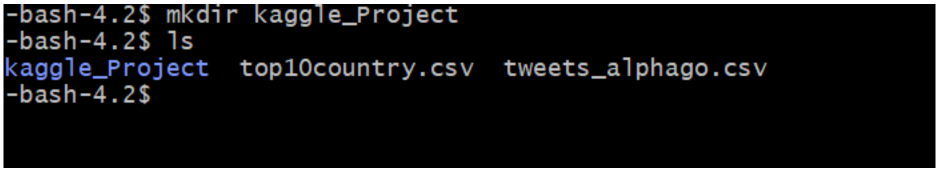
Each row in the file represents an event. All events are related to products and users. There are different types of events.

Please, read dataset description for full information about columns and how to interpret it.

event_time	event_type	product_id	category_id	category_
When event is was happened (UTC)	Event type: one of [view, cart, remove_from_cart, purchase]	Product ID	Product category ID	Category me name (if pres

Step 2: Upload the Dataset to Oracle DBCS

- In Oracle DBCS make a directory named kaggle_project

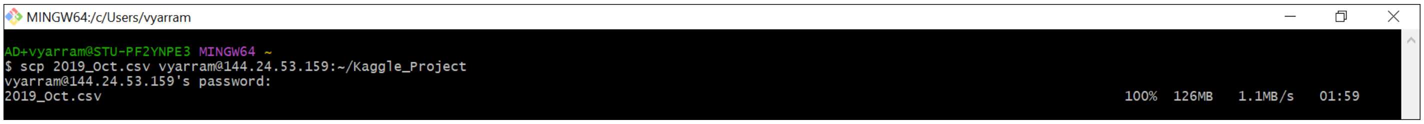


```
-bash-4.2$ mkdir kaggle_Project
-bash-4.2$ ls
kaggle_Project  top10country.csv  tweets_alphago.csv
-bash-4.2$
```

After you locate the file, you just downloaded from Kaggle in your local computer in first step, then upload it Oracle DBCS.

- Open Gitbash, upload the dataset from your local computer directory to Oracle DBCS with SCP.

```
$ scp 2019_Oct.csv vyarram@144.24.53.159:~/kaggle_project
```



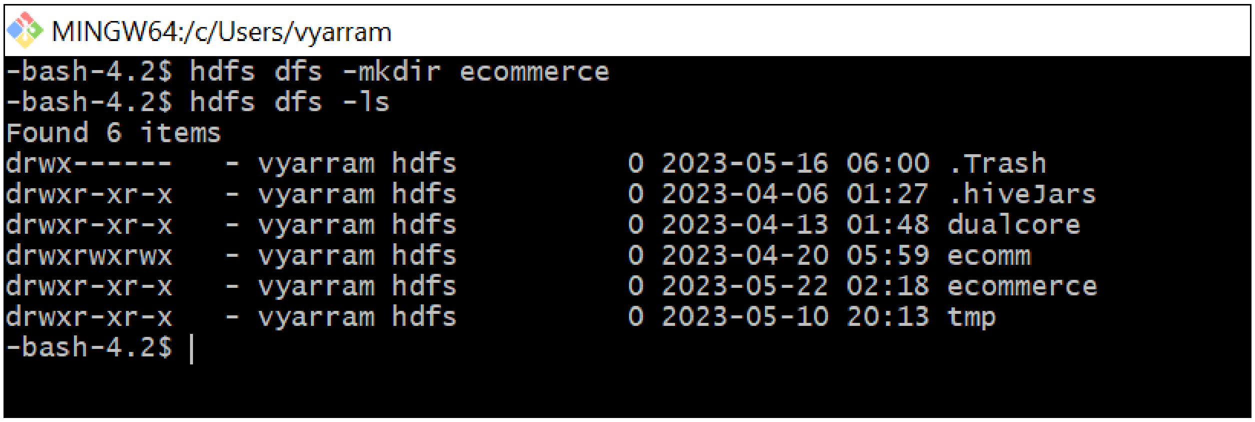
```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ scp 2019_Oct.csv vyarram@144.24.53.159:~/Kaggle_Project
vyarram@144.24.53.159's password:
2019_Oct.csv                                                                                              100%  126MB   1.1MB/s   01:59
```

Step 3: Upload from Oracle DBCS to Hadoop Distributed File System (HDFS)

In this step you will create directory in Hadoop Distributed File System (HDFS), then up upload the dataset from Oracle DBCS to HDFS, then confirm the directory is created.

- Create a directory called ecommerce in HDFS, then ls to check if it is created.

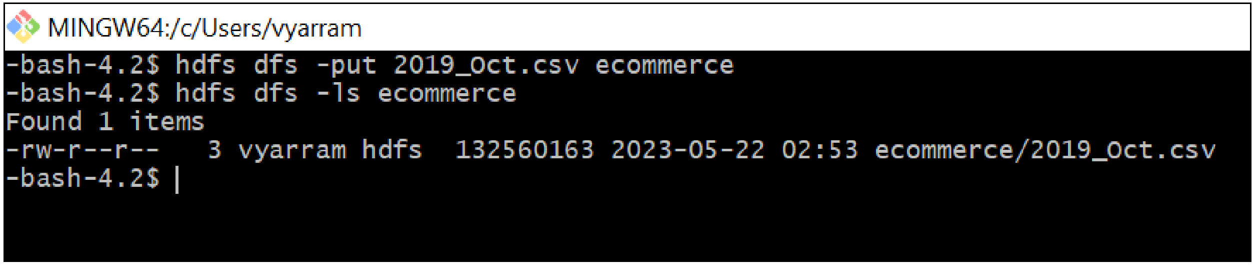
```
$ hdfs dfs -mkdir ecommerce
$ hdfs dfs -ls
```



```
-bash-4.2$ hdfs dfs -mkdir ecommerce
-bash-4.2$ hdfs dfs -ls
Found 6 items
drwx------   - vyarram hdfs          0 2023-05-16 06:00 .Trash
drwxr-xr-x   - vyarram hdfs          0 2023-04-06 01:27 .hiveJars
drwxr-xr-x   - vyarram hdfs          0 2023-04-13 01:48 dualcore
drwxrwxrwx   - vyarram hdfs          0 2023-04-20 05:59 ecomm
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 02:18 ecommerce
drwxr-xr-x   - vyarram hdfs          0 2023-05-10 20:13 tmp
-bash-4.2$ |
```

Upload from Oracle DBCS to Hadoop Distributed File System (HDFS)

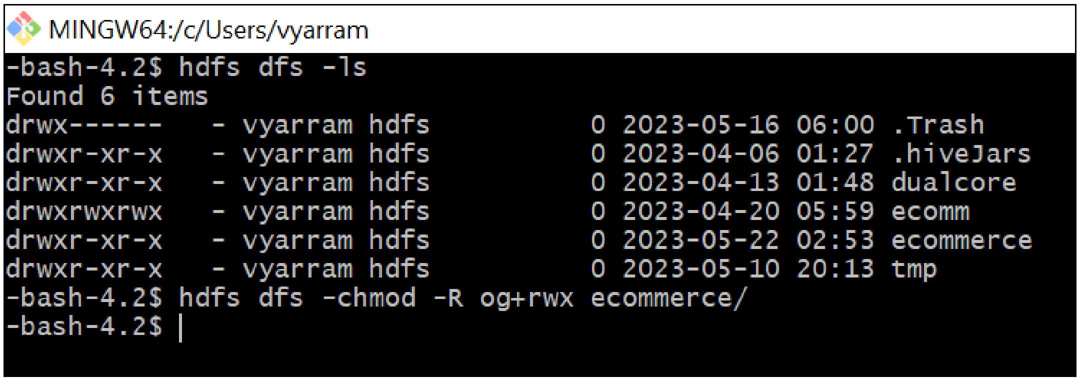
```
$ hdfs dfs -put 2019_Oct.csv ecommerce
$ hdfs dfs -ls ecommerce/
```



```
-bash-4.2$ hdfs dfs -put 2019_Oct.csv ecommerce
-bash-4.2$ hdfs dfs -ls ecommerce
Found 1 items
-rw-r--r--   3 vyarram hdfs  132560163 2023-05-22 02:53 ecommerce/2019_Oct.csv
-bash-4.2$ |
```

- Share the folder with other teammates

```
$ hdfs dfs -chmod -R og+rwx ecommerce/
```



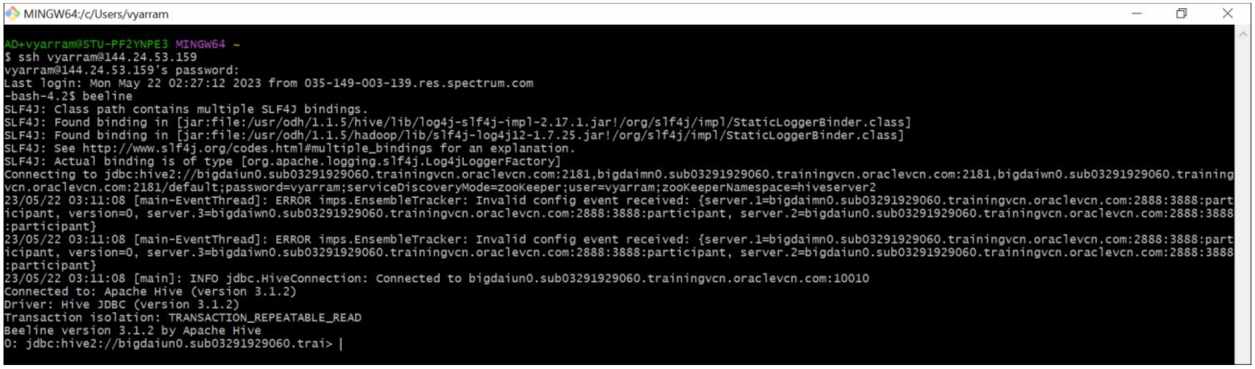
MINGW64:/c/Users/vyarram

```
-bash-4.2$ hdfs dfs -ls
Found 6 items
drwx------   - vyarram hdfs          0 2023-05-16 06:00 .Trash
drwxr-xr-x   - vyarram hdfs          0 2023-04-06 01:27 .hiveJars
drwxr-xr-x   - vyarram hdfs          0 2023-04-13 01:48 dualcore
drwxrwxrwx   - vyarram hdfs          0 2023-04-20 05:59 ecomm
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 02:53 ecommerce
drwxr-xr-x   - vyarram hdfs          0 2023-05-10 20:13 tmp
-bash-4.2$ hdfs dfs -chmod -R og+rwx ecommerce/
-bash-4.2$ |
```

