International Journal of Computational and Applied Mathematics. ISSN 1819-4966 Volume 12, Number 1 (2017)

Evaluation of Telecom Data Analysis Using Business Intelligence

A. Devibala[1], R.Murugesan[2], S.Sivakumar[3]

[1] Research Scholar , Department of Computer Science, CPA College, Bodinayakanur.
[2] Associate Professor , Department of Computer Science , CPA College , Bodinayakanur.
[3] Associate Professor and Head, Department of Computer Science , CPA College, Bodinayakanur.

Abstract

Nowadays the telecom industry is sitting on Terabytes of data. It's a time to dig the data and get some valuable information. In order to ensure a smooth path in customer journey, without any friction points it is essential to estimate customer effort. Business Intelligence (BI) is a set of concepts and methods to improve business decisions. In this research work Apache scope and Apache Flume were used for Data ingestion and Apache Oozie is used for scheduling the jobs. Hadoop five node clusters was installed and for data preprocessing the Apache Hive were using. Our application will be very fast, it is very useful to do analytics for real time data. Our framework is a High end Hadoop cluster, which is reliable, scalable and robust in nature, every process is automated and with alert mechanism. The output is beautifully visualized with IBMCognos. To identify the degree of effort that the customer had to exert in order to get an issue resolved, a request fulfilled, a product purchased or returned. The analysis and analytics are achieved by using the big data application and the outputs are neatly showcased with the reports.

Keywords: Data Analysis, Business Intelligence, Hadoop, Apache sqoop, Apache Flume, Apache Hive

I. INTRODUCTION

Telecom data are vast data and fast growing data [1]. But companies want to get sense of data. Through customer service applications the data is recorded and find out the metrics on the data. Telecom companies want to maximize the business revenue with innovative thoughts and high potential value. They have a vast volume of data daily coming into their organization and leverage on significant insights to get value of data [2].Dig the data and get some knowledge from the data to improve the business. Every telecom operator is always finding newer ways to increase the revenues during a time of stagnant development in the business, through the latest technologies. That is the reason operators trying to make inroads with advances big data technologies and methodology which sets up a business.

In telecom industry BI is used to monitor, analyze and provide performance on sales by product, region, distributor, partner, or sales representative using personalized BI dashboards and intuitive business intelligence reports [2].We need to turn big data into valuable data with quick, easy access to Hadoop and the ability to load to and from relational data sources as well as datasets [3].

Current research in Indian cellular market is focused with customer behavior and it is aggregated and analyzed to gain customer mind map, enabling each business to help make better and quicker business decisions [3]. This information is used by the cellular businesses for direct marketing and customer relationship management. Customer analytics plays a very important role in predicting customer behavior and shaping future customer interactions [4].

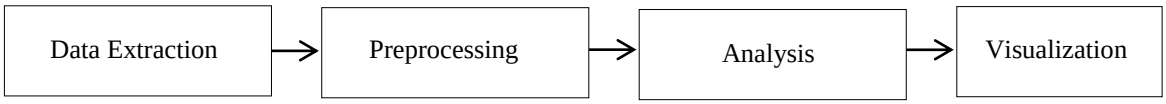

Fig.1 Methodology of telecom data analysis

In this research paper Apache sqoop and Apache Flume were used for Data ingestion and Apache Oozie is used for scheduling the jobs. Hadoop five node clusters were installed and for data preprocessing the Apache Hive were using [4]. For data visualization IBM Cognos was using when the client wants to see the reports so the entire architecture will work from the Data source to IBM Cognos.

In today business, meeting customer satisfaction is a must. The organizations are becoming more and more conscious about the advantages of data and information kept in their organization, the need to integrate these large volumes of data and to utilize this information to support the quality of their decision-making, in order to stay at a competitive advantage and to increase profit.

II. SYSTEM MODEL

We need to turn big data into valuable data with quick, easy access to Hadoop and the ability to load to and from relational data sources as well as datasets. Now business users can profile, transform and cleanse data on Hadoop or anywhere else it may reside using an intuitive user interface. Fig.2 illustrates the architecture of the proposed big data application for telecom.

