

Customer Insights of Omni-Channel Commerce

Divyansh Saxena
Undergraduate Student
Computer Department
V.E.S. Institute of Technology

Vaibhav Malpani
Undergraduate Student
Computer Department
V.E.S. Institute of Technology

Vikramank Singh
Undergraduate Student
Computer Department
V.E.S. Institute of Technology

Mrs Lifna CS
Assistant Professor
Computer Department
V.E.S. Institute of Technology

Abstract – Analytics is a powerful tool that helps in taking profit making decisions in any industry. This paper discusses a methodology to build end to end analytics tool for the ecommerce industry. Selling products requires marketing and thus investment. Using analytics, this marketing can be done in a much directed manner thus reducing the investment and increase the sales. This technique is used and in much demand by marketing heads of various ecommerce websites including the top ones like Amazon and Flipkart. Technique mentioned here describes a process using which a marketing head of any organization can upload its customer data and see different charts on a web or mobile application. In this paper, we have particularly discussed about a use case of ecommerce websites.

I – INTRODUCTION

Customer insights are key ingredients of ecommerce analytics. These insights are created by analysing the customer data, interpreting it and converting it into charts and reports that is displayed on a mobile or web application. End to end analytics software has many parts, right from getting data, transforming data to displaying result. This particular technique establishes a connection with the ftp location where the customer data has been uploaded by the ecommerce. This ftp location can be a Salesforce location, Amazon server location or the local host of the ecommerce system. Whatever the ftp location maybe, first task is to establish a connection with it. Second step of this technique involves the usage of an ETL module. This module is used to extract the customer data from the ftp location, transform the data according to the requirement for performing analytics and loading the transformed data into a database. The last step of this technique is to use the transformed data stored in the database to create the

graphs and reports to provide customer insights to the marketing head of an ecommerce. After generating these graphs and reports, they are supposed to be displayed on a mobile or web application where the insights can be easily accessed and thus important decisions can be made easily. Instead of having two steps of generating reports and displaying the results, there are business intelligence tools that combine the steps into one. There are many open - source as well as enterprise business intelligence tools that should be chosen depending upon the need of marketing head.

II – EXTRACTING THE DATA

Analytics is providing insights of data. Hence, retrieving data from the source is the very first and an important step of the process. FTP (File Transfer Protocol) is a standard network protocol used to transfer computer files between a client and a server on a computer network. In this technique, FTP is used and a connection is established between the client and server system using XAMPP control panel. After establishing the connection with the FTP location, next step is to use ETL module to extract the data from the location. The technique mentioned here involves the usage of an open source ETL tool known as Talend OpenStudio Big Data. Talend OpenStudio is umbrella software that has different versions depending upon the requirement of the user. This particular version is different from other versions because this technique involves loading the data into a NoSQL database. This ETL module helps to develop a java based job that can be run through any website or as a standalone job on a PC. The way it works is that, the user is supposed to use the components that are listed in the software according to the development requirement. These components are connected by the appropriate action that has to be performed, also in an order of action. Below

mentioned is an example of a simple job in Talend Open Studio Big Data.

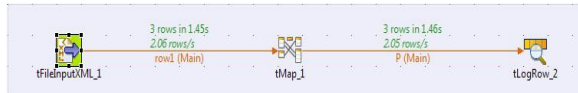


Fig 1 – Sample Talend Job

The above example is a simple job made via Talend Open Studio which takes XML input by tFileInputXML component and displays the result using tLogRow component row wise. This XML input is transformed to some other values using tMap component and these changed values are displayed. For customer insights, similar job is to be designed using this open source ETL tool. The customer insight job is divided into two jobs – Main and Sub Job. Main job is responsible for doing the extraction part of the ETL module.

