Project-1 Documentation

Title: Ecommerce Data – An 

View

Date: 22/11/2016

Developer: Oliver Prasanna R

Abstract: Given a Test Sample of Transaction data from an Ecommerce Portal.

TransactionID	DOT	UserID	Amount	Category	Accessories	Place	City	Payment Mode
0	06-26-2015	4007024	40.33	Exercise & Fitness	Cardio Machine Accessories	Clarksville	Tennessee	credit
1	05-26-2015	4006742	198.44	Exercise & Fitness	Weightlifting Gloves	Long Beach	California	credit

Use-case Generation

Use Case 1: ***Constraint Based Amount Scenario***

- **Input File**: txns-large.dat
- **Input**: Amount [Custom Input from the User]
- **Key**: ID, **Value**: Amount
- **Output**: Set of User ID and Amount Based on the Input provided by the user
- **Data Validation**: Yes.
  Constraint: User Input Can be Only Numbers
- **Concept used** : Advanced Map Reduce

Description: We need to find the Set of User ID and their Associated Amount Based on the Amount Specified by the User.

Screen Shot 1.1: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task1.jar /Oliver/txns-large.da
t /Olive14
Enter the minimum value
100
16/11/21 12:02:33 INFO client.RMProxy: Connecting to ResourceManager at /192.168
.56.123:8032
```

Screen Shot 1.2: Output Window

```
4001371 187.72
4004939 198.32
4005788 159.5
4009362 141.7
4005452 101.34
4002061 175.61
4002286 121.81
4004311 184.18
4009827 142.03
4008449 192.67
4004318 199.07
4008637 198.4
4007202 129.43
4008092 156.38
4007571 123.58
4002940 144.91
4003685 191.29
4002441 139.78
4005772 177.22
4007287 163.81
4007843 180.41
4001406 168.49
```

Screen Shot 1.3: Data Validation

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task1.jar /Oliver/txns-large.da
t /Olive15
Enter the minimum value
hy
Please enter only numbers
hduser@ubuntu64server:~$
```

Output Path:

hadoop fs -cat /Olive15/part-m-00000

Use Case 2: ***A Single Spot with a Range***

- **Input File**: txns-large.dat
- **Input:** Upper and Lower Limit [Custom Input from the User]
- **Key:** ID, Name **Value:** Amount
- **Output:** A Single Count that match the Range
- **Data Validation:** Yes.
  Constraint: User Input Can be Only Numbers
- **Concept :** Advanced Map Reduce

Description: To Obtain the Exact Count of number of Amount Transaction within a Particular Range

Screen Shot 2.1: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task2.jar /Oliver/txns-large.da
t /Olive17
Enter the lower & upper limit
250
500
```

Screen Shot 2.2: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive17/part-r-00000
50000
```

Screen Shot 2.3: Data Validation

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task2.jar /Oliver/txns-large.da
t /Olive16
Enter the lower & upper limit
t
Please enter only numbers
```

Output Path:

Hadoop fs –cat /Olive17/part-r-00000

Use Case 3 & 4: ***An User Aggregate***

- **Input File**: txns-large.dat
- **Input**: ID [Custom Input from the User]
- **Key**: ID, **Value**: Amount
- **Output**: Count, Sum & Average Transaction of a User.
- **Data Validation**: Yes.
  Constraint: User Input Can be Only Numbers
- **Concept** : Advanced Map Reduce

Description: To obtain a summarized data of a user about the transaction completed.

Screen Shot 3.1 & 4. 1: Data Validation

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task3.jar /Oliver/txns-large.da
t /Olive22
tEnter Your Customer ID
t
Please enter only numbers
```

Screen Shot 3.2 & 4.2: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task3.jar /Oliver/txns-large.da
t /Olive21
Enter Your Customer ID
4007024
```

Screen Shot 3.3 & 4.3: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive21/part-r-00000
4007024 Sum960.11Count7Average137.15857142857143
```

Output Path:

Hadoop fs –cat /Olive21/part-r-00000

Use Case 5: ***Quick Month Sales Review***

- **Input File**: txns-large.dat
- **Input**: Month [Custom Input from the User]
- **Key**: Month, **Value**: Amount

- **Output** : Total Sales of Month
- **Data Validation**: Yes.

  Constraints 1: Month can be only between 1-12

  Constraints 2: Month can be only Positive Value

  Constraints 3: Month can be only in Numbers

  - **Concept**: Advanced Map Reduce

Description: To obtain a summarized data of a user about the transaction completed.

