Knowing Your Customer Requirement in E-Commerce (KYCRE)

by

Sakthidevi T

Objective

This project is to develop a general analysis on e-commerce store where sports product (such as Air Sports, Combat Sports, Dancing, Exercise & Fitness, Games, Gymnastics, Indoor Games, Jumping, Outdoor Play Equipment, Outdoor Recreation, Puzzles, Racquet Sports, Team Sports, Water Sports, Winter Sports) can be bought from the comfort of home through the Internet.

This applications support the interaction between different parties participating in a commerce transaction via the network, as well as the management of the data involved in the process.

Scope

The scope of the process is to analyze the requirement of the customer in terms of transaction, category, product, payment mode, city etc. The outcome of this analysis will lead the organization to satisfy the customer as well as in economic benefit.

Background

E-commerce is fast gaining ground as an accepted and used business paradigm. More and more business houses are implementing web sites providing functionality for performing commercial transactions over the web. It is reasonable to say that the process of shopping on the web is becoming commonplace.

Requirements Specifications

This project deals with E-Commerce transaction, we have to handle huge volume of data (which will rise tremendously). Here we are having two kinds of data.

➢ Transactional data which contains transaction details of each customer (such as transaction id, transaction amount, customer id, payment mode, transaction place, transaction date etc.)
➢ Customer data which contains the details of customer (such as customer id, first name, last name, professional and age).

Analysis

Being a data analyzing project, we are going to implement this project with the help HADOOP, an open source Java-based programming framework. There are many Ecosystem tools in HADOOP from there we used **Pig** and **Hive**

Technologies

- ➢ **Map Reduce:** Hadoop Map Reduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner
- ➢ **Pig:** Pig is a high-level platform for creating programs. The language for this platform is called Pig Latin. It can be extended using User Defined Functions (UDFs) which the user can write in Java, Python, JavaScript, Ruby or Groovy and then call directly from the language.
- ➢ **Hive:** Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. The traditional SQL queries must be implemented in the Map Reduce Java API to execute SQL applications and queries over a distributed data.

Use cases:

Project tasks are divided into different use cases based on analysis.

Sales Report

A sales analysis report shows the trends that occur in a company's sales volume over time. In its most basic form, a sales analysis report shows whether sales are increasing or declining. Under this category we have taken two tasks

- i) Calculate total amount sales for each Month.
- ii) Divide the file into 12 files, each file containing each month of data. For e.g. file 1 should contain data of January transaction, file 2 should contain data of February transaction.
- iii) We have implemented Custom Input Format.

Using Map Reduce

Execution Step:

```
hduser@ubuntu64server:~$ hadoop fs -put /home/hduser/txns-large.dat /
hduser@ubuntu64server:~$ hadoop jar Project.jar /txns-large.dat /project
Enter your task
4
```
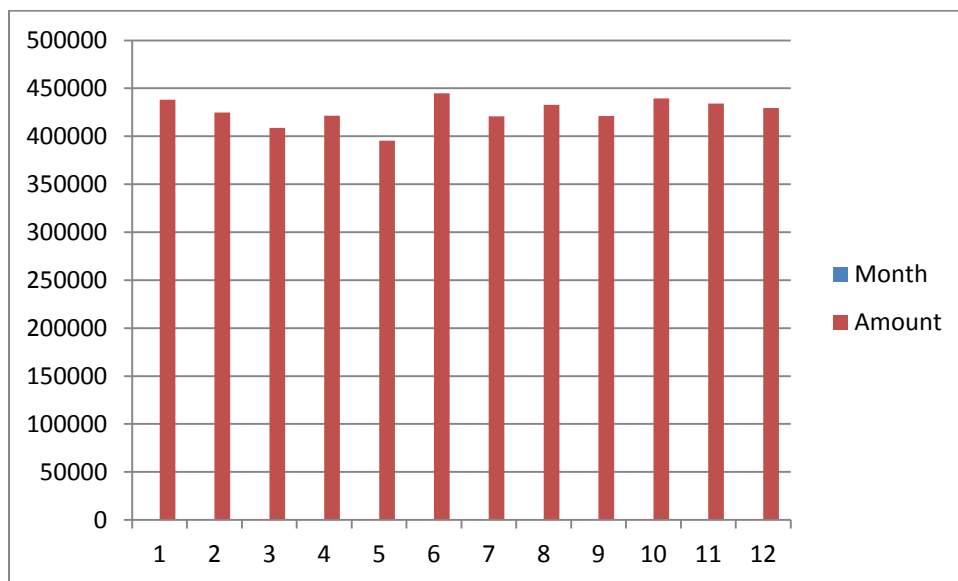
Output

```
1        438165.7600000004
10       424856.2800000005
11       408846.3499999995
12       421490.72999999934
2        395262.3699999991
3        444664.2399999997
4        420695.24000000145
```

