Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Elaborating Role of Big Data Analytics in Transforming Retail Industry

***Ramandeep Kaur and Rajinder Singh**

*Research Scholar, PhD (Comp. Appl.), Guru Kashi University, Talwandi Sabo, Punjab, India.
Assistant Professor, UCCA, Guru Kashi University, Talwandi Sabo, Punjab, India.

ABSTRACT

Data is exploding at an ever-increasing rate. Its growth is accelerated by the Internet of Things (IoT), with every sector creating data and reporting back to the 'brand mothership' – as well as social media. The use of mobile devices is creating data and also changing the type and volume of data created. This new dynamic makes data harder to interpret, as it is. Both, structured and unstructured, and requires a different approach to analysis. Capturing relevant data and executing relevant analytics is an important part of maximizing the ability to interact and influence choices and decisions.Understanding consumer behavior enables organizations to react accordingly and create the best opportunity to influence follow-on actions and choices. The research paper is intended to discuss the importance of big data analytics. The paper also elaborates the working of different tools available within Hadoop framework.

KEYWORDS – Big data, big data analytics, business analytics, Hadoop, retail sector.

***Corresponding author:**

Ramandeep Kaur

Research Scholar,

PhD (Comp. Appl.),

Guru Kashi University, Talwandi Sabo,

Punjab, India.

I. INTRODUCTION

The ease of information availability, alongside the vast number of consumer businesses providing an online presence means we now have more socially informed, and tech-savvy, consumers than ever.This massive change means you need to ensure not just the traditional alignment of people, process and technology but also that business culture, including staff, is flexible enough to attract and retain consumers at the right cost profile. The old adage of 'people buy from people' has long been held to be true. Where this is appropriate, staff need the right skills, information and tools available to maximize every consumer contact point.In the modern world, with so much information available to businesses, finding the right information when it's needed has becomea significant challenge[1, 2].Having the right capabilities available to both staff and consumers has never been more important and having the right type of connection to the consumer has become the challenge.With the potential for significant volumes of consumers, automated solutions with enough information and intelligence to provide the right kind of consumer experience are required.Solutions need to intelligently harvest information from all sources to ensure information is collected and collated, leading to a positive outcome. The digital world has also created businesses that have very few overheads or assets, leading to real challenges for more traditional models.Having the right financial and organizational structures is important to realizing market potential. Without these, you will have the wrong cost or interaction model, leading to poor consumer communications[3, 4]. The availability of information to consumers means businesses now need to cope with change far more frequently.With so many examples of high-flying businesses closing shortly after reaching prominence, there is a need to consider change as the constant.This change should make you consider who your stakeholders are and what their motivation might be.A business that is a partner today can become a competitor overnight. For many, the ability to adapt will have a direct impact on whether they ultimately succeed or fail[5, 6].

II. VALUE CARRIED BY BIG DATA ANALYTICS

Business analytics has the capacity to generate value in a wide range of areas. Here are three most common[7, 8, 9].

- Consumer / Citizen - Insight Whether it is for a commercial organization or a public sector body, understanding the consumer or citizen is vital. Capturing information about their experiences with business, as well as the wider community, can produce insights into what consumers want or don't want. Building a profile of a consumer's journey in order to understand their preferences, likes, dislikes and habits, is an important factor in ensuring their expectations are met.Analytics takes this information to help shape the strategies around what

form of engagement a consumer requires, as well as what action to take next.Giving users the same experience viewed in different ways – 'same User Experience, different User Interface' – helps underpin the relationship between provider and consumer.

- Product / Services Insight - Using analytics to understand which current products or services are right for specific consumers is vital to understanding the health of your business.Demographic insight and application of techniques to focus market messages plays an important part in whether a product or service is hitting its market audience.Using analytics, you can start to model whether new products or services will be more or less applicable, and hence what part they play in the operation.Questionnaire analytics enable one to gain dynamic and immediate insight into which products or services are fulfilling market demand while having no overhead upon the consumer[10].
- Research and Development Insight - Combining consumer insight with product / services insight provides a rich platform of information that any R&D business can use as input into its planning process.Understanding what consumers wish to buy or what issues there are with an existing product helps designers quickly plan new offerings that are more market aligned.For businesses with long product development timescales, understanding the impact of change during the development lifecycle helps them plan and budget accordingly, while setting the correct expectations with their consumers[11, 12].

