



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: VI Month of publication: June 2019

DOI: <http://doi.org/10.22214/ijraset.2019.6123>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Value Proposition for Big Data

Asmita Sharma¹, Drashya Shah², Falguni Mangwani³, Rachana Mudholkar⁴
^{1,2}Computer Department, SPPU

Abstract: Value proposition of the system is done to specify the need of the project. Once the project is approved, spark programs, also called spark jobs, are written to extract data from the databases and stored on Hadoop clusters. The data is then filtered according to the business needs using spark jobs and then this filtered data is placed on the file system again using spark in a format that can be used by SparkML. Spark speeds up the operation by parallelizing the operations. The data stored on Hadoop distributed file system is then used as a test data to be used by SparkML to train. Once the training is done, the machine learning algorithm is then used on production data to forecast the quantity of item required at a particular store at a given date.

Keywords: Elasticsearch, Logstash, Kibana, Hadoop, HDFS.

I. INTRODUCTION

For any e-commerce company which runs to serve people, having out-of-stock condition is both loss of profits and bad name for the company. To not be able to provide the customer with the items they need, causes decline in sales and makes the customer hesitant to come back. Any e-commerce giant, cannot afford to lose customers and the potential sales this way, so this project was needed to solve this issue. A new method which would help the customers as well as the company. The best way to do it is to predict what quantity of what item is needed at what store. Existing system predicts this number by manually looking at the data and uses a formula to obtain the number. This process is slow and does not work at high accuracy. To solve this, this project was started. This project aims at speeding up the existing system by storing all the data on distributed system that can be used with machine learning algorithm to predict the item quantity based on past sales. This project is made to handle large volumes of data stored on system databases. The extraction process is in line with the volume of data. The process is run in parallel so as to perform at a high speed. The proposed system stores the data on clustered storage so that the storage space can be increased if needed.

II. LITERATURE SURVEY

- 1) *Paper Name:* Value proposition discovery in big data enabled business model innovation
 - a) *Author:* De-ning Teng, Peng-yu Lu
 - b) *Paper Explanation:* Business innovation is unavoidable in the intensely competitive marketplace. Under resources constrained condition, the option of innovation perspectives is particularly significant. Business model is the foundation of business goal and business innovation which directly facilitates the innovation of products and business processes. Meanwhile, value proposition is the most significant element in business model innovation. Under such background, an innovative approach to achieve reasonable value proposition from knowledge discovery of enterprise information system is proposed in this paper. The paper also defines major business data sources for information system and presents appropriate big data processing approaches of knowledge discovery. To effectively reveal business value of knowledge discovery, popular data interpretation techniques are presented. Furthermore, in order to facilitate firm leader to prompt value proposition on a better degree, the paper proposes a value proposition generation model based on customers, competitors and profit which called CCP to associate business value discovered from information system with value proposition creation. The pattern of value proposition generation this paper proposed will strengthen competitiveness, and provide a better approach to achieve business goal of enterprises.
- 2) *Paper Name:* Forecasting consumer behavior with innovative value proposition for organizations using big data analytics
 - 1) *Author:* Ankur Balar, Nikita Malviya, Swadesh Pradesh, Ajinkya Gangurde
 - 2) *Paper Explanation:* The term 'Big Data' is used to represent collection of such a huge amount of data that it becomes impossible to manage and process data using conventional database management tools. Big Data is defined by three important parameters 'Volume' - Size of Data, 'Velocity' - Speed of increase of data and 'Variety' - Type of Data. Big data analytics is the process of analyzing this ever growing Big Data. The goal of every organization is to maximize its value for its stakeholders. The paper aims to demonstrate that Big data analytics can be used as a catalyst for generating and increasing value for organizations by improving various business parameters. Furthermore, by utilizing case studies the paper also aims to establish that big data analytics supports creation, enhancement and improvement of various business services to significantly improve customer experience as well as value creation for organizations.

III. PROPOSED SYSTEM

This project is made to handle large volumes of data stored on system databases. The extraction process is in line with the volume of data. The process is run in parallel so as to perform at a high speed. The proposed system stores the data on clustered storage so that the storage space can be increased if needed. The end users of the system are the higher management people who look at the value proposition and decide if the project is necessary. They also decide what data to choose and what attributes are needed. Further along the system, they decide on the biases of each item for the machine learning algorithm and then use the final output to determine the amount of item that is needed to be shipped.

A. Advantages of Proposed System

- 1) Fast data retrieval process.
- 2) Helps to realize ongoing trend.
- 3) Helps to plan business strategy.

IV. IMPLEMENTATION

The system contains three modules, namely extraction of data, value proposition and UI for biases. The data for value proposition is made available in ELK stack. Logstash ingests the data in Elasticsearch which creates inverted index of the documents. Kibana then uses the data in documents to represent the data visually in the form of charts. A UI is developed which takes in user biases which is used to adjust the weights in machine learning algorithm in case any item has special requirements. Extraction of data is done by using Spark to query different databases such as Cassandra and Oracle which returns the required tables. The data is then stored parallelly using Spark on a scalable storage such as Hadoop clusters. Spark is used again to filter out the required attributes from the various available attributes from the data gathered from different tables. This filtered data act as training set for the machine learning algorithm. Once trained, the algorithm is then used for final prediction on the recent item data.

V. CONCLUSIONS

This project tries to speed up the process and make the process more accurate by using the concepts of Big Data and cluster computing frameworks. The data from the databases are huge and, hence, are stored on a scalable storage such as HDFS. Spark is used for extracting the data from databases and storing on scalable storage.

The project then aims to predict the inventory quantity by using SparkML, a machine learning framework for Spark, to train using the training set and then predict the final number of items based on current data. This provides better accuracy as training set is much bigger than the spreadsheets and hence provide a better information about item sale pattern.

VI. ACKNOWLEDGMENT

We would like to take this opportunity to thank the Computer Department of Dr. D.Y. Patil Institute of Engineering, Management and Research for giving us the opportunity to perform this project under the guidance of Ms. Rachana Mudholkar.

We would also like to thank Prof. P.P. Shevatekar, Head of Department for her indispensable support and suggestions.

REFERENCES

- [1] Ankur Balar, Nikita Malviya, Swadesh Pradesh, Ajinkya Gangurde, "Forecasting consumer behavior with innovative value proposition for organizations using big data analytics", IEEE Computer Society, January 2014.
- [2] De-ning Teng, Peng-yu Lu, "Value proposition discovery in big data enabled business model innovation", IEEE Computer Society, August 2016.
- [3] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, "Spark: Cluster computing with working sets", 2nd USENIX Workshop on Hot Topics in Cloud Computing 2017.
- [4] K. Shvachko, H. Kuang, S. Radia, R. Chansler, "The hadoop distributed file system", 2017 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), pp. 1-10, May 2017..



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)