Screen Shot 5.1: Data Validation

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task5.jar /Oliver/txns-large.da
t /Olive25
Enter the month
y
Please enter only numbers
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task5.jar /Oliver/txns-large.da
t /Olive26
Enter the month
-5
Please Enter only Positive numbers
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task5.jar /Oliver/txns-large.da
t /Olive27
Enter the month
45
Please Enter a Valid month(1-12)
```

Screen Shot 5.2: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task5.jar /Oliver/txns-large.da
t /Olive28
Enter the month
8
```

Screen Shot 5.3: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive28/part-r-00000
08      434255.0100000014
```

Output Path:

Hadoop fs –cat /Olive28/part-r-00000

Use Case 6: ***Large Module – Small Module Visualization***

- **Input File**: txns-large.dat
- **Input** : Predefined
- **Key**: Month, **Value**: Entire Line of Data
- **Output** : Partitioned Month

- **Data Validation**: NA
- **Concept** : Map Reduce with Partitioner

Description: To Obtain a Distributed output for every month specific Data. Each of the output will be stored in a separate file based on month.

Screen Shot 6.1: ***Output Window [Multiple Partitioned Files]***

```
hduser@ubuntu64server:~$ hadoop fs -ls /Olive29
Found 13 items
-rw-r--r--   1 hduser supergroup          0 2016-11-21 13:36 /Olive29/_SUCCESS
-rw-r--r--   1 hduser supergroup     377449 2016-11-21 13:35 /Olive29/part-r-000
00
-rw-r--r--   1 hduser supergroup     339311 2016-11-21 13:35 /Olive29/part-r-000
01
-rw-r--r--   1 hduser supergroup     385895 2016-11-21 13:35 /Olive29/part-r-000
02
-rw-r--r--   1 hduser supergroup     368421 2016-11-21 13:35 /Olive29/part-r-000
03
-rw-r--r--   1 hduser supergroup     371798 2016-11-21 13:35 /Olive29/part-r-000
04
-rw-r--r--   1 hduser supergroup     368247 2016-11-21 13:35 /Olive29/part-r-000
05
-rw-r--r--   1 hduser supergroup     375554 2016-11-21 13:36 /Olive29/part-r-000
06
-rw-r--r--   1 hduser supergroup     374305 2016-11-21 13:36 /Olive29/part-r-000
07
-rw-r--r--   1 hduser supergroup     367955 2016-11-21 13:36 /Olive29/part-r-000
08
-rw-r--r--   1 hduser supergroup     368733 2016-11-21 13:36 /Olive29/part-r-000
```

Screen Shot 6.2: ***Output Window [Specific Month View]***

```
00002201,01-05-2015,4007645,022.13,Exercise & Fitness,Abdominal Equipment,Colora
do Springs,Colorado,credit
00006383,01-03-2015,4009779,076.99,Indoor Games,Air Hockey,Flint,Michigan,credit
00032500,01-05-2015,4006784,126.42,Indoor Games,Ping Pong,Eugene,Oregon,credit
00044032,01-18-2015,4004101,049.39,Dancing,Ballet Bars,Pittsburgh,Pennsylvania,c
redit
00044029,01-14-2015,4000585,128.11,Gymnastics,Balance Beams,Jacksonville ,Florid
a,credit
00044028,01-09-2015,4009478,183.65,Team Sports,Cricket,Reno,Nevada,credit
00032514,01-07-2015,4005181,049.23,Water Sports,Swimming,Lexington,Kentucky,cred
it
00016778,01-01-2015,4004887,164.83,Team Sports,Hockey,Jackson,Mississippi,credit
00006370,01-02-2015,4009772,055.41,Jumping,Bungee Jumping,San Antonio,Texas,cred
it
00032523,01-19-2015,4008207,022.03,Games,Portable Electronic Games,Durham,North
Carolina,credit
00032525,01-15-2015,4003481,150.91,Winter Sports,Cross-Country Skiing,Clarksvill
e,Tennessee,credit
00032527,01-02-2015,4003279,145.18,Exercise & Fitness,Weightlifting Machine Acce
ssories,Columbus,Georgia,credit
00011476,01-09-2015,4006467,158.87,Gymnastics,Balance Beams,Jersey City,New Jers
ey,credit
00006368,01-06-2015,4000492,124.45,Winter Sports,Bobsledding,Irving,Texas,credit
00032530,01-10-2015,4009194,113.97,Gymnastics,Gymnastics Rings,Bellevue,Washingt
```

Output Path:

Hadoop fs –cat /Olive29/part-r-00000

Use Case 7: Get the Profession

- **Input File**: txns-large.dat, Customer.dat
- **Input** : ID, AMT, PROFESSION
- **Key**: ID, **Value**: AMT & PROFESSION
- **Output** : Profession
- **Concept** : Mapper Side Join

Description: To Find the Profession of an user who has spent the Max Amount for the Transaction data

Screen Shot 7.1: Input Window