III. TECHNOLOGICAL ASPECT

Basic components of Hadoop architecture are mentioned as under and shown in Fig. 1.

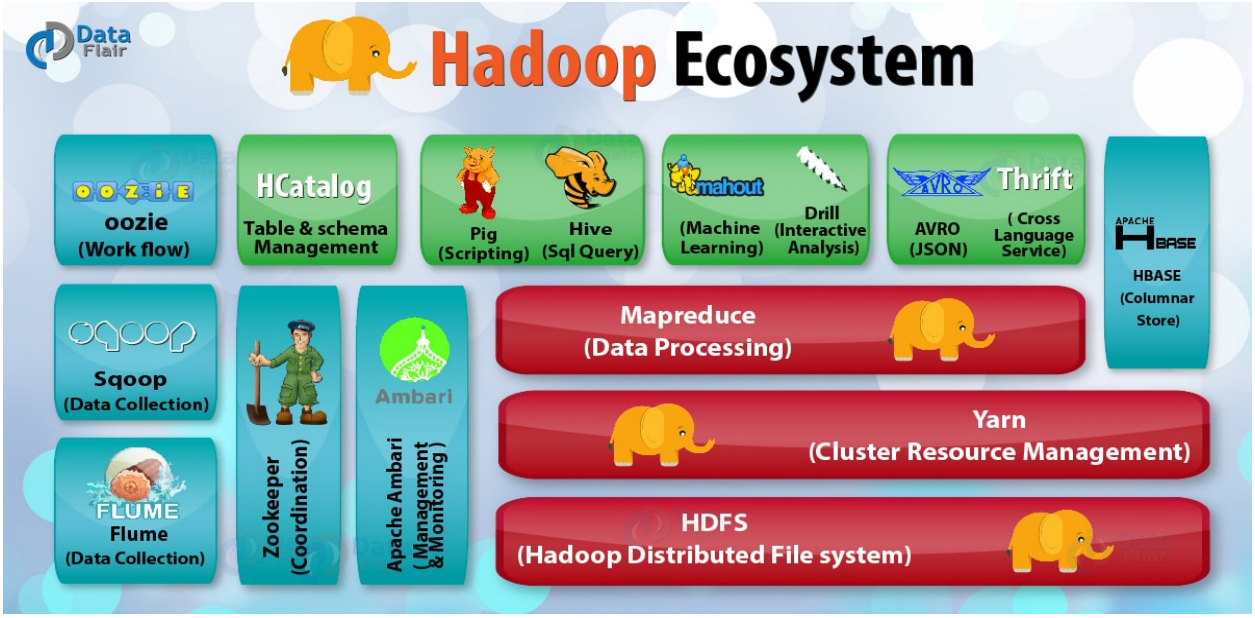

Fig. 1: Figure shows the components of Hadoop ecosystem

Hadoop Distributed File System (HDFS): HDFS is designed in order to provide quick access to data across numerous nodes in a cluster. HDFS basically is a distributed storage system. HDFS is capable of storing enormous amount of data that ranges 100+ terabytes in size and allows streaming this data at very high bandwidth to big data analytics applications.

Map Reduce: Map Reduce is a programming model enabling large data sets to be processed in a distributed manner on compute clusters of commodity hardware. MapReduce operates in three main phases; Map phase, Shuffle phase, and Reduce phase. Mapping involves splitting of large file into pieces to make another set of data. The shuffling phase arranges the sets with same key. The reduce phase considers the output from shuffling phase and assembles the results into consumable solution.Hadoop easily facilitates large-scale data analysis using multiple machines in the cluster[13, 14].

