

Customer Complaint Analysis Using Hadoop (Consumer Analysis)

¹Y. Sudha Laxmi, ²A. Mahendar, ³Dr. Sheikh gouse

^{1,2,3}Department of IT
MLRIT

Hyderabad, India.

sudhasony6@gmail.com

mahe8079@gmail.com

Abstract: Customer complaint Analysis is Handling Customer Dissatisfaction. It may accompanies Web customer complaint management, This may be a critical Issue for the Customer. In this paper we are going to Investigate the current sources of customer complaints and what are the causes of complaints and want to seek the effective ways of handling customer complaints by examining different type of products and Issues. For achieving customer complaint analysis we are using Big data platform for storing of huge amount of Data. These data are stored in the form of traditional Database. Big data mainly focus on the storage of data. It majorly focuses on three attributes they are Data Volume, Data Velocity and Data Variety.

Keywords: Hadoop, Big Data, Cloudera, Hue, Impala

1. Introduction

Gaining the attention of Business managers who want to increasing repeat business and customer loyalty customer complaints plays a vital role in the development of product. The major issue that mainly focuses on Business is the quality and quantity of the product. If the product that does not reach the customer satisfaction it leads to the Dissatisfaction of the customer. To overcome this type of problem we are introducing customer complaint analysis . In the customer will give the product feedback and this leads to the development of the product.

The major key components that have to be discussed in this project are as follows:

- Maximizing the customer Satisfaction and Minimizing the customer Dissatisfaction
- Increasing the product and the service quality
- Increasing Customer Loyalty
- Resolving the customer complaints

i) Maximizing the customer satisfaction

Customer satisfaction/dissatisfaction is determined by how much the customer's expectations differ from the product's or service's.

According to the proposed model of customer complaining behavior, customer dissatisfaction results from unmet expectations about the product and the issue; which may include the product and the information content. Similarly , according to customer metrics, customer satisfaction primarily depends on time, speed , product and service.

ii) Increasing the product and service quality

For the firm's successful complaint management requires quality assurance i.e., QA. It provides feedback in order to improve customer's overall perception of product and service quality. Service quality is generated from feedback mechanisms that serve as Intermediate for Web-based Information markets (or) how product or service quality is evaluated in online business.

iii) Increasing Customer Loyalty

In this it states the establishing effective relationship results in greater customer loyalty and improved data on customer usage. It encourages the use of complaints to improve communication channels between buyers as well as sellers.

iv) Resolving customer complaints

The proper management of customer complaint has a direct impact on customer retention. Complaint management refers to different type of strategies such as to resolve disputes and to improve

ineffective products or services in order to establish a reliability in the eyes of customer.

Complaint data is a key component in the process of problem correction and increase the performance. Customers may file complaints directly to the company product or service is purchased.

Based on these four components, the goals are explore the emerging virtual sites for customer complaints and identifying the common reasons for complaints. All these are analyzed using Hadoop file systems.

All the analysis that are done with the use of online services customer complaint analysis have been analyzed so far. But it take more time for the analysis and it will affect the profits of the organizations. All these problems are faced in Existing System. To overcome these problem we are using Big data.

2.LITERATURE SURVEY

In the year 1979 the author named Richins has fragmented the area of complaints and cohesiveness. The process of complaint management functions is evaluated in various forms.

Later in the year 1990 by the authors Hart and sasser introduced complaint Handling. This refers to the strategies firms use to resolve and learn from the service failures in order to establish the customer reliabilities.

Complaint Handling can be viewed as a sequence of events in which a procedure, beginning with communicating the complaint, generates a process of interaction through the decision and outcomes occurs in the year 1998 by Tax, Brown and ChandraShekaran.

3.PROPOSED SYSTEM

As some of the limitations faced in the online services we are using Big data. In this large amount of Data sets can be analyzed easily. Using different tools such as Hive ,Hue,

Cloudera, Impala, Apache Hadoop we can analyze data easily which are used for future enhancements. To know the speed of tools that are part of the Hadoop Ecosystem i.e., hive and Impala. We are trying to know the process of Execution that is executing with high speed.

Our proposed system aimed to focus on main elements of customer complaints, which are date

-received, Product, Sub-product, Issue, Sub-Issue, State, Complaint-Id and so on.

4. REQUIREMENT ANALYSIS

After analyzing the requirements certain amount of tasks that have to be performed, the next step is to analyze the problem and understanding the context. There are majorly two phases that are present. The first activity in the phase is studying the existing system and the other is to understanding the requirements of the new system.