Step 4: Create the initial table HiveQL

- Open a new bash CLI, ssh into Oracle DBCS, then type in beeline to enter into HiveQL.

```
$ ssh vyarram@144.24.53.159
$ beeline
```

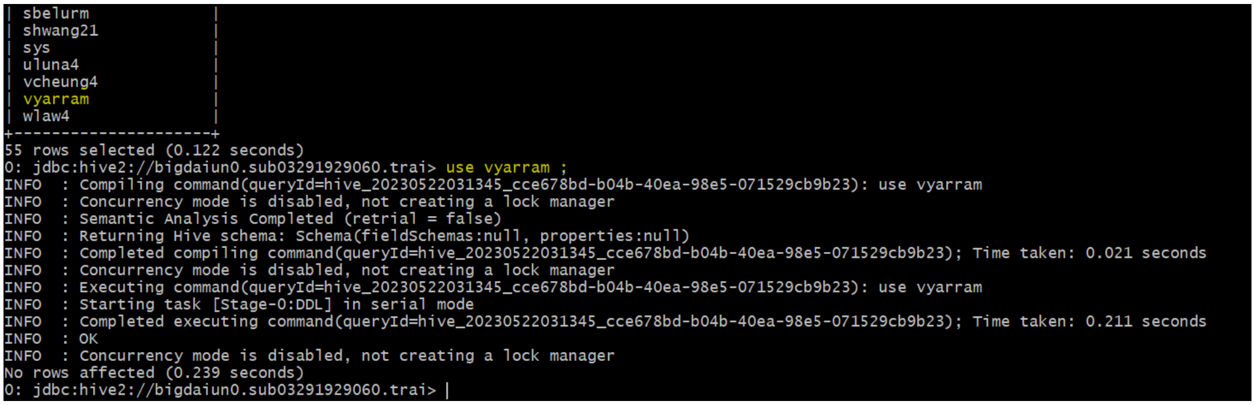


MINGW64:/c/Users/vyarram

```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ ssh vyarram@144.24.53.159
vyarram@144.24.53.159's password:
Last login: Mon May 22 02:27:12 2023 from 035-149-003-139.res.spectrum.com
-bash-4.2$ beeline
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/odh/1.1.5/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/odh/1.1.5/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://bigdaiun0.sub03291929060.trainingvcn.oraclevcn.com:2181,bigdaimn0.sub03291929060.trainingvcn.oraclevcn.com:2181,bigdaiwn0.sub03291929060.training
vcn.oraclevcn.com:2181/default;password=vyarram;serviceDiscoveryMode=zooKeeper;user=vyarram;zooKeeperNamespace=hiveserver2
23/05/22 03:11:08 [main-EventThread]: ERROR imps.EnsembleTracker: Invalid config event received: {server.1=bigdaimn0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888:part
icipant, version=0, server.3=bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888:participant, server.2=bigdaiun0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888
:participant}
23/05/22 03:11:08 [main-EventThread]: ERROR imps.EnsembleTracker: Invalid config event received: {server.1=bigdaimn0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888:part
icipant, version=0, server.3=bigdaiwn0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888:participant, server.2=bigdaiun0.sub03291929060.trainingvcn.oraclevcn.com:2888:3888
:participant}
23/05/22 03:11:08 [main]: INFO jdbc.HiveConnection: Connected to bigdaiun0.sub03291929060.trainingvcn.oraclevcn.com:10010
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.2 by Apache Hive
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> |
```

- Create and use your database. If you have an existing database, you can omit creating database.

```
$ show databases ;
$ use vyarram;
```



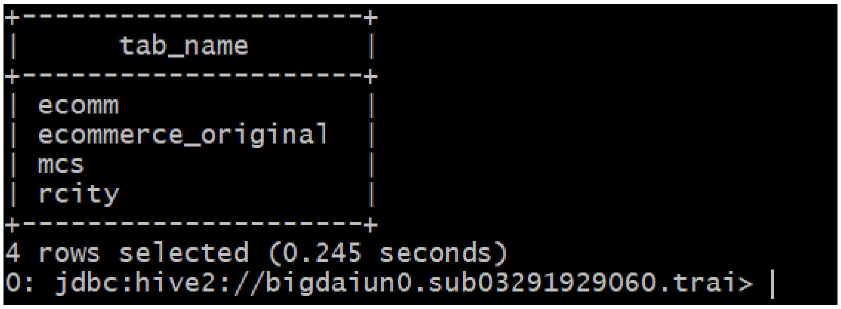
```
| sbelurm              |
| shwang21             |
| sys                  |
| uluna4               |
| vcheung4             |
| vyarram              |
| wlaw4                |
+----------------------+
55 rows selected (0.122 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> use vyarram ;
INFO  : Compiling command(queryId=hive_20230522031345_cce678bd-b04b-40ea-98e5-071529cb9b23): use vyarram
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230522031345_cce678bd-b04b-40ea-98e5-071529cb9b23); Time taken: 0.021 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522031345_cce678bd-b04b-40ea-98e5-071529cb9b23): use vyarram
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230522031345_cce678bd-b04b-40ea-98e5-071529cb9b23); Time taken: 0.211 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.239 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> |
```

- Create an external table named ecommerce_original in HiveQL.

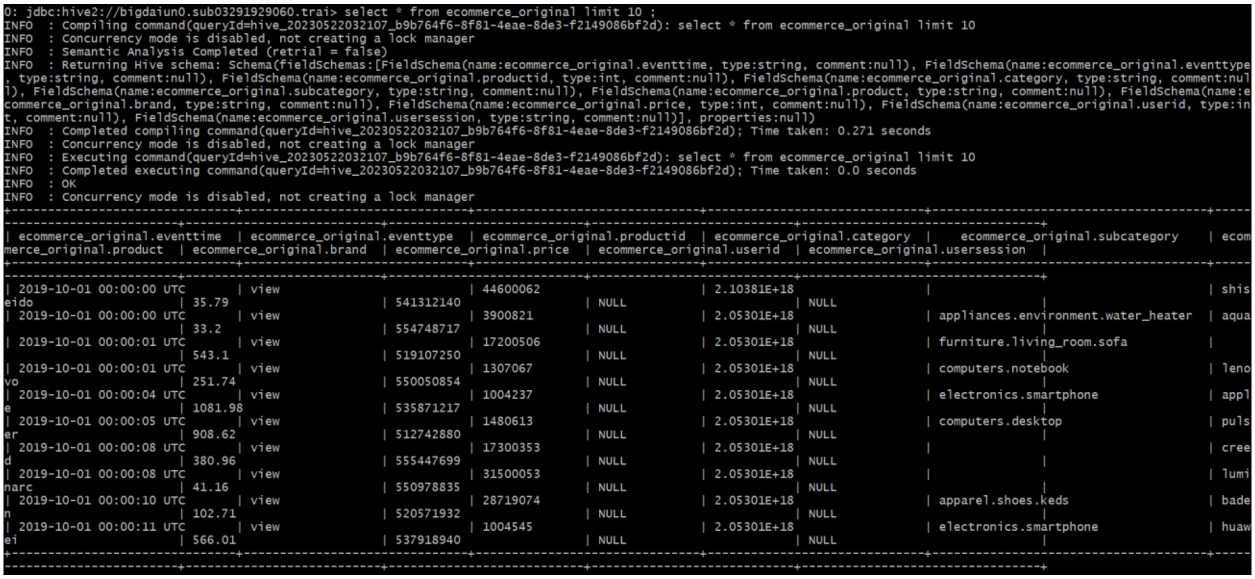
```
DROP TABLE IF EXISTS ecommerce_original;

CREATE EXTERNAL TABLE IF NOT EXISTS ecommerce_original(
eventtime STRING,
eventtype STRING,
productid INT,
category STRING,
subcategory STRING,
product STRING,
brand STRING,
price INT,
userid INT,
usersession STRING
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/'
TBLPROPERTIES ('skip.header.line.count'='1');
```

- Verify if the table is created, then check if there are values in the table ecommerce_original

```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> show tables ;
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> select * from ecommerce_original limit 10 ;
```



```
+----------------------+
|       tab_name       |
+----------------------+
| ecomm                |
| ecommerce_original   |
| mcs                  |
| rcity                |
+----------------------+
4 rows selected (0.245 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> |
```