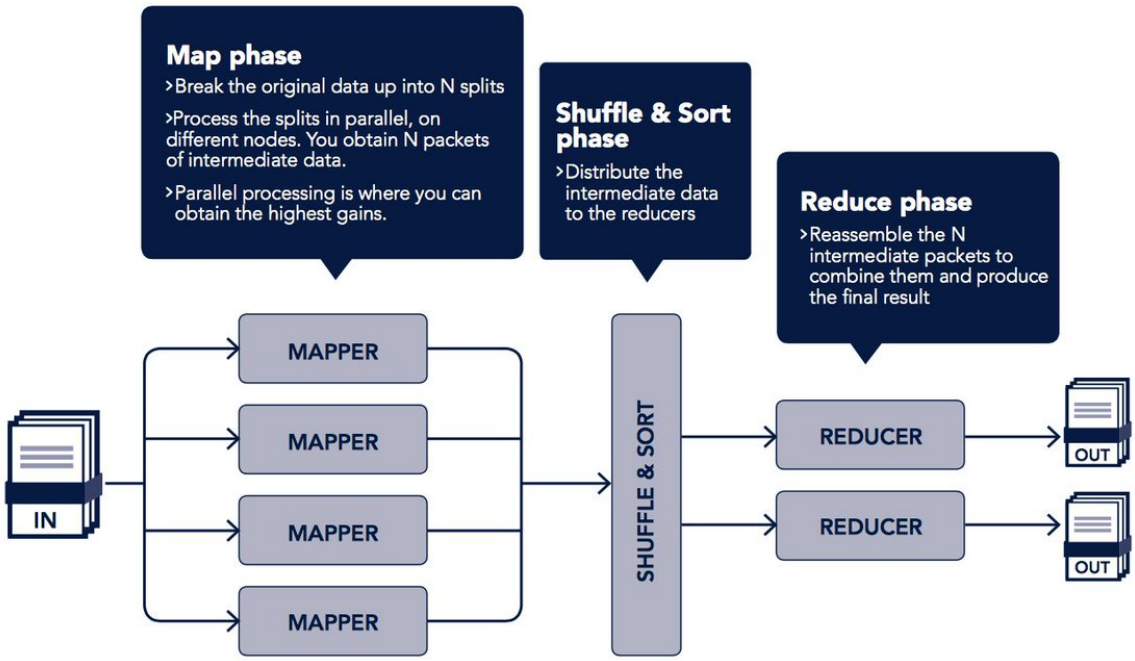

Fig. 2: Figure shows the working concept of MapReduce programming model

YARN: YARN refers to Yet Another Resource Negotiator. YARN brings on the table a clustering platform that helps manage and schedule tasks effectively and efficiently. It was set up to handleboth global and application-specific resource management components[15, 16]. YARN improves utilization over more static MapReduce rules that were rendered in early versions of Hadoop, through dynamic allocation of cluster resources.Every business has different data analytics requirements, which is why the Hadoop ecosystem offers various open-source frameworks to fit your special data analytics needs[17, 18].

