**Title: - Online Transaction Analysis**

Project Author: - Sandeep Kumar Bhandoria

**OBJECTIVE:** I am a Hadoop Analyst in my company OTA Pvt Ltd. My company give the suggestions to other company who shares their transaction data with us.

This time they have given me the Online Transaction data of a company that is planning to surprise their customers for the events for Christmas and New Year and they also want to do some more suggestion so that they can make decision.

My objective is to analyse the data and come up which some use case and solution for that.

**DATA WE HAVE:**

1. Transaction’s Data:

File Used: txns-large.dat

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TransactionID | T\_date | UserId | Price | Product\_Cat | Product |
| 00000000 | 06-26-2015 | 4000003 | 040.33 | Exercise & Fitness | Cardio Machine Accessories |
| 00000001 | 06-01-2015 | 4009775 | 005.58 | Outdoor Recreation | Archery |

2. Customer’s Data:

File Used: Customer.dat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| UserID | FirstName | LastName | Age | Profession |
| 4000001 | Kristina | Chug | 55 | Pilot |
| 4000002 | Paige | Chen | 74 | Teacher |
| 4000003 | Sherri | Melton | 34 | Firefighter |
| 4000004 | Karen | Puckett | 74 | Lawyer |
| 4000005 | Elsie | Hamilton | 43 | Pilot |

**TECHNOLOGY WE USED:**

1) Apache Hadoop

2) Map Reduce programming in Java

**SOFTWARE WE USED:**

1) Virtual Box

2) Eclipse

3) Ubuntu

**PROJECT DESCRIPTION:**

**Use case 1**

**Scenario: - Heavy price based transactions that company have.**

1) We find all the transaction or products based on the user defined prices.

In the case we are expecting input from user for the value of amount on which we have to decide the transaction.

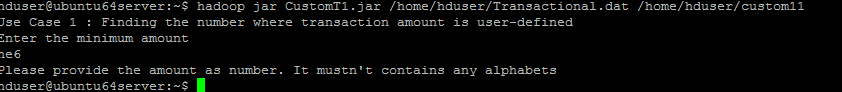
2) This can be used to find the transaction done on a specific price from where we can get products name that the users are interested in for a specific price.

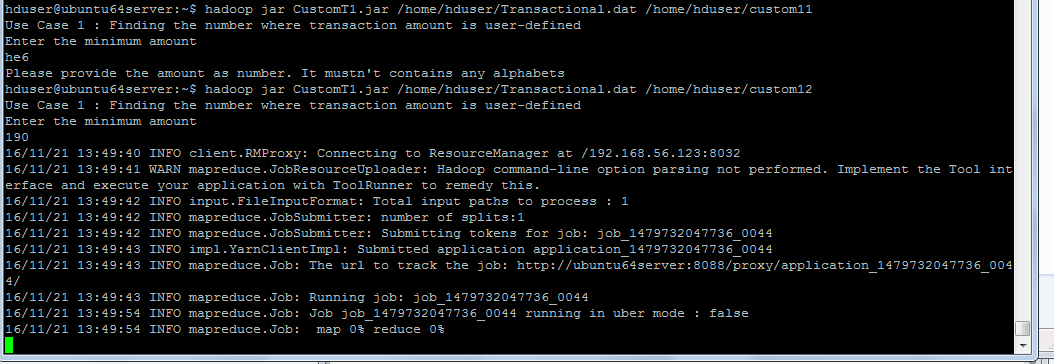
**Validation:**

We have done a check on the user input before processing it further.

1) User can have to specify a minimum price and based on that price all transaction will be filter where price is greater than what user has specified.

2) If the user is passing String in place of number he/she will be displayed a message showing an error message to provide valid input and start the job again.





**Use case 2**

**Scenario: - Price Range Based Products**

1) We will find the number of products we have for a particular range of price.