```
hduser@ubuntu64server:~$ hadoop fs -put /home/hduser/Customer.dat /Oliver
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task6_11.jar /Oliver/txns-large
.dat /Olive30
```

Screen Shot 7.2: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive30/part-r-00000
Pilot   1700.17
```

Output Path:

Hadoop fs –cat /Olive30/part-r-00000

Use Case 8: ***Top 3 Contributors***

- **Input File**: txns-large.dat, Customer.dat
- **Input** : ID, AMT, NAME
- **Key**: NAME, **Value**: AMT
- **Output** : NAME
- **Concept** : Mapper Side Join

Description: To find the top 3 Customers who has spent the MAX Transaction

Screen Shot 8. 1: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task7_11.jar /Oliver/txns-large
.dat /Olive31
```

Screen Shot 8. 2: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive31/part-r-00000
Karen   1080.42
Kristina        980.51
Elsie   719.66
```

Output Path:

Hadoop fs –cat /Olive31/part-r-00000

Use Case 9: ***Rock Star of the Month***

- **Input File**: txns-large.dat, Customer.dat
- **Input** : ID, AMT, NAME
- **Key**: NAME, **Value**: AMT

- **Output** : NAME [BASED ON MONTH]
- **Concept** : Mapper Side Join

Description: To find the User who has Spent MAX Amount in the month of July

Screen Shot 9. 1: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task8_Trans.jar /Oliver/txns-la
rge.dat /Olive32
```

Screen Shot 9.2: Output Window

```
hduser@ubuntu64server:~$ hadoop fs -cat /Olive32/part-r-00000
Karen   155.18
```

Output Path:

Hadoop fs –cat /Olive32/part-r-00000

Use Case 10: Take me to and From the Beginning

- **Input File**: txns-large.dat
- **Input** : AMT
- **Key**: AMT, **Value**: Entire line of a data
- **Output** : SORTED data based by Amount
- **Concept** : Simple Map Reduce

Description: To sort the output based on the Amount

Screen Shot 9.1: Input Window

```
hduser@ubuntu64server:~$ hadoop jar /home/hduser/Task10.jar /Oliver/txns-large.d
at /Olive33
```

Screen Shot 9. 2 : Output Window

```
00034188,02-26-2015,4006041,195.68,Combat Sports,Fencing,Madison,Wisconsin,credi
t
00021379,12-13-2015,4000223,195.68,Outdoor Recreation,Deck Shuffleboard,Cambridg
e,Massachusetts,credit
00020735,04-08-2015,4003036,195.68,Outdoor Recreation,Deck Shuffleboard,Boston,M
assachusetts,credit
00015706,10-13-2015,4007189,195.68,Winter Sports,Luge,Stamford,Connecticut,credi
t
00001412,06-12-2015,4006334,195.69,Water Sports,Water Polo,Fremont,California,cr
edit
00003451,01-30-2015,4003225,195.70,Racquet Sports,Racquetball,Oklahoma City,Okla
homa,credit
00001932,07-05-2015,4007228,195.71,Winter Sports,Downhill Skiing,Reno,Nevada,cre
dit
00004787,12-07-2015,4008646,195.72,Water Sports,Bodyboarding,Jersey City,New Jer
sey,credit
00025794,05-25-2015,4007111,195.72,Games,Poker Chips & Sets,San Francisco,Califo
rnia,credit
00007382,11-22-2015,4001367,195.72,Gymnastics,Vaulting Horses,New York,New York,
credit
00027119,06-22-2015,4001170,195.72,Exercise & Fitness,Exercise Balls,Seattle,Was
hington,credit
00046726,06-19-2015,4001182,195.72,Exercise & Fitness,Cardio Machine Accessories
,St. Louis  ,Missouri,credit
```

Output Path:

Hadoop fs –cat /Olive33/part-r-0000

Conclusion:

Thus, we have used the concept of Map Reduce to Analyze the real –time data.