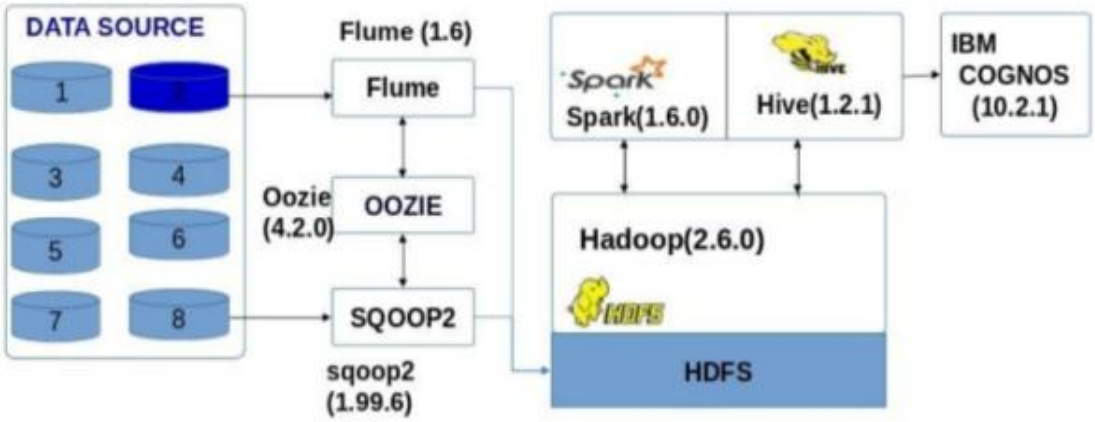

Fig.2: Architecture of big data application for telecom.

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. . It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs. The promise of low cost, high availability storage and processing power has drawn many organizations to Hadoop.

The entire project is divided into four modules first module extract the data from the client source IE. (Extract, Transform and Load)ETL layer, ETL layer will extract and transfer and load the data from client data source to our local HDFS data source. The second module is consolidating the data IE and preprocessing the data according to our requirement. The third module is analysis the data and does analytics on that data. The fourth module is visualized the data by using IBM Cognos.

Data Source: Our client is a telecom operator so most of the data are in structure format only so our client databases are RDBMS (relational database management system), except click-stream data IE. The data are in log format is unstructured. We will import the data like interactive voice response (IVR), automatic call distributor (ACD), and data. So for data extraction, both Flume and Scoop were used Flume for unstructured data and scope for structured data.

Sqoop Extraction: Sqoop is used to extract the data from any RDBMS to HDFS or vice versa, in our case we will extract the data from client data source to our local data repository i.e. HDFS. Through Sqoop will extract the data called IVR, ACD data. Sqoop have a good feature called while extraction the data we can transform the data for our requirement. Sqoop provides a pluggable mechanism for optimal connectivity to external systems. The Sqoop extension API provides a convenient framework for building new connectors which can be dropped into Sqoop installations to provide connectivity to various systems. Sqoop itself comes bundled with various connectors that can be used for popular database and data warehousing systems.

Flume Data Extraction: Flume is used to extract the data from client log directory to our local HDFS. In our case we will extract the click stream log data and Ingestion to our local HDFS. Flume has many agents for transferring the data here we are using the SpoolSource. First, we have to mount the directory in configuration. So when the data is ingested to that particular directory that will automatically copy into our local server in checkpoint directory. From that checkpoint directory the data are automatically copied to HDFS and the file extension will change. We can easily find out which file is copied properly into the HDFS. Once the server get down the data copping will stop and once the server get on it will automatically copy the data to HDFS. A flow in Flume starts from the client. The client transmits the event to a source operating within the agent. The Source receiving this event then delivers it to one or more channels. One or more sinks operating within the same agent drains these channels. Channels decouple the ingestion rate from drain rate using the familiar producer consumer model of data exchange. The sink of one Agent can be chained to the source of another agent. This chaining enables the creation of complex data flow topologies.