Execution Step:

```
hduser@ubuntu64server:~$ hadoop jar Project.jar /txns-large.dat /project
Enter your task
5
```

```
01       128.62
01       117.25
01       93.15
01       179.88
```

Report

Gift card

Gift cards can be an effective way to increase online sales, acquire new customers, and encourage return visits. It's even discounted ones, may also contribute to margins. Kickback gift card promotions give money, in the form of a second gift card, back to the giver as a reward for buying the gift card. Under this category we have taken three tasks

> i)   Transaction made above 160.
> ii) Transaction made between the range 175 to 200
> iii) Number of transaction, Sum of transaction and average of transaction calculated
>   for each customer.
> iv) We have implemented Custom Input Format.

Map Reduce

Execution Step:

E-Commerce Task1

```
hduser@ubuntu64server:~$ hadoop jar Project.jar /txns-large.dat /project
Enter your task
1
Enter the minimum value
160
```

Output

```
4006924 177.9
4009373 181.2
4000172 198.33
4005005 169.85
4000041 178.06
```

E-Commerce Task2

```
hduser@ubuntu64server:~$ hadoop jar Project.jar /txns-large.dat /project
Enter your task
2
Enter the minimum value
175
Enter the minimum value
200
```

Output

```
hduser@ubuntu64server:~$ hadoop fs -cat /project/p*
6524
hduser@ubuntu64server:~$
```

E-Commerce Task3

```
hduser@ubuntu64server:~$ hadoop jar Project.jar /txns-large.dat /project
Enter your task
3
Enter the user id
4000002
```

Output

```
hduser@ubuntu64server:~$ hadoop fs -cat /project/p*
4000002 112.33 2 56.165
```

Hive

Execution Step

E-Commerce Task1

```
ockford Illinois          credit
Time taken: 34.798 seconds
hive> select uid, amt from transaction where amt>160;
```

Output

```
                    cloudera@localhost:~                      _ □ ×
 File  Edit  View  Search  Terminal  Help
4009373 181.2
4000172 198.33
4005005 169.85
4000041 178.06
4002205 186.38
4009081 182.06
4000454 199.35
4008385 163.0
4007113 189 15
```

E-Commerce Task2

```
4007843 180.41
4001406 168.49
Time taken: 279.567 seconds
hive> select count(amt) from transaction where amt between 175 and 200;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
```
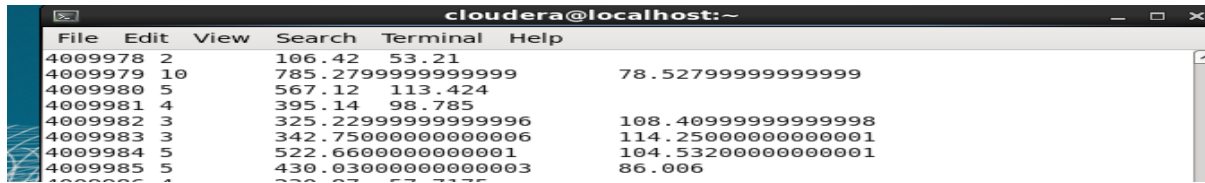
Output

```
ite: 5 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 630 msec
OK
6528
Time taken: 163.668 seconds
```
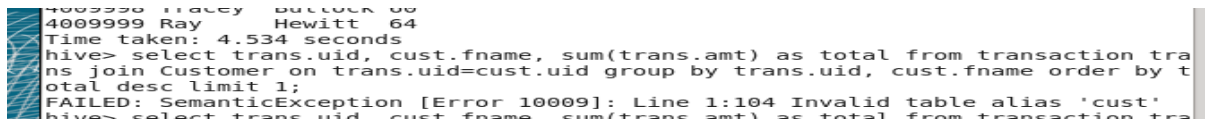
E-Commerce Task3

```
Time taken: 163.668 seconds
hive> select uid, count(amt), sum(amt), avg(amt) from transaction group by uid;

   hive
                        Firefox automatically sends some data to Mozilla so that
```