Understanding the properties and requirements of a system is more difficult and requires creative thinking and understanding of existing systems.

Requirements which we are using in this project include both Hardware and software requirements for the process.

(a)Hardware Requirements

- Dual Quad-core CPU
- 4-8 GB of memory per processor core
- 1 Gigabit Ethernet

(b)Software Requirements

- Hadoop 2.x
- MySQL
- VMware
- HDFS
- Hive
- Hue
- Impala.

5. METHODOLOGY

In this project, we explain how we prepared our datasets. After that, we provide how we analyzed the data using some statistical analysis. Then, we introduce how we constructed some of the models to achieve our purpose.

a)How HDFS used in project:

HDFS is a java based file system that provides scalable and reliable data storage and it was designed to span large clusters of commodity servers. The quality and quantity of enterprise data is available in HDFS. It is scalable, fault-tolerant, distributed storage system that works closely with wide variety of concurrent data applications.

HDFS cluster is comprised of Name node manages the clusters in metadata and Data Nodes that stores the data. These Attributes like

- Permissions
- Modification
- Access time
- Name space

5.2 Hive is used in project

The data which is present in MySQL is imported to Hive using Sqoop command. Steps that are involved in hive are,

- Start installation
- Preparing to use a MySQL streaming Result set.
- Beginning code generation.
- Transferring of Data
- Retrieving of Records
- Execute SQL
- Loading Uploaded data into Hive.

5.3 Cloudera used in Project

Cloudera is revolutionizing enterprise data management by offering the first unified platform for Big data.

CDH contains the main, core elements of Hadoop that provides reliable, scalable data for processing of large datasets. It provide security, high availability, and integration with Hardware and software.

It contains single Hardware and Software system

- Single Management Model
- Single Security Model
- Common Storage
- Single Meta Data

5.4 Hue used in project

Hue includes web Applications that let you browse HDFS and H Base files, write apache Hive and Impala queries, Export data with Apache Sqoop, Submit Map Reduce programs.

Hue is used to get results in different types of charts like pie charts, Bar charts in the result sets. This is used to analyze the results easily.

5.5 Impala used in project

- Impala gives parallel processing Database technology on top of Hadoop eco-system. It allows users to perform low latency queries interactively.
- Hive Map Reduce job will take some minimum time in launching and processing

of queries where as impala gives results in seconds.

- Impala being real-time query engine best suited for analytics and for data scientists to perform analytics on data stored in Hadoop File System.

5.6 Sqoop used in project

Sqoop is a command line interface application for transferring data between relational data base and Hadoop.

It Supports incremental loads of a single table or a free from SQL query as well a saved jobs which can run multiple times to import updates made to a database.

It supports both Sqoop import and Export.

6.IMPLEMENTATION

Collection of Data sets according to customer complaints like product, Sub-product, Issue ,Sub-Issue, State, Company Name, Zip-code, Date-received and so on.

These are different types of attributes which are collected from the various Organization. For e.g., for each complaint there is a unique Zip-code is present and the complaint-Id is present and Data-received and State.

To implement customer complaint analysis the following steps are to be followed.

Step1: Collect the data sets of customer complaint Analysis.

Step 2: Format the data file in windows by removing the header.

Step 3: Convert the database into CSV format and copy the file into Cloudera.

Step4: Create the Data base with the file name and use the database.

Example: create database table name;

Use table name;

Step 5: Creating the tables un Sequential Query Language i.e., SQL

Step 6: Loading the data into MYSQL.

Step 7: Importing Data from MYSQL to Hive using Sqoop.

Step 8: Check the tables in database.

Step 9: Execute the commands in Impala for getting results.

7. RESULTS

After execution of Impala command:

Enter into the Browser and then select Hue Impala in the query Editor and write then write the command for Execution

For example when the limit is 20 then the query is shown as

Select * from table limit 20;

Example: select * from customer complaints limit 20;

From the above figure 7.1 shows the results that are present in customer complaints with the help of Impala tool with respect to the product

	date_received	product	sub_product	issue
1	5/16/2014	Credit reporting		Incorrect information on
2	11/13/2015	Debt collection	Payday loan	Cont'd attempts collect d
3	10/3/2014	Credit card		Credit line increase/decr
4	12/7/2014	Credit card		Identity theft / Fraud / Em
5	1/28/2015	Credit reporting		Incorrect information on
6	7/22/2013	Credit reporting		Unable to get credit repo
7	10/6/2015	Credit card		Other
8	12/18/2014	Debt collection	Credit card	Communication tactics
9	12/19/2014	Debt collection	Other (i.e. phone, health club, etc.)	Communication tactics
10	3/24/2015	Credit card		APR or interest rate

Fig.7.1 Results in Impala

To see the results in the form of charts such as Bar-graphs or pie-charts we can analyze the data with the help of X-axis and Y-axis.