**Validation:**

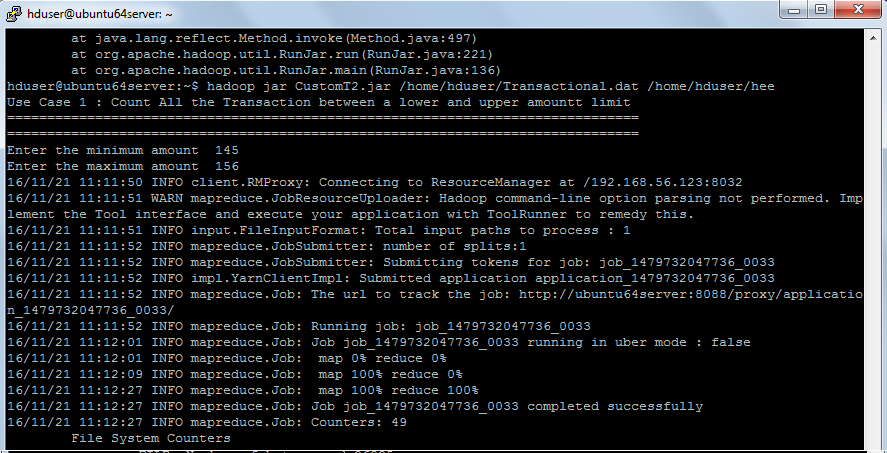
1) We will be accepting user input for minimum and maximum limit for price.

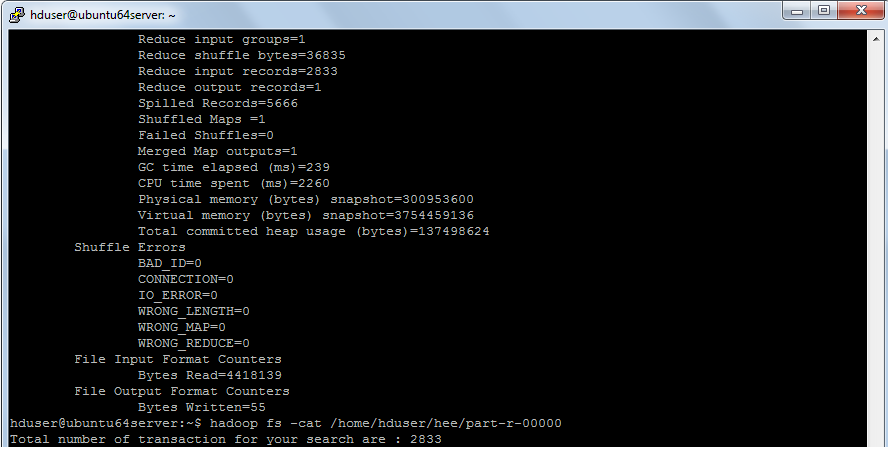
2) The maximum price can’t be less than minimum price we user pass inputs. The will be showed a message for this. And also tells the user to run the task again with proper inputs.

3) Minimum amount can’t be less than 0. Message will be displayed for the same.

4) Maximum amount can’t be less than 0. Message will be displayed for the same.

**Output screenshot: -**





**Use case 3**

**Scenario: - Customers wise transaction and purchase**

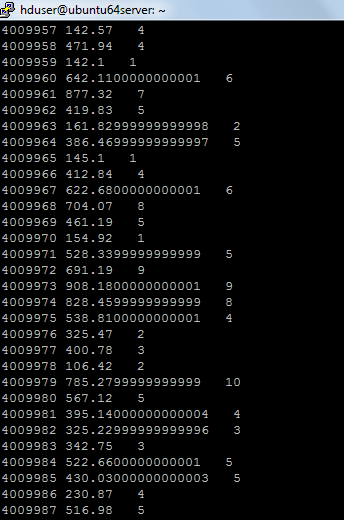
1) The Company is planning for a scheme to give offers to customers based on

a) Their past number of transactions.

b) Total purchase they did.

This requires an analysis to prepare a report per each user.

