

Real-time analytics in Ecommerce Transaction

Project Author : Niyaz Ahmed

Date : 22/11/2016

Objective : Real-time analytics in the use of the capacity to use, data and related resources as soon as the data enters the system

Test Data :

1) Ecommerce Transaction Data (txns-large.dat)

TransID	Date	UserId	Amount	Cat	Machine	Sate	City	Payment

2) Customer Details (custs-large.dat)

UserID	First Name	Last Name	Age	Position

Use Case 1 : Constraint Based Amount Scenario

- To find the product based on the user search or product the user has purchased.
- Whenever user purchases a product of a particular price or within range of amount than at time the user will provide with similar type of product within the same range.
- **Data Validation : Yes.**

Constraint : User Input Can be Only Numbers

OUTPUT

```
Last login: Mon Nov 21 04:31:42 2016 from 192.168.56.1
hduser@ubuntu64server:~$ hadoop jar Ctask1.jar /17Sep/txns-large.dat/ ct1
Enter the amount
170
```

```
hduser@ubuntu64server:~$ hadoop fs -ls /ct1
Found 2 items
-rw-r--r--  1 hduser supergroup          0 2016-11-21 05:27 /ct1/_SUCCESS
-rw-r--r--  1 hduser supergroup    82866 2016-11-21 05:27 /ct1/part-m-00000
hduser@ubuntu64server:~$ hadoop fs -cat /ct1/p*
```

```
00046831      191.38
00046832      198.49
00046836      199.41
00046842      189.37
00046846      184.89
00046849      186.0
00046860      196.9
```

```
hduser@ubuntu64server:~$ hadoop jar Ctask2.jar /17Sep/txns-large.dat/ ct2
Enter the lower limit
120
Enter the upper limit
140
```

```
hduser@ubuntu64server:~$ hadoop fs -ls /ct2
Found 2 items
-rw-r--r--  1 hduser supergroup          0 2016-11-21 05:44 /ct2/_SUCCESS
-rw-r--r--  1 hduser supergroup          5 2016-11-21 05:44 /ct2/part-r-00000
hduser@ubuntu64server:~$ hadoop fs -cat /ct2/p*
6524
hduser@ubuntu64server:~$
```

Use Case 2 : Top Three Spender or in a particular month

- During a e-shopping festival many people shop on a particular day so based on amount spend by the customer, the top three customer who will spend maximum money on shopping will be provided with gift hamper.
- The same scenario can be applied when the shopping festival is for the entire month.

OUTPUT

```
hduser@ubuntu64server: ~  
hduser@ubuntu64server:~$ hadoop jar Ctask7.jar /17Sep/txns-large.dat /cttt7
```

```
hduser@ubuntu64server: ~  
hduser@ubuntu64server:~$ hadoop fs -ls /cttt7  
Found 2 items  
-rw-r--r--  1 hduser supergroup      0 2016-11-21 22:54 /cttt7/_SUCCESS  
-rw-r--r--  1 hduser supergroup    77 2016-11-21 22:54 /cttt7/part-r-00000  
hduser@ubuntu64server:~$ hadoop fs -cat /cttt7/p*  
Ted      16991.8700000000006  
Calvin  16891.9200000000006  
Gretchen      16762.3900000000003  
hduser@ubuntu64server:~$
```

Use Case 3 : Revenue made via transaction

- This Use Case is use by the E-Commerce Company to check the revenue they have generated in a single day, month or any month or a complete year.
- Revenue sorted via them can be categorized on their needs.
- **Data Validation** : Yes.

Constraint : User Input Can be Only Numbers for Months

OUTPUT

```
hduser@ubuntu64server:~$ hadoop jar Ctask4.jar /17Sep/txns-large.dat/ ct4
Enter the Months
11
```

```
hduser@ubuntu64server:~$ hadoop fs -ls /ct44
Found 2 items
-rw-r--r-- 1 hduser supergroup          0 2016-11-21 09:03 /ct44/_SUCCESS
-rw-r--r-- 1 hduser supergroup        22 2016-11-21 09:03 /ct44/part-r-00000
hduser@ubuntu64server:~$ hadoop fs -cat /ct44/p*
12      421490.72999999876
hduser@ubuntu64server:~$
```

```
hduser@ubuntu64server:~$ hadoop fs -cat /ct55/part-r-00002
03      444664.23999999935
hduser@ubuntu64server:~$ hadoop fs -cat /ct55/p*
01      438165.76000000004
02      395262.37000000005
03      444664.23999999935
04      420695.23999999999
05      432627.57999999967
06      421074.54999999976
07      439560.80000000002
08      434255.01000000106
09      429321.63000000105
10      424856.28000000009
11      408846.3499999989
12      421490.72999999997
hduser@ubuntu64server:~$
```

Use Case 4 : Customer Dilemma

- Normally people who do shopping online face some situation regarding the product, delivery or any other problem.
- So in that scenario the customer calls the customer care and the people there addresses the problem. For which they ask for Customer Id No so that all the data regarding the user can fetch and problem regarding the issues can resolved.
- **Data Validation :** Yes.

Constraint : For Customer Care People Input Can be Only Numbers

OUTPUT

```
hduser@ubuntu64server:~$ hadoop jar Ctask3.jar /17Sep/txns-large.dat/ ct3
Enter the User Id
4006742
```

CONCLUSION : Thus a Real Time Analysis of E-Commerce happens as soon as the Data Enters the System and Output is generated fast as per desire or needs.