International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

ENHANCED ECOMMERCE FOR BUSINESS PRODUCTIVITY

Mrs. Maithili. K[1], Mrs. Priyaadarshini. B.S[2]

[1]Assistant Professor, Dept. of Computer Science and Engineering, Kingston Engineering college, Vellore, Tamil Nadu, India - 632059
[2]Student , Dept. of Computer Science and Engineering, Kingston Engineering college, Vellore, Tamil Nadu, India - 632059

Abstract - *In order to promote business and to develop a website similar to E-Commerce depth analysis is required which can be performed using Big Data Analytics (BDA) tools upon powerful framework called Hadoop by Scraping data from commonly known E-Commerce website Flipkart using Robotic Process Automation (RPA) Tools Robotic Enterprise Framework. The main objective of this paper is to scrape data from flipkart and to perform analysis upon those data to identify which product is widely sold and the customers are interested in purchasing to promote the business online and to develop new retail store online after performing depth analysis. This ensures that the revenue and production of the company is increased.*

Key Words: E -Commerce, Flip cart, Online transaction, Types, RAP tools, BDA tools

1. INTRODUCTION

E-commerce (electronic commerce) is the buying and selling of goods and services, or the transmitting of funds or data, over an electronic network, primarily the internet. These business transactions occur either as business-to-business (B2B), business-to-consumer (B2C), consumer-to-consumer or consumer-to-business. The terms e-commerce and e-business are often used interchangeably. The term e-tail is also sometimes used in reference to the transactional processes for online shopping.

After the number of individual users sharing electronic documents with each other grew in the 1980s, the rise of eBay and Amazon in the 1990s revolutionized the e-commerce industry. Consumers can now purchase endless amounts of items online, from typical brick and mortar stores with e-commerce capabilities and one another.

Existing concept deals with providing backend by using MySQL which contains lot of drawbacks i.e. data limitation is that processing time is high when the data is huge and once data is lost we cannot recover so thus we proposing concept by using Hadoop tool. It employs My SQL DB. Technique - MySQL is a relational database .in these we can store the data and processing but some limitations . We can process limitation of data. We get results with take more time and maintenance cost is very high.

Proposed concept deals with providing database by using Hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster and we get results with less time, high throughput and maintenance cost is very less and we are using joints, partitions and bucketing techniques in Hadoop. This uses Hadoop for the implementation. Hadoop is open source framework which as overseen by the apache software foundation and it is used for storing and processing huge datasets with a cluster of commodity hardware. We use Hadoop tool contains two things one is hdfs and map reduce. We also use Hadoop ecosystems like Sqoop, Hive and pig.

1.1 Advantages of proposed system:

- Huge Volume of Data related to E-Commerce is collected from resource center such as Flipkart by using scraper.
- RPA Tool is used for Scraping and Cleaning the Data where REFramework plays a vital role.
- Data is preprocessed carefully by predicting the outcome of these datasets.
- In order to analyze huge volume of data Hadoop framework is used to analyze and store data.
- This paper can promote the business by knowing the highly used products by the customers in E-Commerce Websites.
- No data loss problem
- Efficient data processing.

The Data can be collected from multiple resources such as Kaggle, UCI and .gov websites appropriately or can be directly scraped from E-Commerce Websites using Robotic Process Automation Tool. And can be preprocessed to perform analysis using Hive and Pig tool by retrieving data from HDFS. To predict those products which has high rating and reviews in order to promote the business. We are analyzing the e-commerce products removing the product description and the technological feasibility of this project is that we are using known technologies as well as open source codes and the operational feasibility is that we can be able to do as per the business requirement, and the economic feasibility is that we are preparing the project as per the partial fulfillment of the BCIS so the economic feasibility is fulfilled.

2. SYSTEM DESIGN OF MODULES

- Data Scraping Module Using UiPath
- Data Preprocessing Module
- Data Migration Module With Sqoop
- Data Analytic Module With Hive
- Data Analytic Module With Pig
- Implementation of MapReduce Algorithm
- Data Scraping Module
- In this module we scrape E-Commerce data using RPA tool i.e., using UiPath from Flipkart directly.