Output

E-commerce Task using Join


```
4009998 Tracey   Bullock 66
4009999 Ray      Hewitt  64
Time taken: 4.534 seconds
hive> select trans.uid, cust.fname, sum(trans.amt) as total from transaction tra
ns join Customer on trans.uid=cust.uid group by trans.uid, cust.fname order by t
otal desc limit 1;
FAILED: SemanticException [Error 10009]: Line 1:104 Invalid table alias 'cust'
hive> select trans uid   cust fname   sum(trans amt) as total from transaction tra
```
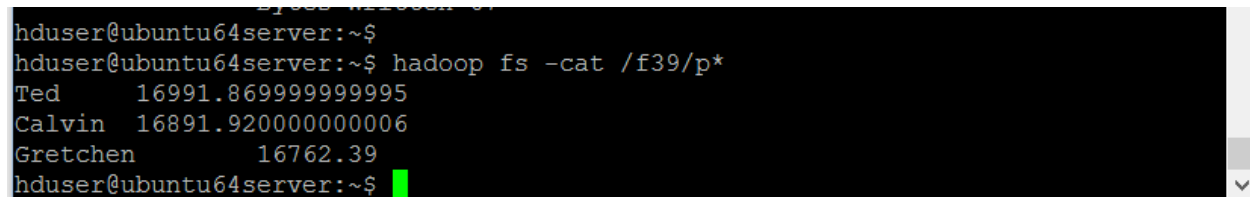
Output


```
Total MapReduce CPU Time Spent: 20 seconds 950 msec
OK
4009485 Stuart  1973.3
Time taken: 423.874 seconds
hive>
```

Honors and Waiver

In this E-Commerce website the customer who made more transaction in one day and in vocational period is honored as "champions" and provide waiver for the next transaction. From this we can attract the customer and we can have hold with customer. Under this we have taken three tasks
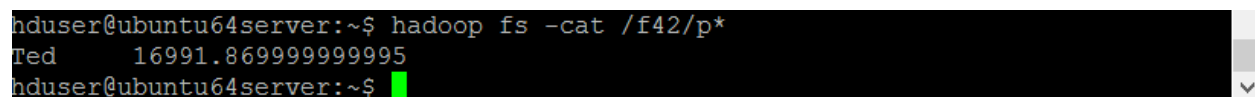
     i)   Find the name of top 3 spenders.

     ii)   Find the name of user who has spends the maximum amount.

     iii)   Find the user who has spends the max amount in July month.

Output: Find the name of top 3 spenders.


```
hduser@ubuntu64server:~$
hduser@ubuntu64server:~$ hadoop fs -cat /f39/p*
Ted     16991.869999999995
Calvin  16891.920000000006
Gretchen        16762.39
hduser@ubuntu64server:~$
```

Output: Find the name of user who has spends the maximum amount.


```
hduser@ubuntu64server:~$ hadoop fs -cat /f42/p*
Ted     16991.869999999995
hduser@ubuntu64server:~$
```

Output: Find the user who has spends the max amount in July month.

```
                    Bytes Written=13
hduser@ubuntu64server:~$ hadoop fs -cat /f37/p*
Toni    2082.44
hduser@ubuntu64server:~$
```

Extras

Payment mode

Ecommerce is the exchange of goods and services enabled through an electronic method. It is also known as a sample of Electronic Data Interchange (EDI), e-commerce payment systems have become increasingly popular due to the widespread use of the internet-based shopping and banking. There are various forms of payment for ecommerce. From the given data we are having

> ➢ Credit: The easiest form of electronic money that is available and most widely used today.
> ➢ Cash: Cash on delivery has emerged as one of the most sought after services for e-commerce entities and it is reported that in some cases as high as 50 per cent of orders are placed with various online retailers with this payment option, while the remaining opt for credit card or bank payments.

Conclusion

The rapid growth in electronic commerce around world has prompted many to look for better ways of measuring the phenomenon. As more and more countries and international agencies become involved, it is important to develop plans to ensure that there is no unnecessary duplication of effort and that users have the data necessary for informed decision making at the earliest possible opportunity. From the above analysis, it appears that there are a number of actions to be taken to develop the organization economic and to satisfy the customer.