**Output screenshot: -**

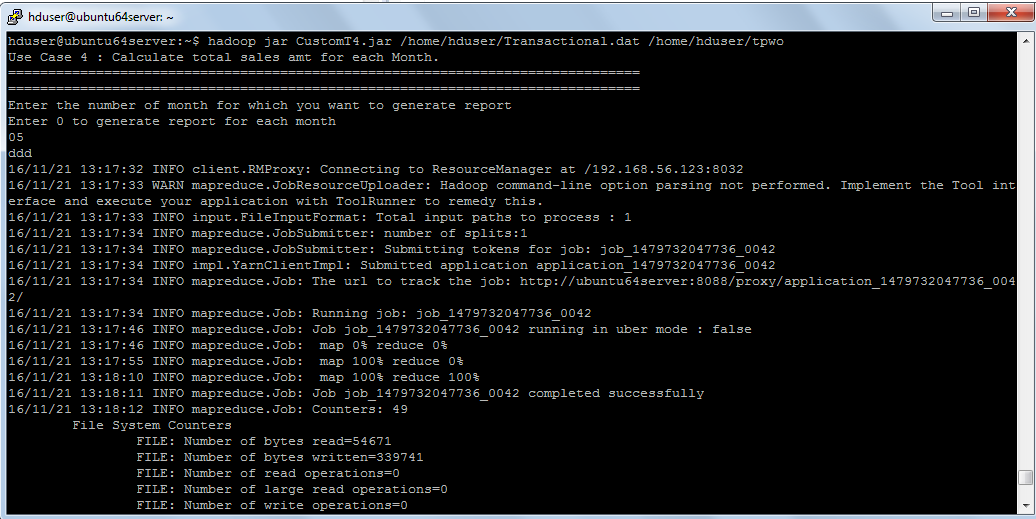


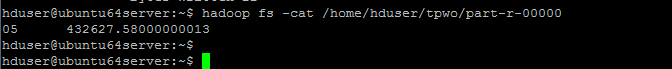
**Use case 4**

**Scenario: - Monthly Wise Revenue**

At the end of every year your company wants to do an analysis to know in which month people usually comes for shopping.

**Output screenshot: -**



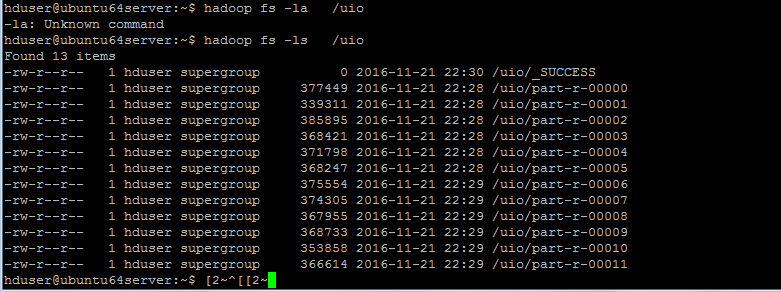


**Use Case 5**

**Scenario: - Monthly Wise Transaction Summary**

Being a Hadoop Developer and Admin, You may need to partition your final data to make further processing easy.

We have been asked to divide all the transaction based on the month and store each transaction according to the months. So12 files are created for this one for each month.

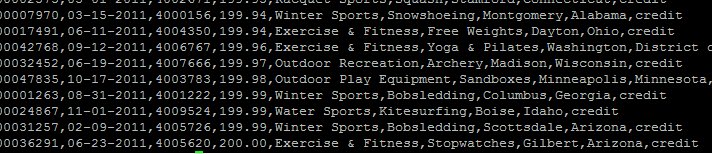


**Use case 6**

**Scenario: -File sorting based on price**

We have the transaction file and this file will be sorted based on the amounts available in each transaction.

**Output screenshot: -**

****

**Use Case 7**

**Scenario: - Top profession who does shopping the most**

1) Company wants to target the particular area where people are more interested in their products so we have analysed the top profession.

**Output screenshot: -**



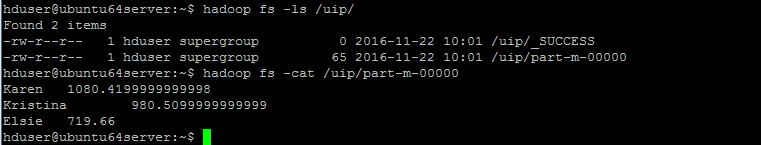
The customers who are pilot are doing more transactions.

**Use Case 8**

**Scenario: -Analyze Top 3 customers to give additional rewards.**

Our online shopping website wants to give rewards to some top 3 customers.

**Output screenshot: -**

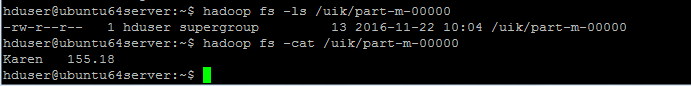


**Use Case 9**

**Scenario: - Month Wise top customer**

1) We have analysed the data to get the top customer for a specific month July.

**Output screenshot: -**



Karen is the top customer who spent the most for online shipping.

**CONCLUSION -** Above Data Analysis shows that we can get various information using map reduce Hadoop processing to make better decision in E-commerce Industry which will help the website owner in providing better service for their customers.