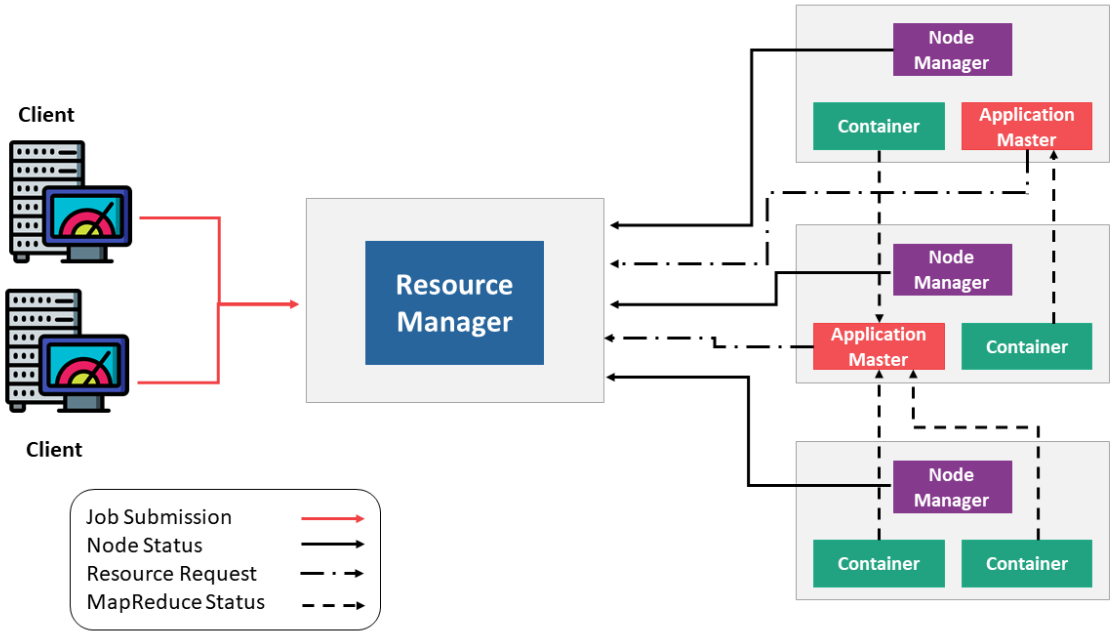

Fig. 3: Figure shows the components of YARN

Hive : Hive is an open-source data warehousing framework that structures and queries data using a SQL-like language called Hive QL. Hadoop allows developers to write complex MapReduce applications over structured data in a distributed system[19, 20]. If a developer can't express a logic using Hive QL, Hadoop allows choosing traditional map/reduce programmers to plug in their custom mappers and reducers. Hive is a very good relational-database framework and can accelerate queries using indexing feature.

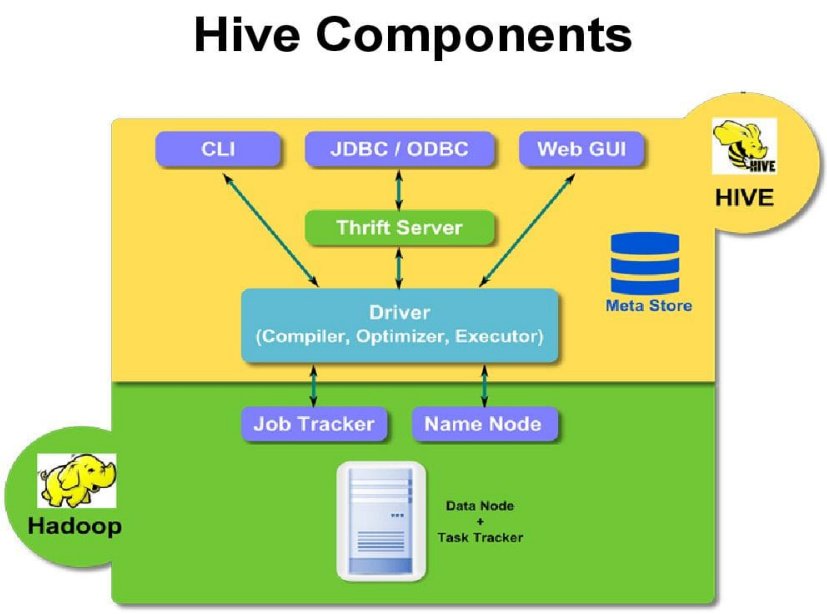

Fig. 4: Figure shows the components of Hive

Ambari : Ambari was designed to remove complexities of Hadoop management by providing a simple web interface that can provision, manage and monitor Apache Hadoop clusters. Ambari, which is an open-source platform, makes it simple to automate cluster operations via an intuitive Web UI as well as a robust REST API. The core benefits of Ambari are mentioned as under.