Oozie Scheduler: Apache Oozie is a workflow scheduler system to manage Apache Hadoop jobs. Oozie Workflow jobs are directed a cyclical graph (DAGs) of actions. Oozie Coordinator jobs are recurrent Oozie Workflow jobs triggered by time (frequency) and data availability. Oozie is a scalable, reliable and extensible system. In our Project Everything is automated for our requirement data used to ingest for every 2 hrs. , 24hrs like that so the jobs were scheduled accordingly. The Oozie have the xml pages we have to configure and that should be placed in HDFS the Oozie should read the file from HDFS only.

Apache Hive data warehousing: Apache Hive is used for data warehouse and consolidation the data i.e., preprocessing the data. Hive supports analysis the large amount of structured data. Hive has the SQL like language called HiveQL. In hive external tables, partitioning, clustering, group by like queries were used for preprocessing. For fast querying the spark engine were used.

High Availability Hadoop Cluster: In a typical HA cluster two separate machines are configure as name nodes. At any point in time, exactly one of the name nodes is an active state, and the other is in a standby state. The active name node is responsible for all client operations in the cluster, while the standby is simply acting as a slave, maintaining enough state to provide a fast failover if necessary.

In the event of a failover, the standby will ensure that it has read all of the edits from the journal nodes before promoting itself to the active state. This ensures that the namespace state is fully synchronized before a failover occurs. In order to provide a fast failover, it is also necessary that the Standby node have up-to-date information regarding the location of blocks in the cluster. In order

to achieve this, the data nodes are configured with the location of both NameNodes, and send block location information and heartbeats to both.

IBMCognos: IBMCognos is a web based visualization tool. It provides a tool set for reporting, analysis and monitoring of events and metrics. IBMCognos is integrated with hive by using JDBC connector. IBM Framework Manager is used to build the packages for processed data and allows you to add additional logic and details. IBM Report Studio is used to build the reports and make the individual queries. We can edit and enhance the reports individual.

III.RESULT AND DISCUSSION

Automatic Call distributor (ACD) Call Pattern Analysis: The metrics qualified in this work depict the ACD call preference rate of the customers across Circles, Hubs, etc. there by pinpointing to areas of lower Interactive Voice Response (IVR), efficiency. ACD call preference rate is defined on IVR calls getting transferred to agents within a short time interval of less than 10 seconds. The Operator can easily identify the IVR, call transfers for circle and Hubs wise.

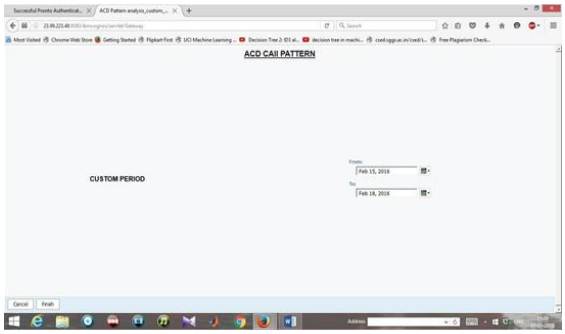

Fig.3: custom period menu selection

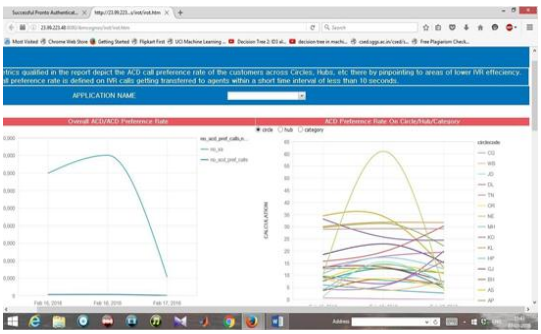

Fig.4: ACD preference rate

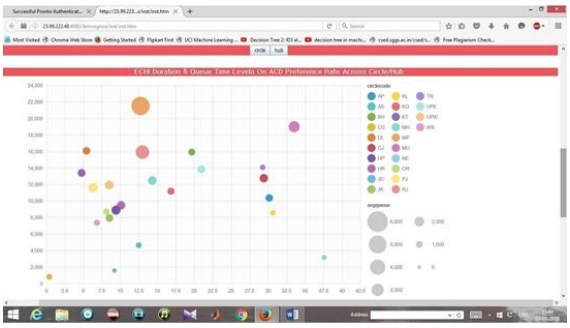

Fig.5: Bubble chart for ECHI Duration and Queue time Queue time.