From the above figure.7.2 it shows the results in Bar-graphs. The results is based on Zip-code and date-received. Date-received is taken on X-axis where as zip-code is given on y-axis.

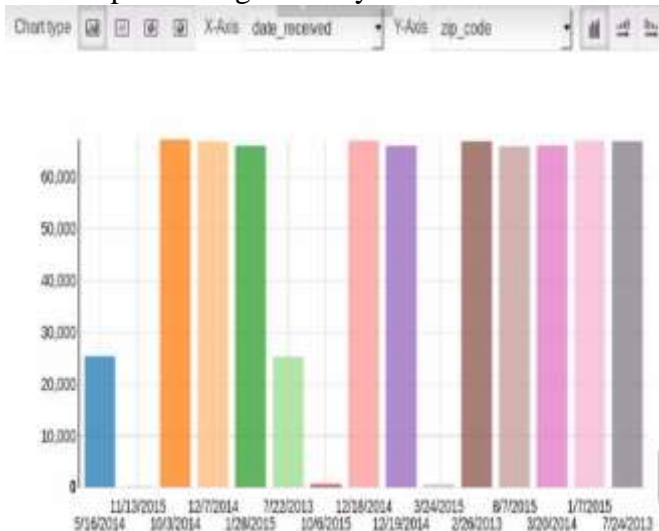


Fig.7.2 results in Bar graphs

Similar to the Bar-graphs we can create a pie-charts with the following Attributes.

From the above fig.7.3 the results are shown in the form of pie-chart with the attributes date-received and zip-code. Where as zip-code is unique to each and every product.

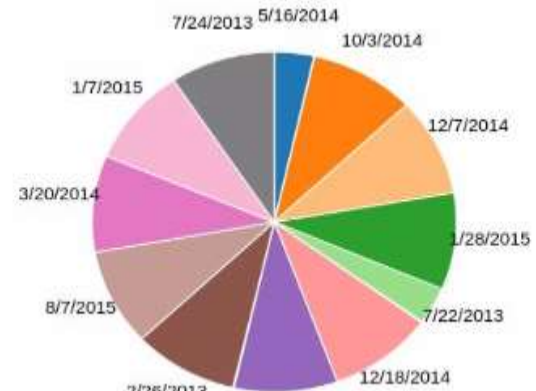


Fig.7.3 Results in pie-charts

8. CONCLUSION

Customer complaint analysis are important to find and there's no better way to collect direct feedback from the customers and improve your product or service. However, the way you Handle a complaint is the difference between keeping a customer or losing one. So, the next time you receive a customer complaint, listen to what the customer has to say, apologize find a solution and follow up to see if he or she is happy with the way you are handling it.

As earlier loading large amount of data is very difficult. By using Big data complexity of loading large amount of data can be reduced. The proposed tool enables agencies too easily and economically clean, characterize and analyze the data to identify actionable patterns and trends.

9.REFERENCES

[1] Anderson, R. E., "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance," Journal of Marketing Research, February 1973, pp. 38-44.
 [2] Barbara, S., "Consumer Complaint Handling as a Strategic Marketing Tool," The Journal of Consumer Marketing (2:4), Fall 1985, pp. 5-17.
 [3] Betrand, K., "Marketers Discover What 'Quality' Really Means," Business Marketing (72), 1987, pp. 58-72.
 [4] Blodgett J. G., Donald. H. G., and Walters, R. G., "The Effects of Perceived Justice on Negative

Word-of-Mouth and Repatronage Intentions,”
Journal of Retailing (69), Winter 1993, pp. 399-428.

[5] Cho, Y., Im, I., Ferjemstad, J., and Hiltz, R., “Causes and Outcomes of Online Customer Complaining Behavior: Implications for Customer Relationship Management (CRM),” *Proceedings of the 2001 Americas Conference on Information Systems*, Boston, August 2001.