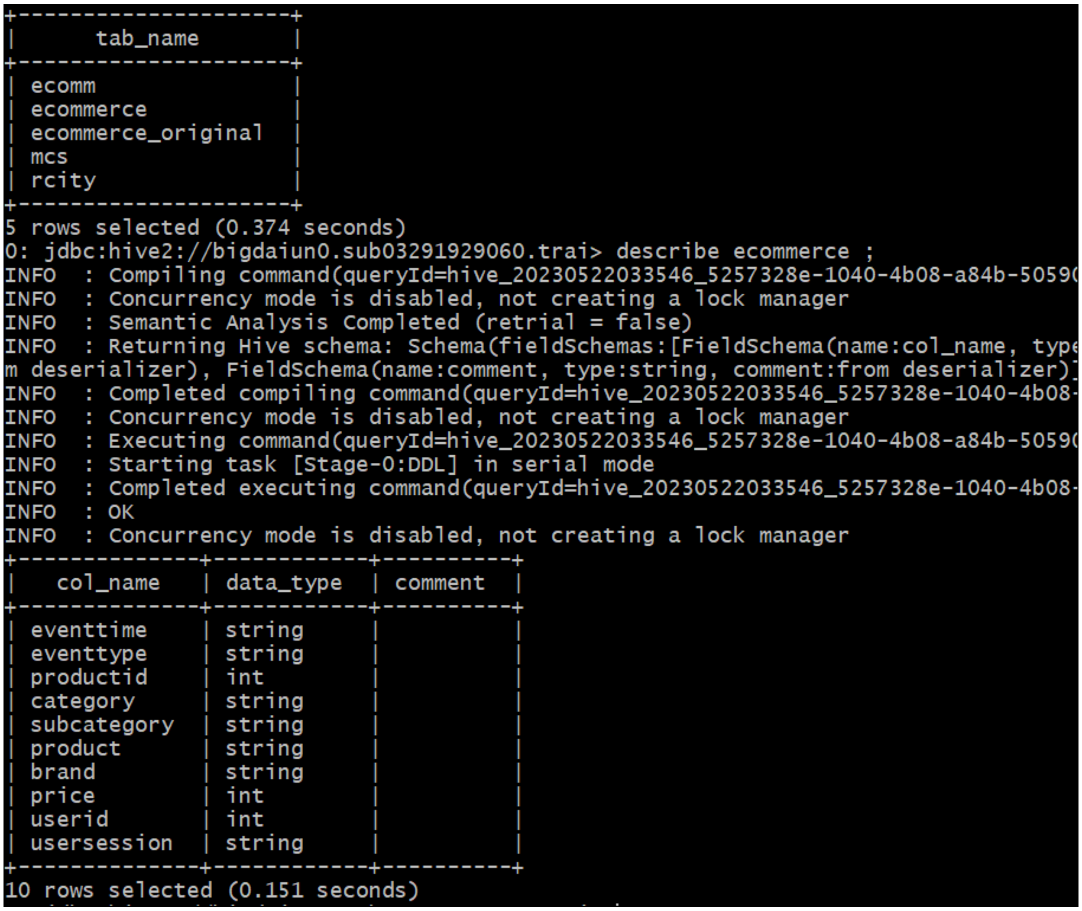


```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> select * from ecommerce_original limit 10 ;
INFO  : Compiling command(queryId=hive_20230522032107_b9b764f6-8f81-4eae-8de3-f2149086bf2d): select * from ecommerce_original limit 10
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:ecommerce_original.eventtime, type:string, comment:null), FieldSchema(name:ecommerce_original.eventtype
, type:string, comment:null), FieldSchema(name:ecommerce_original.productid, type:int, comment:null), FieldSchema(name:ecommerce_original.category, type:string, comment:nul
l), FieldSchema(name:ecommerce_original.subcategory, type:string, comment:null), FieldSchema(name:ecommerce_original.product, type:string, comment:null), FieldSchema(name:e
commerce_original.brand, type:string, comment:null), FieldSchema(name:ecommerce_original.price, type:int, comment:null), FieldSchema(name:ecommerce_original.userid, type:in
t, comment:null), FieldSchema(name:ecommerce_original.usersession, type:string, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230522032107_b9b764f6-8f81-4eae-8de3-f2149086bf2d); Time taken: 0.271 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522032107_b9b764f6-8f81-4eae-8de3-f2149086bf2d): select * from ecommerce_original limit 10
INFO  : Completed executing command(queryId=hive_20230522032107_b9b764f6-8f81-4eae-8de3-f2149086bf2d); Time taken: 0.0 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+-------------------------------+-------------------------------+-------------------------------+-------------------------------+--------------------------------------+-----
-----------------------+-----------------------------+-----------------------------+-----------------------------+-------------------------------------+
| ecommerce_original.eventtime  | ecommerce_original.eventtype  | ecommerce_original.productid  | ecommerce_original.category  |    ecommerce_original.subcategory    | ecom
merce_original.product  | ecommerce_original.brand  | ecommerce_original.price  | ecommerce_original.userid  | ecommerce_original.usersession  |
+-------------------------------+-------------------------------+-------------------------------+-------------------------------+--------------------------------------+-----
-----------------------+-----------------------------+-----------------------------+-----------------------------+-------------------------------------+
| 2019-10-01 00:00:00 UTC       | view                          | 44600062                      | 2.10381E+18                   |                                      | shis
eido                   | 35.79                       | 541312140                   | NULL                        | NULL                                |
| 2019-10-01 00:00:00 UTC       | view                          | 3900821                       | 2.05301E+18                   | appliances.environment.water_heater  | aqua
                       | 33.2                        | 554748717                   | NULL                        | NULL                                |
| 2019-10-01 00:00:01 UTC       | view                          | 17200506                      | 2.05301E+18                   | furniture.living_room.sofa           |
                       | 543.1                       | 519107250                   | NULL                        | NULL                                |
| 2019-10-01 00:00:01 UTC       | view                          | 1307067                       | 2.05301E+18                   | computers.notebook                   | leno
vo                     | 251.74                      | 550050854                   | NULL                        | NULL                                |
| 2019-10-01 00:00:04 UTC       | view                          | 1004237                       | 2.05301E+18                   | electronics.smartphone               | appl
e                      | 1081.98                     | 535871217                   | NULL                        | NULL                                |
| 2019-10-01 00:00:05 UTC       | view                          | 1480613                       | 2.05301E+18                   | computers.desktop                    | puls
er                     | 908.62                      | 512742880                   | NULL                        | NULL                                |
| 2019-10-01 00:00:08 UTC       | view                          | 17300353                      | 2.05301E+18                   |                                      | cree
d                      | 380.96                      | 555447699                   | NULL                        | NULL                                |
| 2019-10-01 00:00:08 UTC       | view                          | 31500053                      | 2.05301E+18                   |                                      | lumi
narc                   | 41.16                       | 550978835                   | NULL                        | NULL                                |
| 2019-10-01 00:00:10 UTC       | view                          | 28719074                      | 2.05301E+18                   | apparel.shoes.keds                   | bade
n                      | 102.71                      | 520571932                   | NULL                        | NULL                                |
| 2019-10-01 00:00:11 UTC       | view                          | 1004545                       | 2.05301E+18                   | electronics.smartphone               | huaw
ei                     | 566.01                      | 537918940                   | NULL                        | NULL                                |
+-------------------------------+-------------------------------+-------------------------------+-------------------------------+--------------------------------------+-----
-----------------------+-----------------------------+-----------------------------+-----------------------------+-------------------------------------+
```

- STEP 5: Data Cleaning

Removed all the Null which are present in the Category Code and Brand column

```
Create table ecommerce as select *  from ecommerce_original where (category != '' or brand
!= '') and  (category != '' and  brand != '') ;
show tables ;
 describe ecommerce ;
```

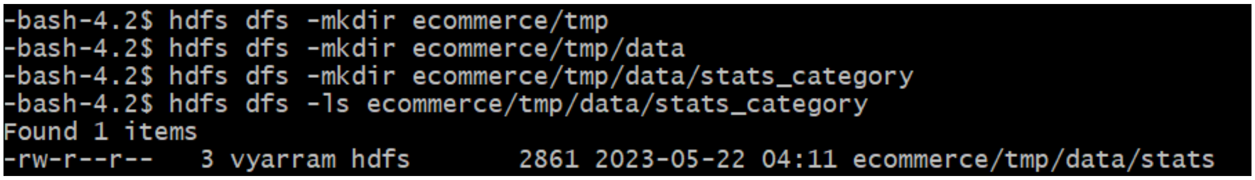


```
+----------------------+
|       tab_name       |
+----------------------+
| ecomm                |
| ecommerce            |
| ecommerce_original   |
| mcs                  |
| rcity                |
+----------------------+
5 rows selected (0.374 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> describe ecommerce ;
INFO  : Compiling command(queryId=hive_20230522033546_5257328e-1040-4b08-a84b-5059
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type
m deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)]
INFO  : Completed compiling command(queryId=hive_20230522033546_5257328e-1040-4b08-
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522033546_5257328e-1040-4b08-a84b-5059
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230522033546_5257328e-1040-4b08-
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+--------------+------------+----------+
|   col_name   | data_type  | comment  |
+--------------+------------+----------+
| eventtime    | string     |          |
| eventtype    | string     |          |
| productid    | int        |          |
| category     | string     |          |
| subcategory  | string     |          |
| product      | string     |          |
| brand        | string     |          |
| price        | int        |          |
| userid       | int        |          |
| usersession  | string     |          |
+--------------+------------+----------+
10 rows selected (0.151 seconds)
```

Step 6: Creating HiveQL

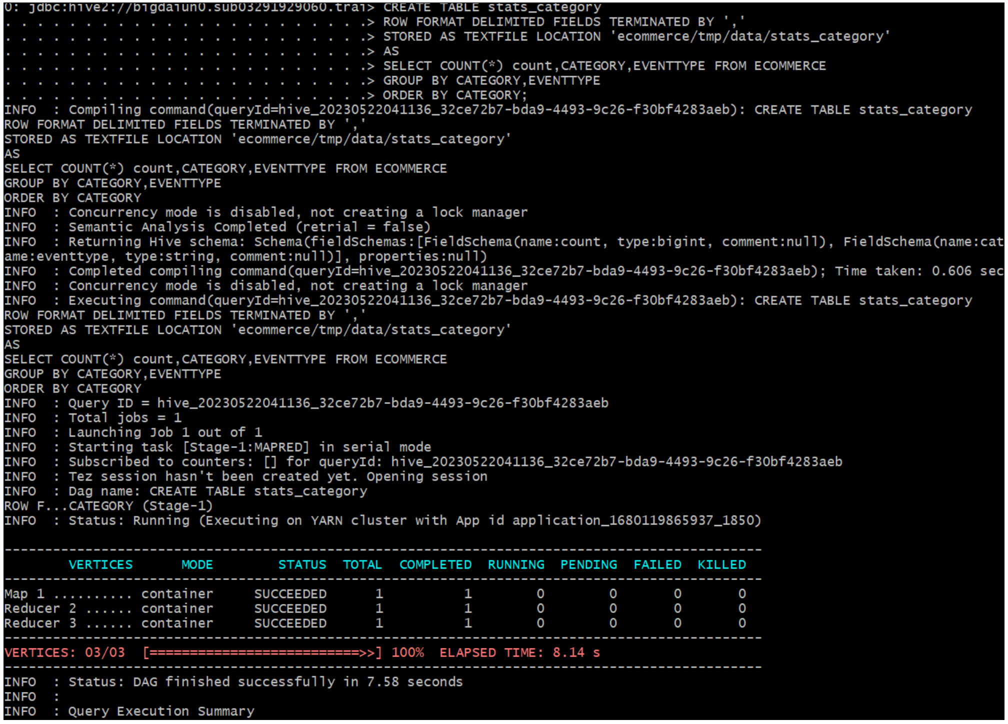
- Count of Event Type in Each `Category

Make stats_category Directory in Hadoop File system
Hdfs dfs -mkdir ecommerce/tmp
Hdfs dfs -mkdir ecommerce/tmp/data
Hdfs dfs -mkdir ecommerce/tmp/data/stats_category