2.1 Module description

➢ **Data Scraping Module Using UiPath**

UiPath can extract literally anything you can see in a web browser. This includes statistics, finance and stock info, real-estate data, product catalogs, search-engine results, job listings, social networks feeds, customer opinions, and competitive pricing. Within a company, you can find even an even larger variety of data formats that UiPath can handle reports, dashboards, customers, employees, finance, and medical data that you need to transform and migrate. In our work we will extract the product details from ecommerce sites for scraping.

➢ **Data Preprocessing Module**

In this module we have to create Data set for E-Commerce dataset it contain set of table such that product name, product price, ratings and reviews for certain period, this data set will be first provide into MySQL database with help of this dataset we analysis this project.

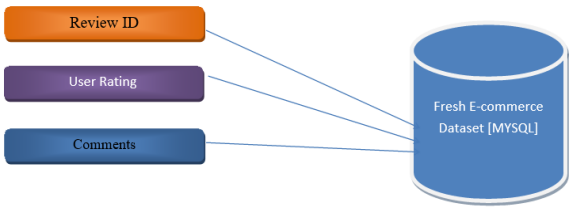


Fig – 1: Data processing module

➢ **Data Migration Module with Sqoop**

Now we are ready with dataset. So now our aim is to transfer the dataset into hadoop (HDFS), that will be happen in this module. Sqoop is a command-line interface application for transferring data between relational databases and Hadoop.In this module we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).

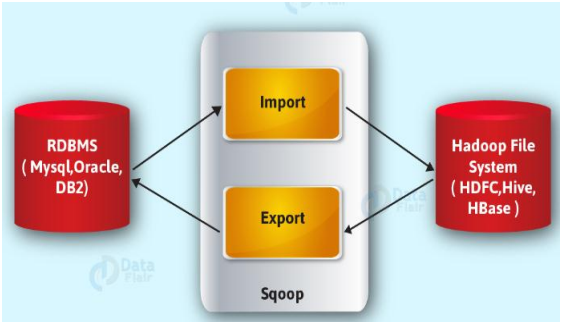


Fig – 2: Data Migration Module with Sqoop

➢ **Data Analytic Module with Hive**

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions.In this module we have to analysis the dataset using HIVE tool which will be stored in hadoop (HDFS).For analysis dataset HIVE using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.