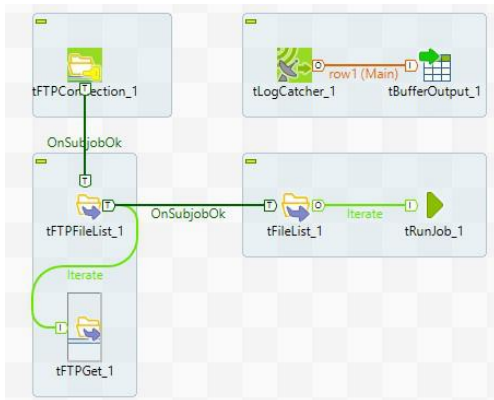


Fig 2 – Main Job

The above figure depicts the main job designed in Talend OpenStudio for extracting the data from FTP location. The first component is tFTPConnection that establishes a connection between ETL and FTP location. Next component in sequence is tFTPFileList which displays the list of all files with a particular extension which can be explicitly mentioned in the properties of this component. Another component used sequentially is tFTPGet. This component plays an important role in the design as it extracts the data from the selected file from the file list got from previous component. As there can be more than one file that are to be extracted from the FTP location, there is a need of a component that will process a particular file and keeping other files into a queue. Talend OpenStudio provides a component for the same process that is tFileList. After getting the file from FTP location provided by the ecommerce, the

process of getting or extracting data is completed. Few extra components are used in the design like tLogCatcher and tBufferOutput that are responsible for displaying the errors, if any during the extraction process.

III – TRANSFORMING THE DATA

After acquiring the appropriate data, next step is to manipulate and transform the data according to the need of insights. For this transformation, another job is designed which is named as Sub Job and is connected to the main job. The data extracted via the Main job is transferred to this job using the component tRunJob. Now, the command of the process is transferred to the Sub job which is responsible for transformation.

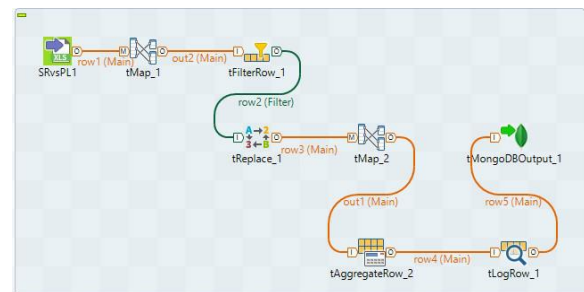


Fig 3 – Sub Job

The above figure depicts the sub job designed in Talend OpenStudio for transforming the data. Every different customer insight will require different transformation design. This particular job is designed to get a insight of number of customers via Search Result of the browser versus the number of customers via Product Listing on the ecommerce website itself. This insight is very important for any ecommerce as it shows which area to focus for marketing. If search result numbers are more, then website filtration should be improved or vice versa. To enable this kind of insight, an appropriate job design is to be built. The first component of this job design is tFileExcelInput which will load the data. Next component is tMap which is responsible for some filtration using regular expressions. These results in a list of rows that have products bought using search result or product listing. Rest other rows are filtered out using another component tFilterRow. Next component is tReplace which is used to replace the columns or add or delete columns. Using this component, two columns are added to the data –

Search Result and Product Listing. The values of these columns are Boolean depending upon the result. Use tMap to convert these Boolean values to integer 0 and 1 values for addition. The next component is tAggregateRow which is the key component of this design. It calculates the number of products sold using search result and product listing date wise by aggregating the data using the Date column. Final component is tLogRow using which the results are displayed. These results are to be stored in a NoSQL database for which tMongoOutput is used and discussed in next section.