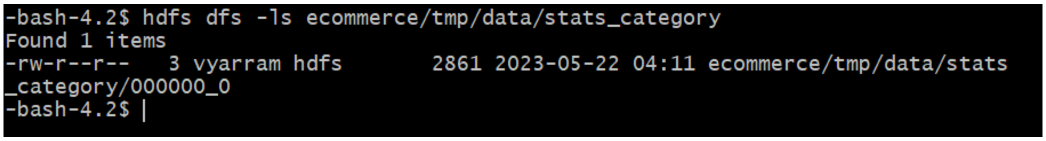


```
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp/data
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp/data/stats_category
-bash-4.2$ hdfs dfs -ls ecommerce/tmp/data/stats_category
Found 1 items
-rw-r--r--   3 vyarram hdfs       2861 2023-05-22 04:11 ecommerce/tmp/data/stats
```

```
CREATE TABLE STATS_CATEGORY
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/stats_category'
AS
SELECT COUNT (*) COUNT, CATEGORY, EVENTTYPE FROM ECOMMERCE
GROUP BY CATEGORY, EVENTTYPE
ORDER BY CATEGORY.
```

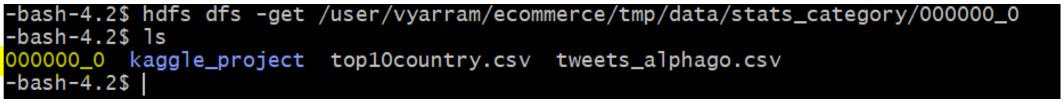


```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE TABLE stats_category
. . . . . . . . . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . . . . . . . . . . . . . . . . . . . .> STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/stats_category'
. . . . . . . . . . . . . . . . . . . . . . .> AS
. . . . . . . . . . . . . . . . . . . . . . .> SELECT COUNT(*) count,CATEGORY,EVENTTYPE FROM ECOMMERCE
. . . . . . . . . . . . . . . . . . . . . . .> GROUP BY CATEGORY,EVENTTYPE
. . . . . . . . . . . . . . . . . . . . . . .> ORDER BY CATEGORY;
INFO  : Compiling command(queryId=hive_20230522041136_32ce72b7-bda9-4493-9c26-f30bf4283aeb): CREATE TABLE stats_category
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/stats_category'
AS
SELECT COUNT(*) count,CATEGORY,EVENTTYPE FROM ECOMMERCE
GROUP BY CATEGORY,EVENTTYPE
ORDER BY CATEGORY
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:count, type:bigint, comment:null), FieldSchema(name:cat
ame:eventtype, type:string, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230522041136_32ce72b7-bda9-4493-9c26-f30bf4283aeb); Time taken: 0.606 sec
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522041136_32ce72b7-bda9-4493-9c26-f30bf4283aeb): CREATE TABLE stats_category
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/stats_category'
AS
SELECT COUNT(*) count,CATEGORY,EVENTTYPE FROM ECOMMERCE
GROUP BY CATEGORY,EVENTTYPE
ORDER BY CATEGORY
INFO  : Query ID = hive_20230522041136_32ce72b7-bda9-4493-9c26-f30bf4283aeb
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230522041136_32ce72b7-bda9-4493-9c26-f30bf4283aeb
INFO  : Tez session hasn't been created yet. Opening session
INFO  : Dag name: CREATE TABLE stats_category
ROW F...CATEGORY (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1680119865937_1850)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        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: 8.14 s
----------------------------------------------------------------------------------------------
INFO  : Status: DAG finished successfully in 7.58 seconds
INFO  :
INFO  : Query Execution Summary
```



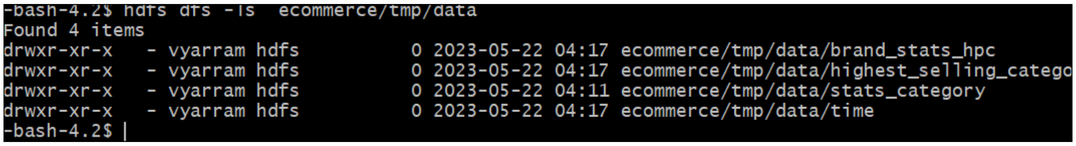
```
-bash-4.2$ hdfs dfs -ls ecommerce/tmp/data/stats_category
Found 1 items
-rw-r--r--   3 vyarram hdfs       2861 2023-05-22 04:11 ecommerce/tmp/data/stats
_category/000000_0
-bash-4.2$ |
```

hdfs dfs -get /user/palla3/ecommerce/tmp/data/stats_category/000000_0



```
-bash-4.2$ hdfs dfs -get /user/vyarram/ecommerce/tmp/data/stats_category/000000_0
-bash-4.2$ ls
000000_0  kaggle_project  top10country.csv  tweets_alphago.csv
-bash-4.2$ |
```

hdfs dfs -ls ecommerce/



```
-bash-4.2$ hdfs dfs -ls  ecommerce/tmp/data
Found 4 items
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 04:17 ecommerce/tmp/data/brand_stats_hpc
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 04:17 ecommerce/tmp/data/highest_selling_catego
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 04:11 ecommerce/tmp/data/stats_category
drwxr-xr-x   - vyarram hdfs          0 2023-05-22 04:17 ecommerce/tmp/data/time
-bash-4.2$ |
```

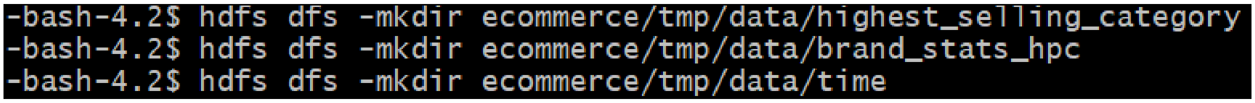
- Creating Table for highly Purchase category

Create more directories in the tmp/data

hdfs dfs -mkdir ecommerce/tmp/data/highest_selling_category

hdfs dfs -mkdir ecommerce/tmp/data/time

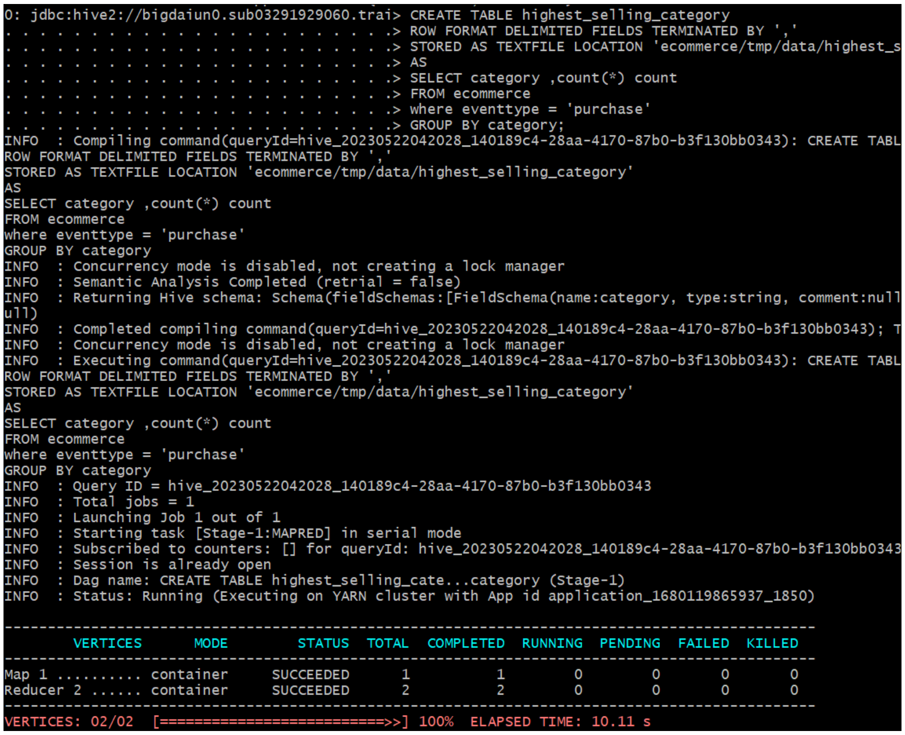
hdfs dfs -mkdir ecommerce/tmp/data/brand_stats_hpc



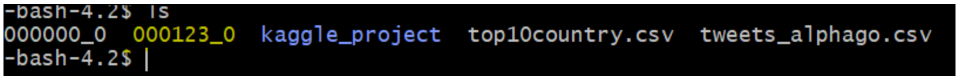
```
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp/data/highest_selling_category
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp/data/brand_stats_hpc
-bash-4.2$ hdfs dfs -mkdir ecommerce/tmp/data/time
```

```
CREATE TABLE highest_selling_category
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/highest_selling_category'
AS
SELECT category, count (*) count
FROM ecommerce
where eventtype = 'purchase'
GROUP BY category.
```