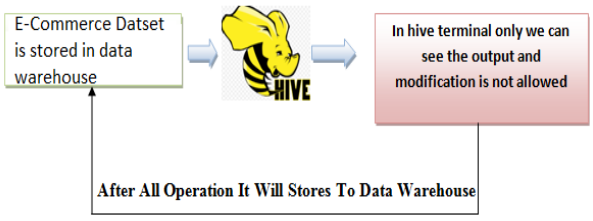
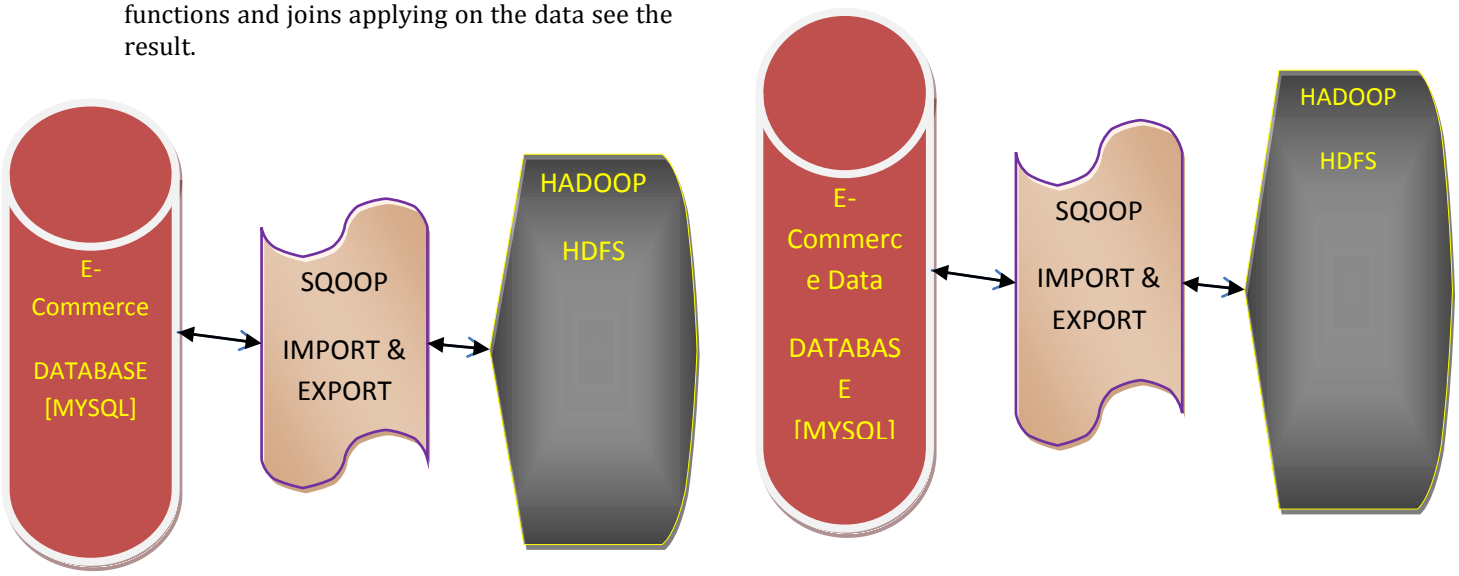


Fig – 3: Data Analytic Module with Hive

➢ **Data Analytic Module with Pig**

Apache Pig is a high level data flow platform for execution Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is also top of the map reduce process running background. In this module also used for analyzing the Data set through Pig using Latin Script data flow

language.in this also we are doing all operators, functions and joins applying on the data see the result.



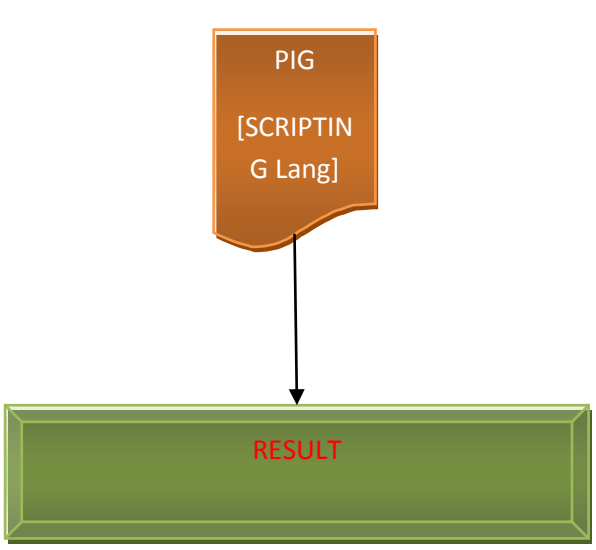


Fig – 4: Data Analytic Module with Pig

➢ **Implementation of MapReduce Algorithm**

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce.In this module also used for analyzing the data set using MAP REDUCE. Map Reduce Run by Java Program.