- Simplified Installation, Configuration and Management
- Centralized Security Setup
- Full Visibility into Cluster Health
- Highly Extensible and Customizable

HBase : HBase is an open-source, distributed, versioned, non-relational database model that provides random, real-time read/write access to your big data. HBase is a NoSQL Database for Hadoop. It's a great framework for businesses that have to deal with multi-structured or sparse data. HBase makes it possible to push the boundaries of Hadoop that runs processes in batch and doesn't allow for modification. With HBase, you can modify data in real-time without leaving the HDFS environment.HBase is a perfect fit for the type of data that fall into a big table. HBase first performs the task of storing and searching billions of rows and millions of columns. It then shares the table across multiple nodes, paving the way for MapReduce jobs to run locally.

Pig : Pig is an open-source technology that enables cost-effective storage and processing of large data sets, without requiring any specific formats. Pig is a high-level platform and uses Pig Latin language for expressing data analysis programs. Pig also features a compiler that creates sequences of MapReduce programs.The framework processes very large data sets across hundreds to thousands of computing nodes, which makes it amenable to substantial parallelization. In simple words, we can consider Pig as a high-level mechanism that is suitable for executing MapReduce jobs on Hadoop clusters using parallel programming.

ZooKeeper : Zoo Keeper is an open-source platform that offers a centralized infrastructure for maintaining configuration information, naming, providing distributed synchronization, and providing group services. The need of a centralized management arises when a Hadoop cluster spans 500 or more commodity servers, which is why Zookeeper has become so popular.ZooKeeper also avoids the single point of failure situation as it replicates data over a set of hosts, and the servers are in sync with each other. Although Java and C are currently used for ZooKeeper applications, Python, Perl, and REST interfaces could also be used someday for ZooKeeper applications.

IV. IMPLEMENTATION AND CONTRIBUTION

This section elaborates the implementation of Hadoop using Hortonworks Sandbox 2.2.0.

Fig. 5 shows the queries written in query editor of Hortonworks Sandbox 2.2.0.