hdfs dfs -getmerge ecommerce/tmp/data/highest_selling_category/ ~/000123_0



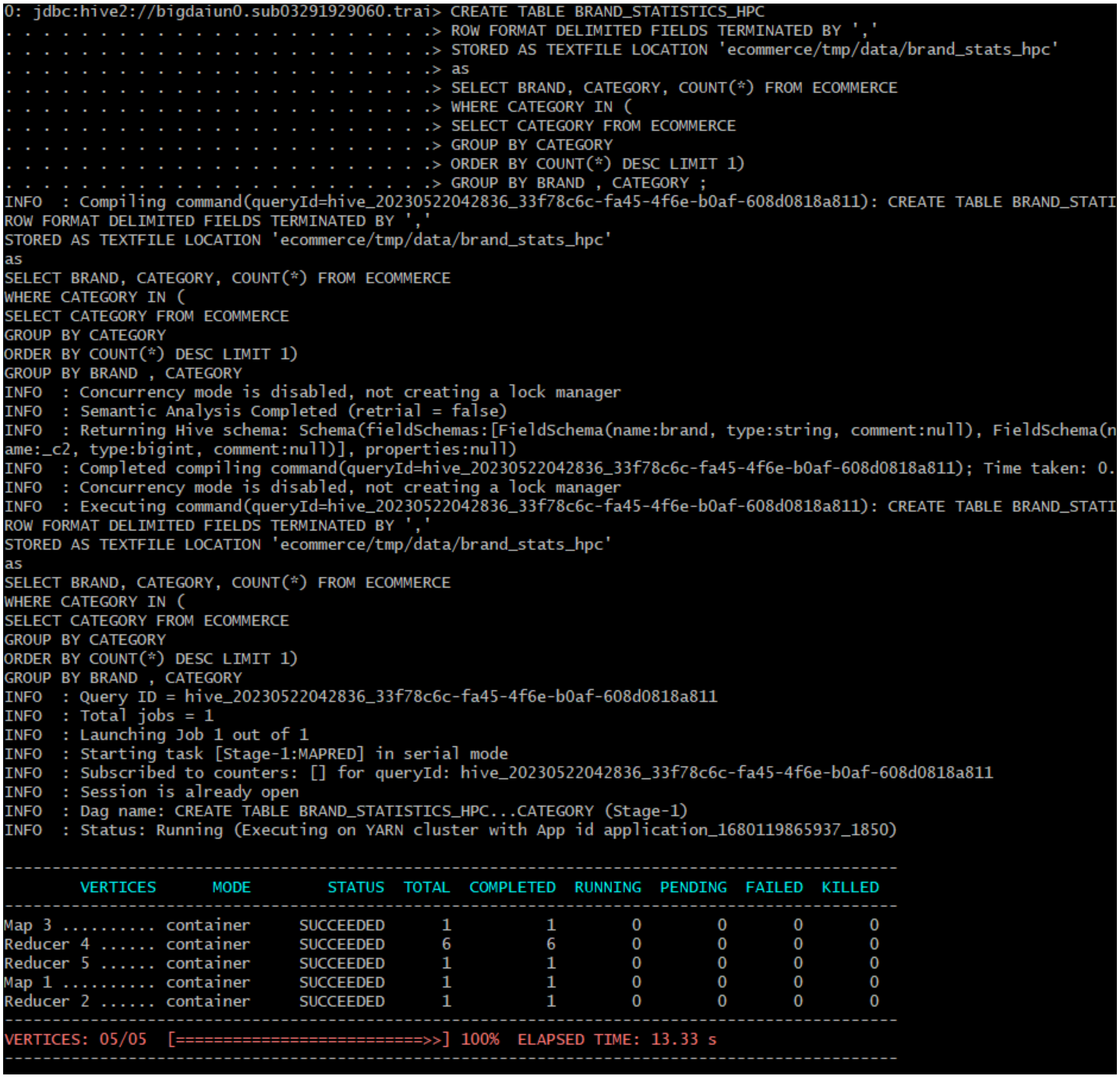
```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE TABLE highest_selling_category
. . . . . . . . . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . . . . . . . . . . . . . . . . . . . .> STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/highest_s
. . . . . . . . . . . . . . . . . . . . . . .> AS
. . . . . . . . . . . . . . . . . . . . . . .> SELECT category ,count(*) count
. . . . . . . . . . . . . . . . . . . . . . .> FROM ecommerce
. . . . . . . . . . . . . . . . . . . . . . .> where eventtype = 'purchase'
. . . . . . . . . . . . . . . . . . . . . . .> GROUP BY category;
INFO  : Compiling command(queryId=hive_20230522042028_140189c4-28aa-4170-87b0-b3f130bb0343): CREATE TABL
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/highest_selling_category'
AS
SELECT category ,count(*) count
FROM ecommerce
where eventtype = 'purchase'
GROUP BY category
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:category, type:string, comment:null
ull)
INFO  : Completed compiling command(queryId=hive_20230522042028_140189c4-28aa-4170-87b0-b3f130bb0343); T
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522042028_140189c4-28aa-4170-87b0-b3f130bb0343): CREATE TABL
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/highest_selling_category'
AS
SELECT category ,count(*) count
FROM ecommerce
where eventtype = 'purchase'
GROUP BY category
INFO  : Query ID = hive_20230522042028_140189c4-28aa-4170-87b0-b3f130bb0343
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230522042028_140189c4-28aa-4170-87b0-b3f130bb0343
INFO  : Session is already open
INFO  : Dag name: CREATE TABLE highest_selling_cate...category (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1680119865937_1850)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 10.11 s
```



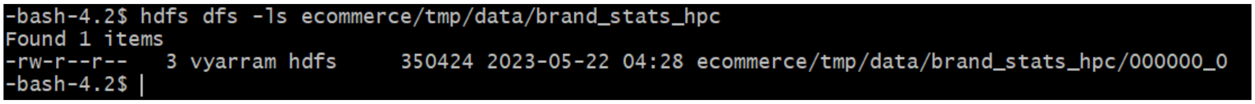
```
-bash-4.2$ ls
000000_0  000123_0  kaggle_project  top10country.csv  tweets_alphago.csv
-bash-4.2$ |
```

- Creating Table for the Popular Brands In the highly Purchasing category

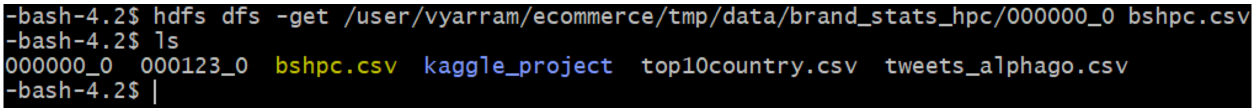
```
CREATE TABLE BRAND_STATISTICS_HPC
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/brand_stats_hpc'
as
SELECT BRAND, CATEGORY, COUNT(*) FROM ECOMMERCE
WHERE CATEGORY IN (
SELECT CATEGORY FROM ECOMMERCE
GROUP BY CATEGORY
ORDER BY COUNT(*) DESC LIMIT 1)
GROUP BY BRAND , CATEGORY ;
```



```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE TABLE BRAND_STATISTICS_HPC
. . . . . . . . . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . . . . . . . . . . . . . . . . . . . .> STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/brand_stats_hpc'
. . . . . . . . . . . . . . . . . . . . . . .> as
. . . . . . . . . . . . . . . . . . . . . . .> SELECT BRAND, CATEGORY, COUNT(*) FROM ECOMMERCE
. . . . . . . . . . . . . . . . . . . . . . .> WHERE CATEGORY IN (
. . . . . . . . . . . . . . . . . . . . . . .> SELECT CATEGORY FROM ECOMMERCE
. . . . . . . . . . . . . . . . . . . . . . .> GROUP BY CATEGORY
. . . . . . . . . . . . . . . . . . . . . . .> ORDER BY COUNT(*) DESC LIMIT 1)
. . . . . . . . . . . . . . . . . . . . . . .> GROUP BY BRAND , CATEGORY ;
INFO  : Compiling command(queryId=hive_20230522042836_33f78c6c-fa45-4f6e-b0af-608d0818a811): CREATE TABLE BRAND_STATI
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/brand_stats_hpc'
as
SELECT BRAND, CATEGORY, COUNT(*) FROM ECOMMERCE
WHERE CATEGORY IN (
SELECT CATEGORY FROM ECOMMERCE
GROUP BY CATEGORY
ORDER BY COUNT(*) DESC LIMIT 1)
GROUP BY BRAND , CATEGORY
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:brand, type:string, comment:null), FieldSchema(n
ame:_c2, type:bigint, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230522042836_33f78c6c-fa45-4f6e-b0af-608d0818a811); Time taken: 0.
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522042836_33f78c6c-fa45-4f6e-b0af-608d0818a811): CREATE TABLE BRAND_STATI
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/brand_stats_hpc'
as
SELECT BRAND, CATEGORY, COUNT(*) FROM ECOMMERCE
WHERE CATEGORY IN (
SELECT CATEGORY FROM ECOMMERCE
GROUP BY CATEGORY
ORDER BY COUNT(*) DESC LIMIT 1)
GROUP BY BRAND , CATEGORY
INFO  : Query ID = hive_20230522042836_33f78c6c-fa45-4f6e-b0af-608d0818a811
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230522042836_33f78c6c-fa45-4f6e-b0af-608d0818a811
INFO  : Session is already open
INFO  : Dag name: CREATE TABLE BRAND_STATISTICS_HPC...CATEGORY (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1680119865937_1850)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 3 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 4 ...... container     SUCCEEDED      6          6        0        0       0       0
Reducer 5 ...... container     SUCCEEDED      1          1        0        0       0       0
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 05/05  [==========================>>] 100%  ELAPSED TIME: 13.33 s
----------------------------------------------------------------------------------------------
```



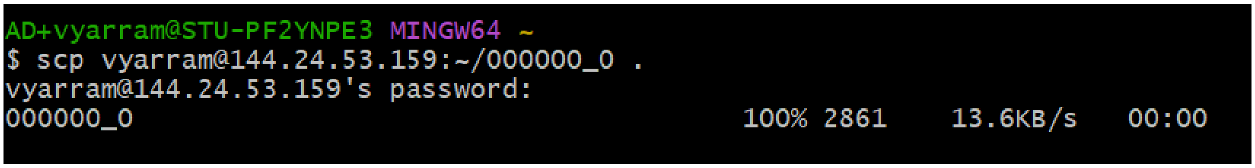
```
-bash-4.2$ hdfs dfs -ls ecommerce/tmp/data/brand_stats_hpc
Found 1 items
-rw-r--r--   3 vyarram hdfs     350424 2023-05-22 04:28 ecommerce/tmp/data/brand_stats_hpc/000000_0
-bash-4.2$ |
```



```
-bash-4.2$ hdfs dfs -get /user/vyarram/ecommerce/tmp/data/brand_stats_hpc/000000_0 bshpc.csv
-bash-4.2$ ls
000000_0  000123_0  bshpc.csv  kaggle_project  top10country.csv  tweets_alphago.csv
-bash-4.2$ |
```

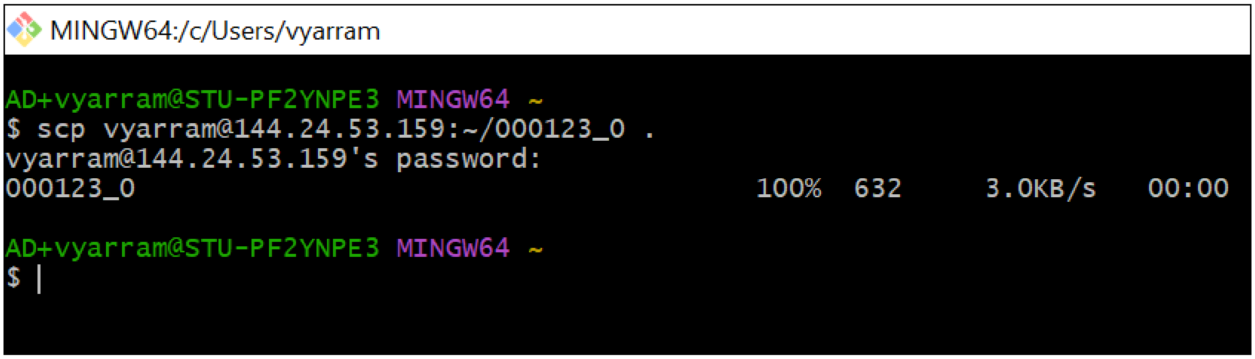
Download the 000000_0, 000123.csv, bshpc.csv from Linux Server to your local computer directory with scp.

```
$ scp vyarram@144.24.53.159:~/000000_0 .
```



```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ scp vyarram@144.24.53.159:~/000000_0 .
vyarram@144.24.53.159's password:
000000_0                                    100% 2861    13.6KB/s    00:00
```