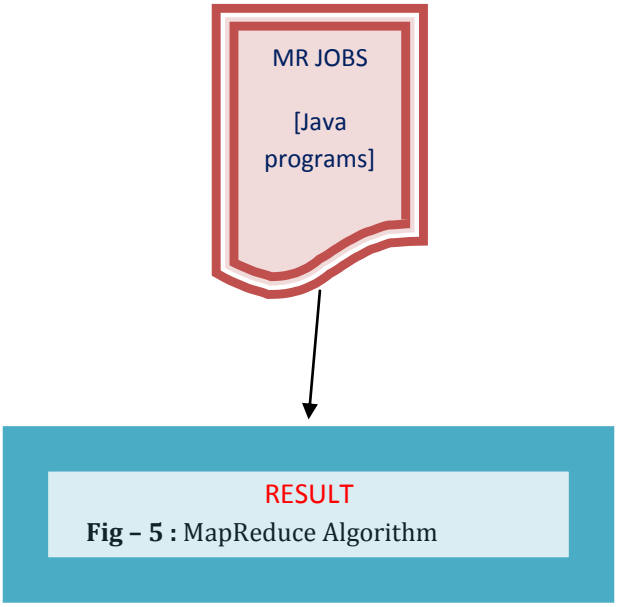


Fig – 5 : MapReduce Algorithm

3. PERFORMANCE METRICS

The analysis is completely based on the tools such as Hive and Pig which makes use of HQL(Hive Query Language) and Pig Latin Scripting Language and upon terminal the analysis logic is performed based on use cases. The result is same upon both the tools whereas the performance may vary and has both merits and demerits on comparison with those analysis tools. The predicted output can be which product is commonly purchased by many customers with their review and rating for promoting the business.The performance absolutely cannot be predicted because it may

vary from user to user since the use case or the scenario can vary and all the outcome depends on the HQL and Pig Latin Script Queries and the data.

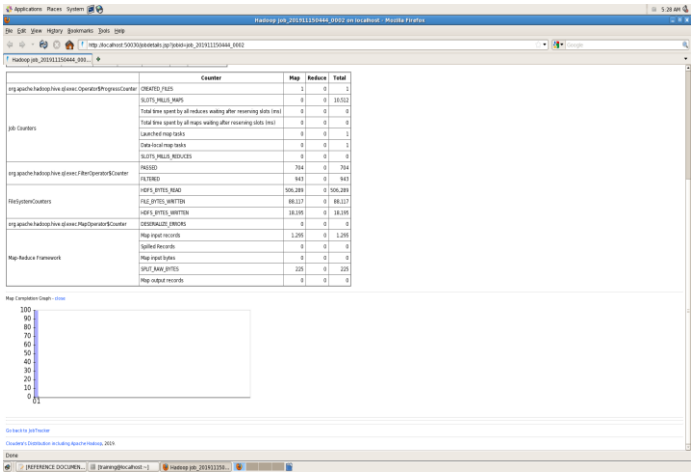


Fig – 6: MapReduce Algorithm

This Graph is one such example of Mapping Process and time taken in order to execute a query and its configuration triggered to execute that query.

4. CONCLUSIONS

In this paper we focused on categorization of customers based on the products purchased to increase the quantity and quality of specified products to earn more profit in online business. To yield better results we used huge volume of data hence Hadoop framework which is used for Big Data was focused and for scraping huge volume of data Robotic Process Automation was used which reduces manual effort. Spark is a framework - in the same way that Hadoop is - which provides a number of inter-connected platforms, systems and standards for Big Data projects. Like Hadoop, Spark is open-source and under the wing of the Apache Software Foundation.

Essentially, open-source means the code can be freely used by anyone. Beyond that, it can also be altered by anyone to produce custom versions aimed at particular problems, or industries. Volunteer developers, as well as those working at companies which produce custom versions, constantly refine and update the core software adding more features and efficiencies. In fact Spark was the most active project at Apache last year. It was also the most active of all of the open source Big Data applications, with over 500 contributors from more than 200 organizations. Unlike Hadoop, Spark does not come with its own file system - instead it can be integrated with many file systems including Hadoop's HDFS, MongoDB and Amazon's S3 system.

REFERENCES

[1] TITLE: Revisiting Semi-Supervised Learning for Online Deceptive Review [2017]

[2] TITLE: RESEARCH OF DECEPTIVE REVIEW DETECTION BASED ON TARGET PRODUCT IDENTIFICATION AND METAPATH FEATURE WEIGHT CALCULATION [2018]

[3] TITLE: A Network-Based Spam Detection Framework for Reviews in Online Social Media [2017]

[4] TITLE A Review of Natural Language Processing Research [2014]