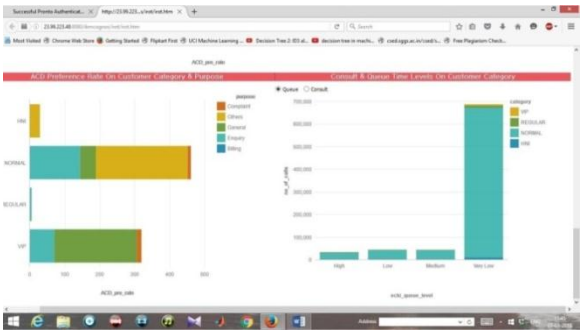

Fig.6: Bubble chart for ECHI Duration and

Fig-2 Describes that the Line Chart for ACD Call pattern analysis based on the ACD preference Rate On Circle, Hub, Category and Overall ACD preference Rate. Fig-3 describes the Bubble chart for ECHI duration and queue time levels on ACD preference rate across circle or hub. Fig-4 describes that Bar graph for customer category of ACD Preference Rate on Customer Category and Purpose and Consult and Queue Time Levels on Customer category.

ACD Call Pattern Analysis that qualifies the overall patterns and trends available for ACD calls at circles, hubs, customer category level. It focuses on metrics like transfer rate, ACD preference rate, ACD queue time levels, consult time across circles, hubs, purpose customer category, etc. In this report, ACD preference rate which is defined on IVR calls with less than 10 seconds duration ending in transfer disposition. The report qualifies on ACD call pattern and duration metrics at the application level. ACD preferred calls monitored over a period of time. Proportions of the transferred calls are monitored by circle, hub or category. ACD preference rate, consult time and assist are tracked across circles. ACD preference rate, consult time and assist are tracked across hubs.

IV. CONCLUSION

This paper presents a big data application for telecom. We are acquiring the knowledge and insights of data and get the value. So this result improves the business metrics. The aim of the project is to advance analysis application for telecom, which will give the results in the form of reports. So, anyone can understand how the things going on without having any knowledge on that. This framework is pretty good for the analysis in fraction of minutes. There is a big productivity gain for the operator they no need to maintain the application everything we automated and if anything go wrong the alert mechanism is there to warn the employees through mail or short message service. End users can easily run the reports whatever they want within minimum time. The report results will display very clearly with different charts like bubble chart, line chart, bar charts.

References

[1] Dr. Poonam Kumar "Impact of Business Intelligence Systems in Indian Telecom Industry" Asia Pacific Institute of Management, New Delhi (INDIA), Business Intelligence Journal - July, 2012 Vol.5 No.2

[2] Monica LIA "Customer Data Analysis Model using Business Intelligence Tools in Telecommunication Companies " University of Economic Studies, Bucharest, Romania Database Systems Journal vol. VI, no. 2/2015.

[3] Pareek, D.: Business Intelligence for Telecommunications. Auer Bach Publications, 1. Taylor & Francis Group LLC (2007).

[4] A Guadalupe J. Torres, Ram B. Basnet, Andrew H. Sung, Srinivas Mukkamala, Bernardete M. Ribeiro "Hadoop Ecosystem: An Introduction" Volume 5 Issue 6, June 2016.

[5] Sqoop- Apache Software Foundation project home page http://sqoop.apache.org/

[6] Flume - Apache Software Foundation project home page https://flume.apache.org

[7] Oozie – Apache Software Foundation Project home page http://Oozie.apache.org

[8] Hadoop - Apache Software Foundation project home page http://hadoop.apache.org.

[9] Hive - Apache Software Foundation project home page http://hive.apache.org.