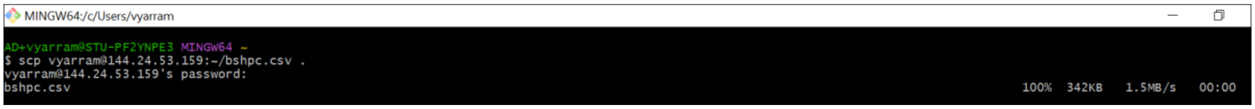
```
$ scp vyarram@144.24.53.159:~/000123_0 .
```



MINGW64:/c/Users/vyarram

```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ scp vyarram@144.24.53.159:~/000123_0 .
vyarram@144.24.53.159's password:
000123_0                                    100%   632      3.0KB/s    00:00

AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ |
```

```
$ scp vyarram@144.24.53.159:~/bshpc.csv .
```



MINGW64:/c/Users/vyarram

```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ scp vyarram@144.24.53.159:~/bshpc.csv .
vyarram@144.24.53.159's password:
bshpc.csv                                   100%  342KB   1.5MB/s   00:00
```

The category code has been segregated into three different Category, Subcategory and Product

Example:

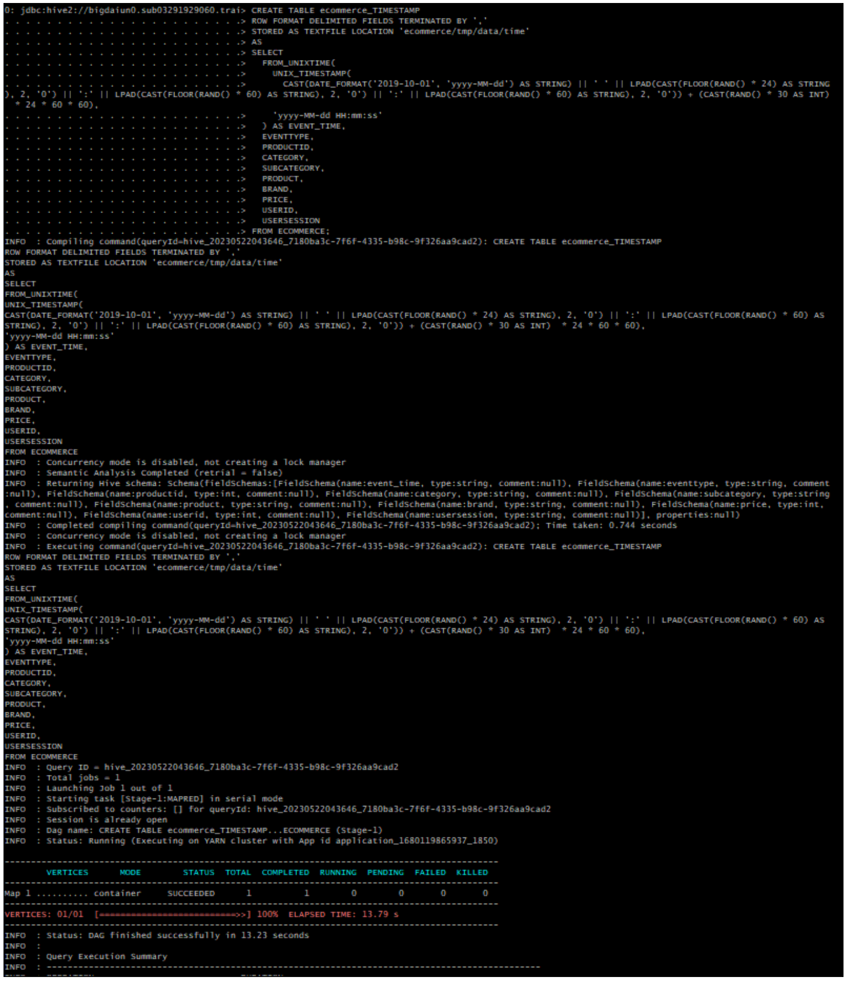
Category Code: appliances.environment.water_heater

Category: Appliances

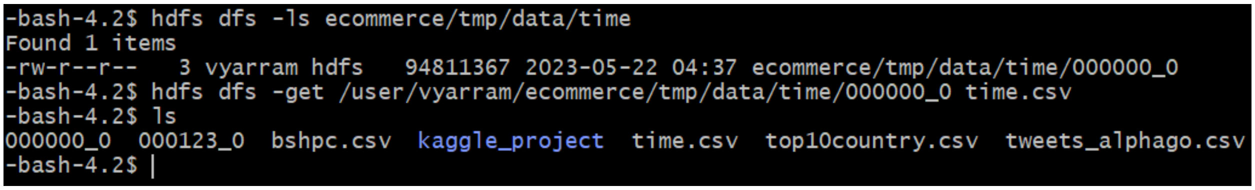
Subcategory: Environment

Product: Water Heater

```
CREATE TABLE ecommerce_TIMESTAMP
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/time'
AS
SELECT
  FROM_UNIXTIME(
    UNIX_TIMESTAMP(
      CAST(DATE_FORMAT('2019-10-01', 'yyyy-MM-dd') AS STRING) || ' ' ||
LPAD(CAST(FLOOR(RAND() * 24) AS STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() *
60) AS STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS STRING), 2, '0')) +
(CAST(RAND() * 30 AS INT)  * 24 * 60 * 60),
    'yyyy-MM-dd HH:mm:ss'
  ) AS EVENT_TIME,
  EVENTTYPE,
  PRODUCTID,
  CATEGORY,
  SUBCATEGORY,
  PRODUCT,
  BRAND,
  PRICE,
  USERID,
  USERSESSION
FROM ECOMMERCE;
```

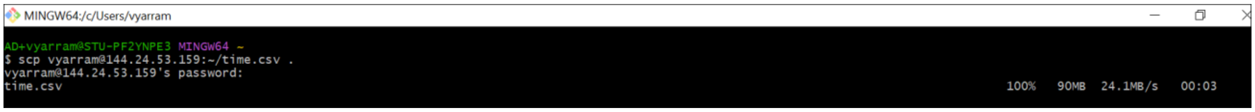


```
0: jdbc:hive2://bigdaiun0.sub03291929060.trai> CREATE TABLE ecommerce_TIMESTAMP
. . . . . . . . . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . . . . . . . . . . . . . . . . . . . .> STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/time'
. . . . . . . . . . . . . . . . . . . . . . .> AS
. . . . . . . . . . . . . . . . . . . . . . .> SELECT
. . . . . . . . . . . . . . . . . . . . . . .>   FROM_UNIXTIME(
. . . . . . . . . . . . . . . . . . . . . . .>     UNIX_TIMESTAMP(
. . . . . . . . . . . . . . . . . . . . . . .>       CAST(DATE_FORMAT('2019-10-01', 'yyyy-MM-dd') AS STRING) || ' ' || LPAD(CAST(FLOOR(RAND() * 24) AS STRING
), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS STRING), 2, '0')) + (CAST(RAND() * 30 AS INT)
  * 24 * 60 * 60),
. . . . . . . . . . . . . . . . . . . . . . .>     'yyyy-MM-dd HH:mm:ss'
. . . . . . . . . . . . . . . . . . . . . . .>   ) AS EVENT_TIME,
. . . . . . . . . . . . . . . . . . . . . . .>   EVENTTYPE,
. . . . . . . . . . . . . . . . . . . . . . .>   PRODUCTID,
. . . . . . . . . . . . . . . . . . . . . . .>   CATEGORY,
. . . . . . . . . . . . . . . . . . . . . . .>   SUBCATEGORY,
. . . . . . . . . . . . . . . . . . . . . . .>   PRODUCT,
. . . . . . . . . . . . . . . . . . . . . . .>   BRAND,
. . . . . . . . . . . . . . . . . . . . . . .>   PRICE,
. . . . . . . . . . . . . . . . . . . . . . .>   USERID,
. . . . . . . . . . . . . . . . . . . . . . .>   USERSESSION
. . . . . . . . . . . . . . . . . . . . . . .> FROM ECOMMERCE;
INFO  : Compiling command(queryId=hive_20230522043646_7180ba3c-7f6f-4335-b98c-9f326aa9cad2): CREATE TABLE ecommerce_TIMESTAMP
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/time'
AS
SELECT
FROM_UNIXTIME(
UNIX_TIMESTAMP(
CAST(DATE_FORMAT('2019-10-01', 'yyyy-MM-dd') AS STRING) || ' ' || LPAD(CAST(FLOOR(RAND() * 24) AS STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS
STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS STRING), 2, '0')) + (CAST(RAND() * 30 AS INT)  * 24 * 60 * 60),
'yyyy-MM-dd HH:mm:ss'
) AS EVENT_TIME,
EVENTTYPE,
PRODUCTID,
CATEGORY,
SUBCATEGORY,
PRODUCT,
BRAND,
PRICE,
USERID,
USERSESSION
FROM ECOMMERCE
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:event_time, type:string, comment:null), FieldSchema(name:eventtype, type:string, comment
:null), FieldSchema(name:productid, type:int, comment:null), FieldSchema(name:category, type:string, comment:null), FieldSchema(name:subcategory, type:string
, comment:null), FieldSchema(name:product, type:string, comment:null), FieldSchema(name:brand, type:string, comment:null), FieldSchema(name:price, type:int,
comment:null), FieldSchema(name:userid, type:int, comment:null), FieldSchema(name:usersession, type:string, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230522043646_7180ba3c-7f6f-4335-b98c-9f326aa9cad2); Time taken: 0.744 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20230522043646_7180ba3c-7f6f-4335-b98c-9f326aa9cad2): CREATE TABLE ecommerce_TIMESTAMP
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'ecommerce/tmp/data/time'
AS
SELECT
FROM_UNIXTIME(
UNIX_TIMESTAMP(
CAST(DATE_FORMAT('2019-10-01', 'yyyy-MM-dd') AS STRING) || ' ' || LPAD(CAST(FLOOR(RAND() * 24) AS STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS
STRING), 2, '0') || ':' || LPAD(CAST(FLOOR(RAND() * 60) AS STRING), 2, '0')) + (CAST(RAND() * 30 AS INT)  * 24 * 60 * 60),
'yyyy-MM-dd HH:mm:ss'
) AS EVENT_TIME,
EVENTTYPE,
PRODUCTID,
CATEGORY,
SUBCATEGORY,
PRODUCT,
BRAND,
PRICE,
USERID,
USERSESSION
FROM ECOMMERCE
INFO  : Query ID = hive_20230522043646_7180ba3c-7f6f-4335-b98c-9f326aa9cad2
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230522043646_7180ba3c-7f6f-4335-b98c-9f326aa9cad2
INFO  : Session is already open
INFO  : Dag name: CREATE TABLE ecommerce_TIMESTAMP...ECOMMERCE (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1680119865937_1850)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 13.79 s
----------------------------------------------------------------------------------------------
INFO  : Status: DAG finished successfully in 13.23 seconds
INFO  :
INFO  : Query Execution Summary
INFO  : ----------------------------------------------------------------------------------------------
```



```
-bash-4.2$ hdfs dfs -ls ecommerce/tmp/data/time
Found 1 items
-rw-r--r--   3 vyarram hdfs   94811367 2023-05-22 04:37 ecommerce/tmp/data/time/000000_0
-bash-4.2$ hdfs dfs -get /user/vyarram/ecommerce/tmp/data/time/000000_0 time.csv
-bash-4.2$ ls
000000_0  000123_0  bshpc.csv  kaggle_project  time.csv  top10country.csv  tweets_alphago.csv
-bash-4.2$ |
```