IV – LOADING THE DATA

The transformed data that is achieved using the sub job design is to be stored into a NoSQL database as it is very flexible and scalable as compared to SQL database. In this technique, MongoDB is used as the NoSQL database. Hence, there is a need to establish a connection between Talend and MongoDB. Before this connection, another connection is to be established, that is, between MongoDB and PC. To do this, go to Command Prompt and change the directory to the one with MongoDB folder. Then type, mongod. This will establish a connection between MongoDB database and PC on port 27017.

```
C:\Program Files\MongoDB\Server\3.0\bin>mongod
2016-01-31T21:44:13.575+0530 I JOURNAL [initandlisten] journal dir=C:\data\db\j
urnal
2016-01-31T21:44:13.578+0530 I JOURNAL [initandlisten] recover : no journal fil
es present, no recovery needed
2016-01-31T21:44:13.639+0530 I JOURNAL [durability] Durability thread started
2016-01-31T21:44:13.641+0530 I JOURNAL [journal writer] Journal writer thread s
tarted
2016-01-31T21:44:13.799+0530 I CONTROL [initandlisten] MongoDB starting : pid=
78348 port=27017 dbpath=C:\data\db\ 64-bit host=Divyansh_Saxena
2016-01-31T21:44:13.801+0530 I CONTROL [initandlisten] targetMinOS: Windows 7/W
indows Server 2008 R2
2016-01-31T21:44:13.802+0530 I CONTROL [initandlisten] db version v3.0.6
2016-01-31T21:44:13.803+0530 I CONTROL [initandlisten] git version: 1ef45a23ad0
5e3480ac919b28afcbac615488f2
2016-01-31T21:44:13.804+0530 I CONTROL [initandlisten] build info: windows sys.
setwindowsversion(major=6, minor=1, build=7601, platform=2, service_pack='Servic
e Pack 1') BOOST_LIB_VERSION=1_49
2016-01-31T21:44:13.805+0530 I CONTROL [initandlisten] allocator: tcmalloc
2016-01-31T21:44:13.805+0530 I CONTROL [initandlisten] options: {}
2016-01-31T21:44:17.562+0530 I NETWORK [initandlisten] waiting for connections
on port 27017
```

Fig 4 – Establishing MongoDB Connection

Next step is to establish a connection between MongoDB and Talend. This is done using the component known as tMongoOutput. This is configured to the appropriate database and collection where the data has to be stored. This configuration is done in the properties section of the component. After running the job, the transformed job is stored in MongoDB in the predetermined database and collection. Now that, data is stored in the database, a BI tool is to be used to create required insights.

V – CREATING INSIGHTS

Now that the data is stored in MongoDB, it can be used to create insights, reports and charts. There are two ways to achieve this. One can be to analyse the data using some statistical tools like R and then displaying the result on a web or mobile application. Another and more efficient way is to create insights and displaying results via single software itself. Such softwares are commonly known as Business Intelligence Tool. In this technique, Microstrategy tool is used to create the charts. First step is to establish a connection between MongoDB database and Microstrategy so that the data stored in it can be retrieved by the BI Tool and further steps can be taken. The data retrieved by Microstrategy is stored in a cache memory called as Intelligent Cube. Now, once the data is retrieved, charts and reports can be created using the inbuilt features and finally be displayed onto Web and Mobile Applications. Here is the output graph –

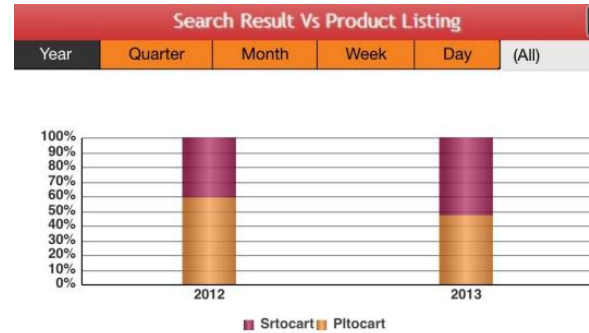


Fig 5 – Customer Insight

VI – REFERENCES

- [1] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey MIS Quarterly Vol. 36 No. 4, pp. 1165-1188/December 2012
- [2] Nikhilesh Dholakia, Dong Zhang Qualitative Research in the Age of E-Commerce: Data Sources and Approaches QUALITATIVE SOCIAL RESEARCH SOZIALFORSCHUNG Volume 5, No. 2, Art. 29 May 2004
- [3] Satkaur, Anuj Mehta scope of ETL in retail domain IJARCSS Volume 3, Issue 5, May 2013 ISSN: 2277 128X
- [4] Li Jian, Xu Bihua Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on Volume 6