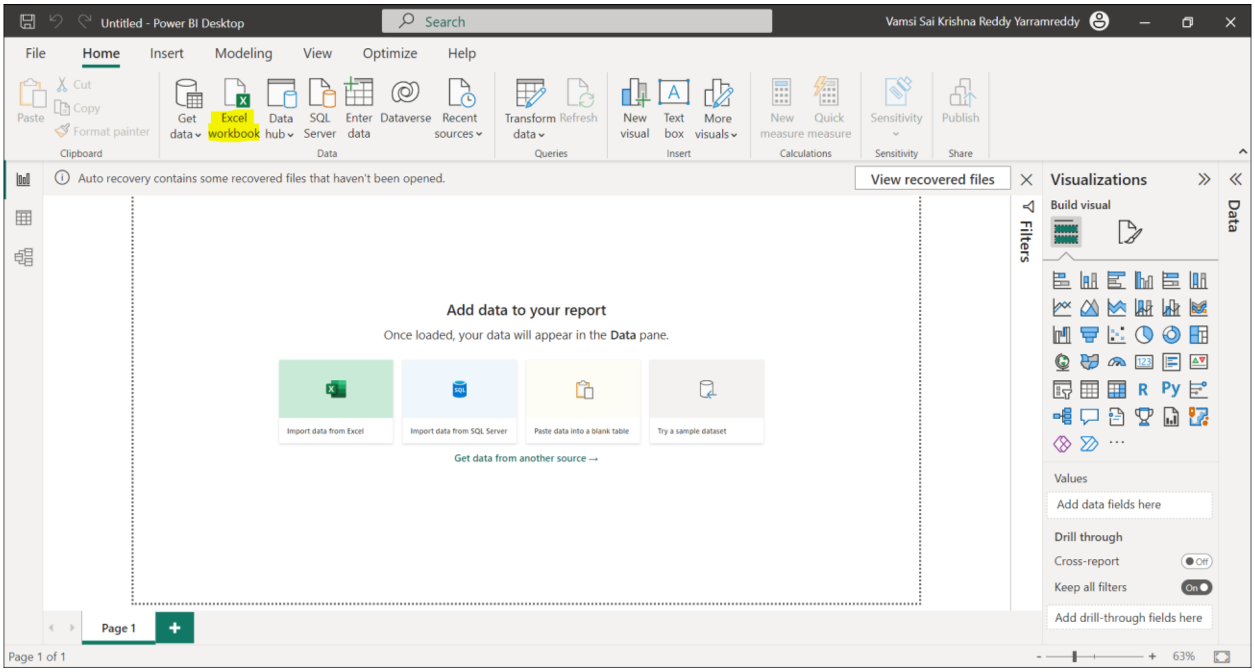
$ scp vyarram@144.24.53.159:~/time.csv .



```
AD+vyarram@STU-PF2YNPE3 MINGW64 ~
$ scp vyarram@144.24.53.159:~/time.csv .
vyarram@144.24.53.159's password:
time.csv                                                                  100%   90MB  24.1MB/s   00:03
```

Step: Visualization

Now upload the Csv's file into Power Bi to create the Visualization charts.

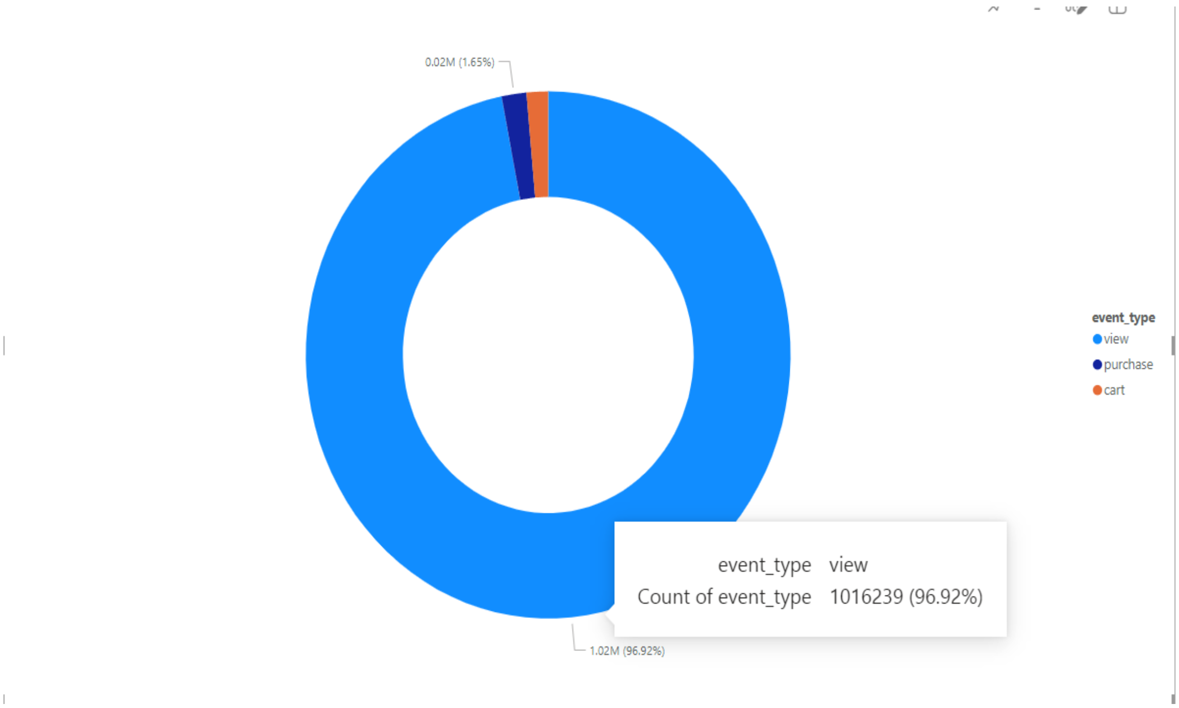


Auto recovery contains some recovered files that haven't been opened.

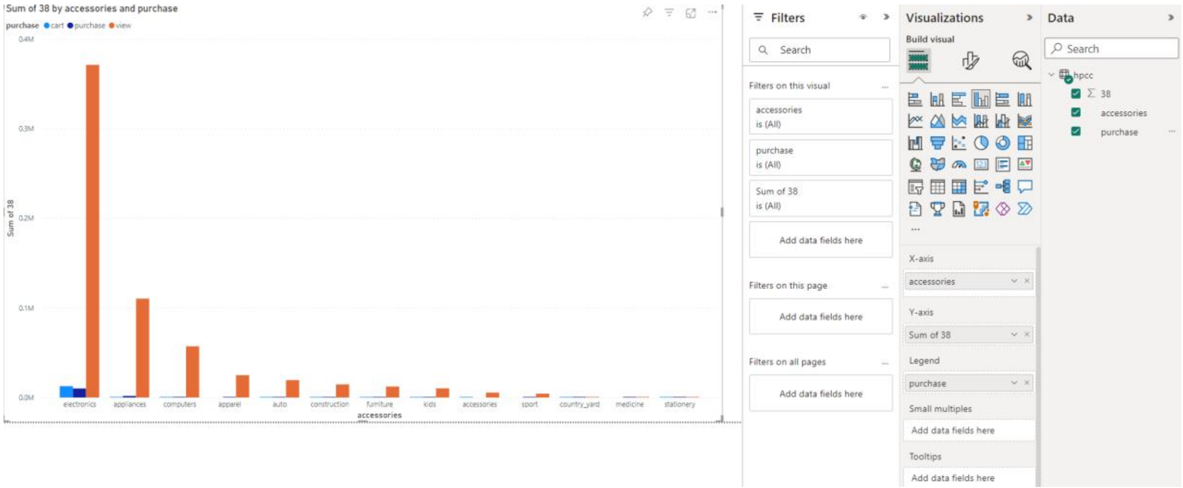
Add data to your report

Once loaded, your data will appear in the **Data** pane.

- STATISTICS OF VIEWS, CART AND PURCHASES



- Count of Event Type of each Category



Filters

Search

Filters on this visual

accessories
is (All)

purchase
is (All)

Sum of 38
is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

Add data fields here

Visualizations

Build visual

X-axis

accessories

Y-axis

Sum of 38

Legend

purchase

Small multiples

Add data fields here

Tooltips

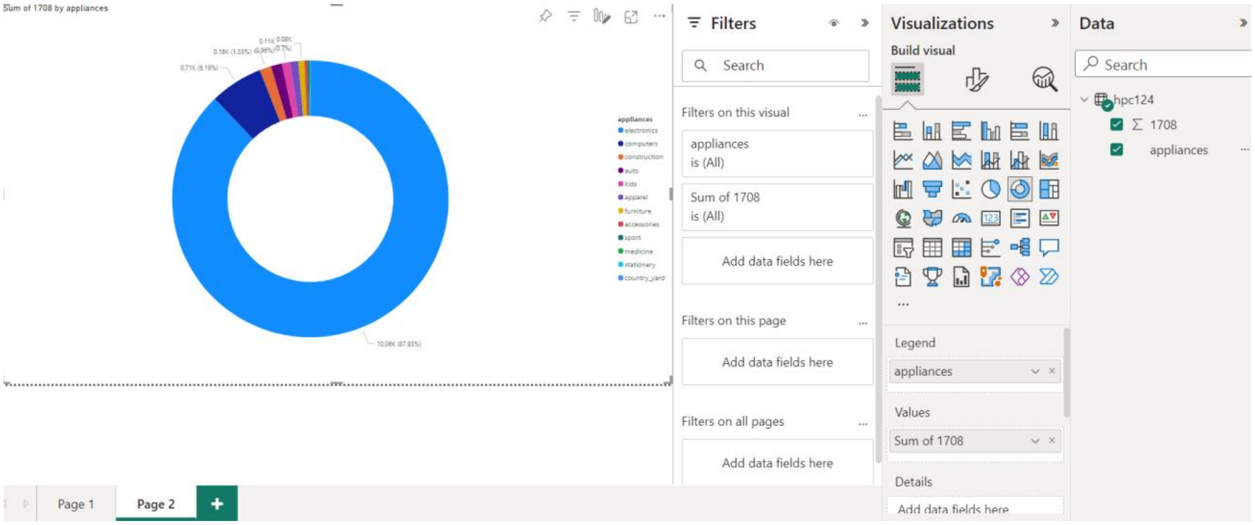
Add data fields here

Data

Search

hpcc
- ☑ Σ 38
- ☑ accessories
- ☑ purchase

- Highly Purchased Categories



Filters on this visual

appliances is (All)

Sum of 1708 is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

Add data fields here

Legend

appliances

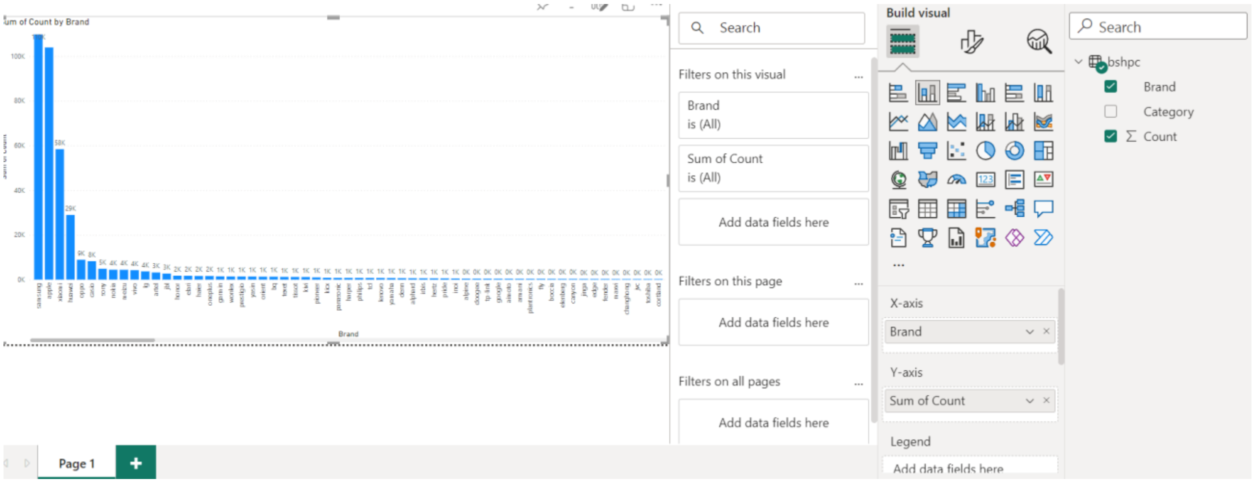
Values

Sum of 1708

Details

Add data fields here

- Popular Brands in Highly Purchased Category



Filters on this visual

Brand is (All)

Sum of Count is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

Add data fields here

X-axis

Brand

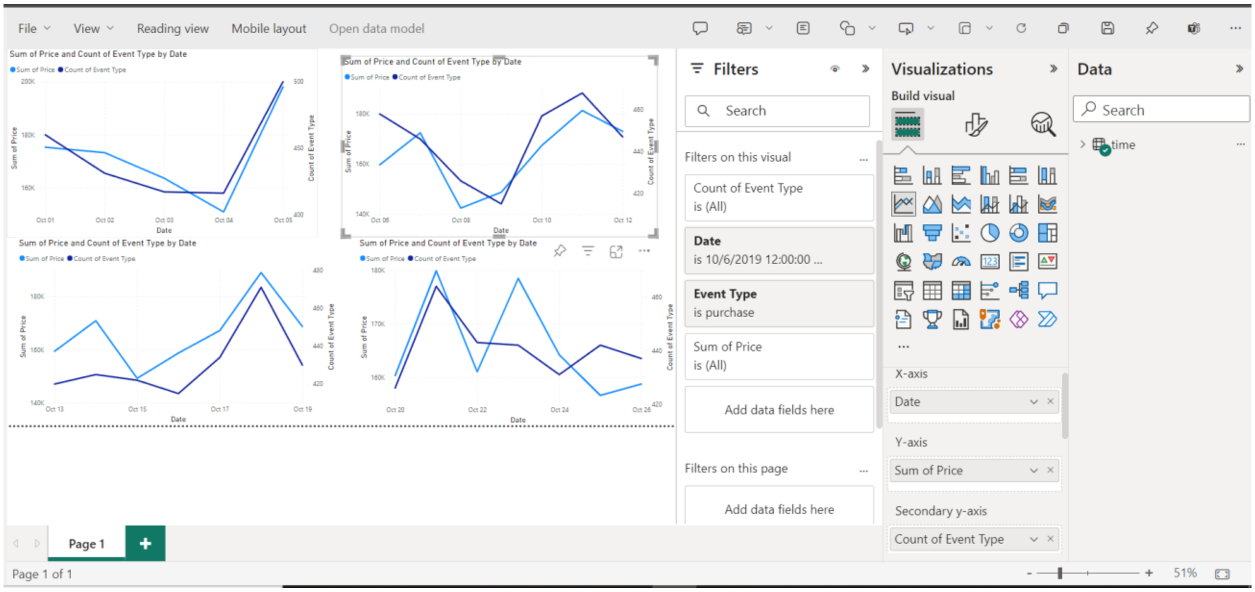
Y-axis

Sum of Count

Legend

Add data fields here

- 4-Week Data Report for the month of Oct-2019



Filters on this visual

Count of Event Type
is (All)

Date
is 10/6/2019 12:00:00 ...

Event Type
is purchase

Sum of Price
is (All)

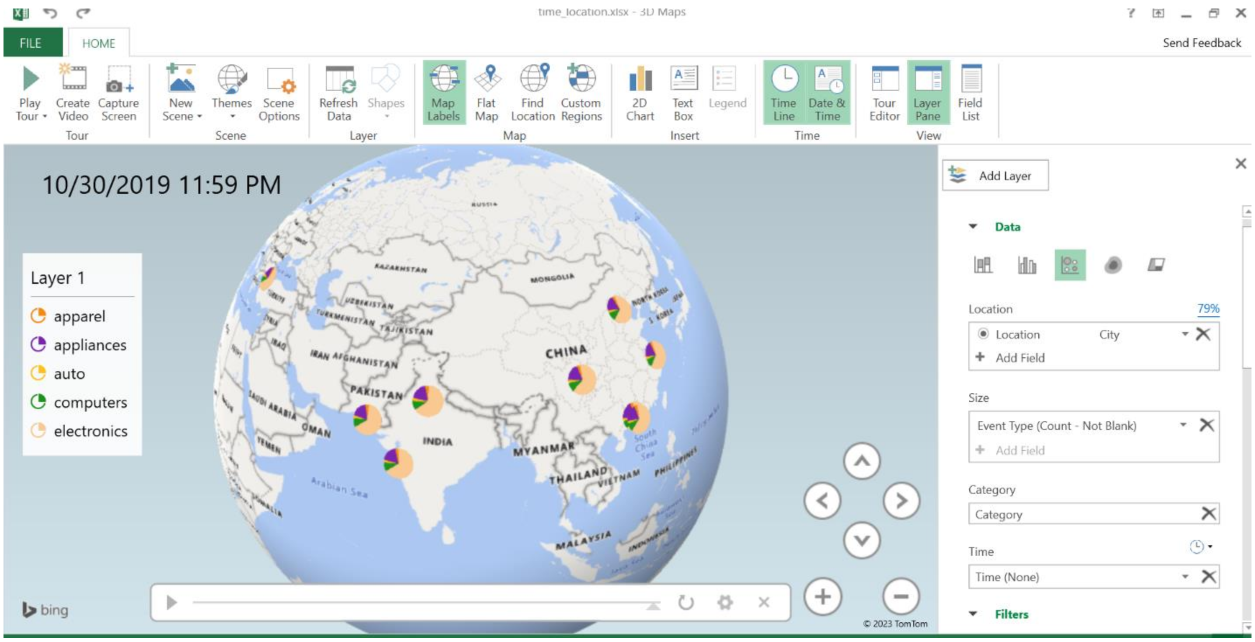
Filters on this page

X-axis
Date

Y-axis
Sum of Price

Secondary y-axis
Count of Event Type

- Spatial Analysis



10/30/2019 11:59 PM

Layer 1

- ◔ apparel
- ◔ appliances
- ◔ auto
- ◔ computers
- ◔ electronics

Data

Location 79%

○ Location City

+ Add Field

Size

Event Type (Count - Not Blank)

+ Add Field

Category

Category

Time

Time (None)

Filters

References

1. URL of Data Source : [eCommerce behavior data from multi category store | Kaggle]

2. Github : https://github.com/prathimasarvani/5200-System-Analysis-and-Design