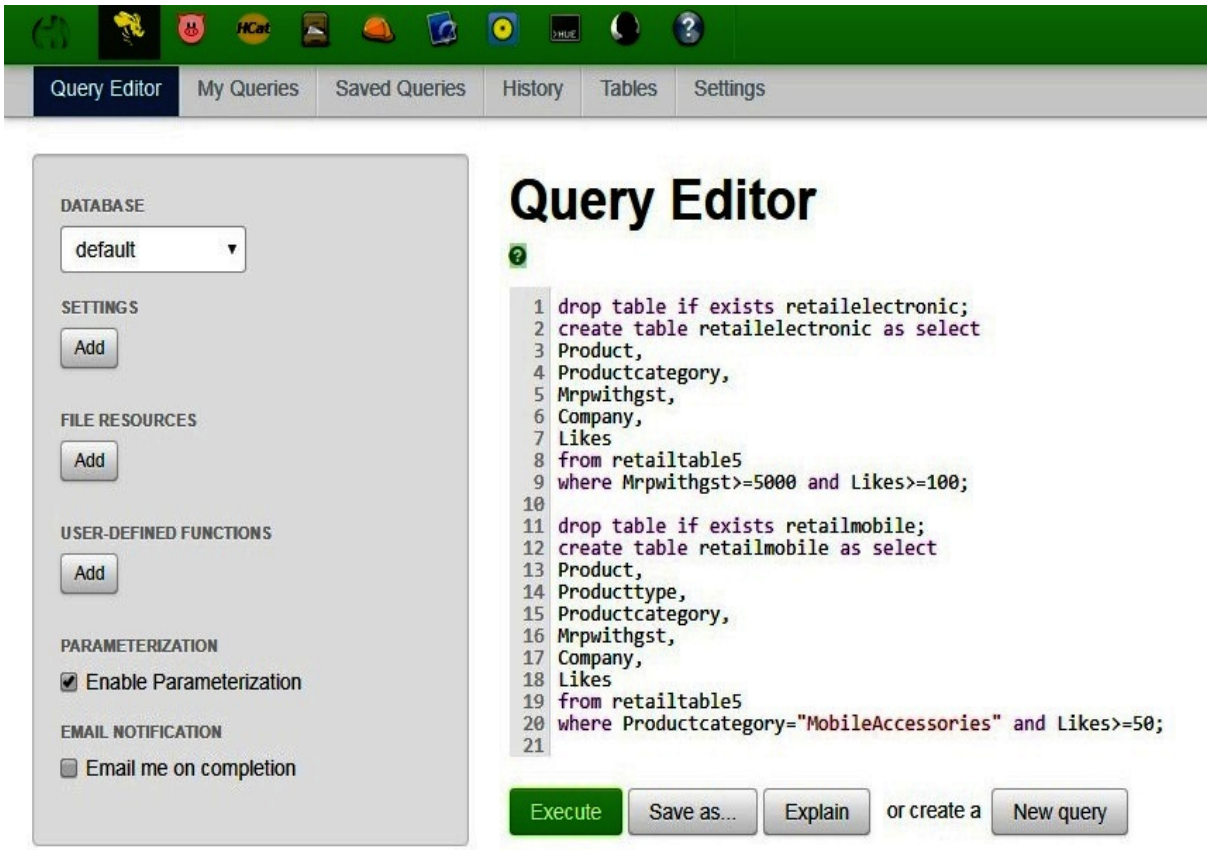

Fig. 5: Figure shows the queries written in query editor

Fig. 6 below shows the two created tables "retailelectronic" and "retailmobile" in the Tables list.

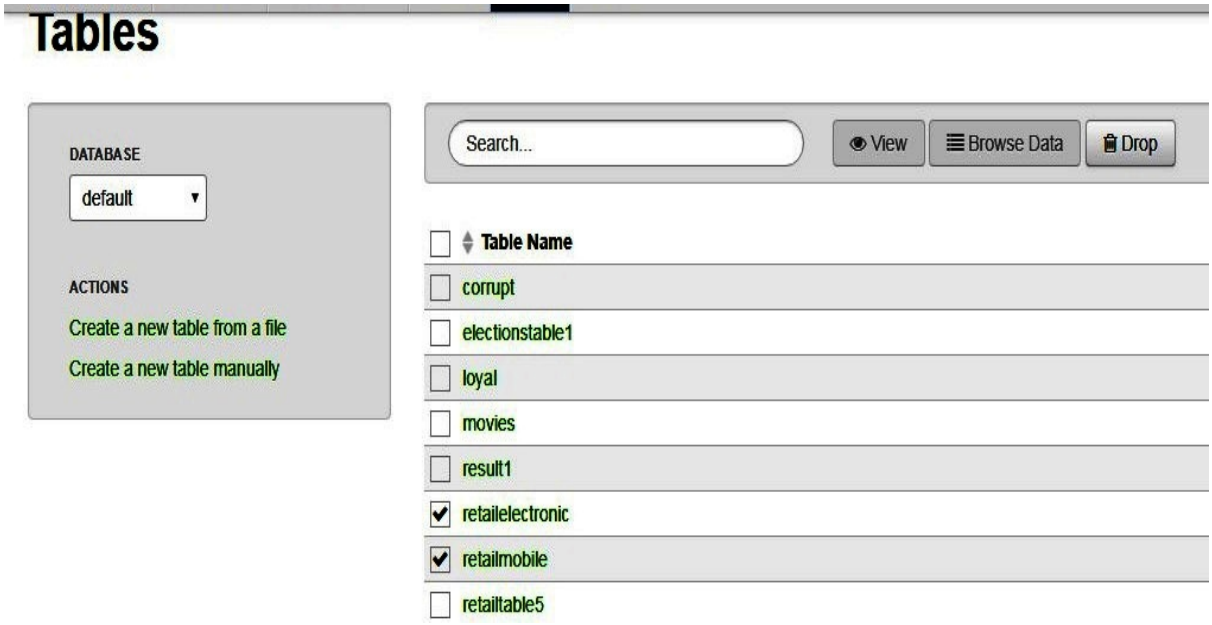

Fig. 6: Figure shows the two created tables "retail electronic" and "retail mobile" in the Tables list

Fig. 7 shows the contents of the table "retail electronic" below and shows the visualization results of the same table in Fig. 8.

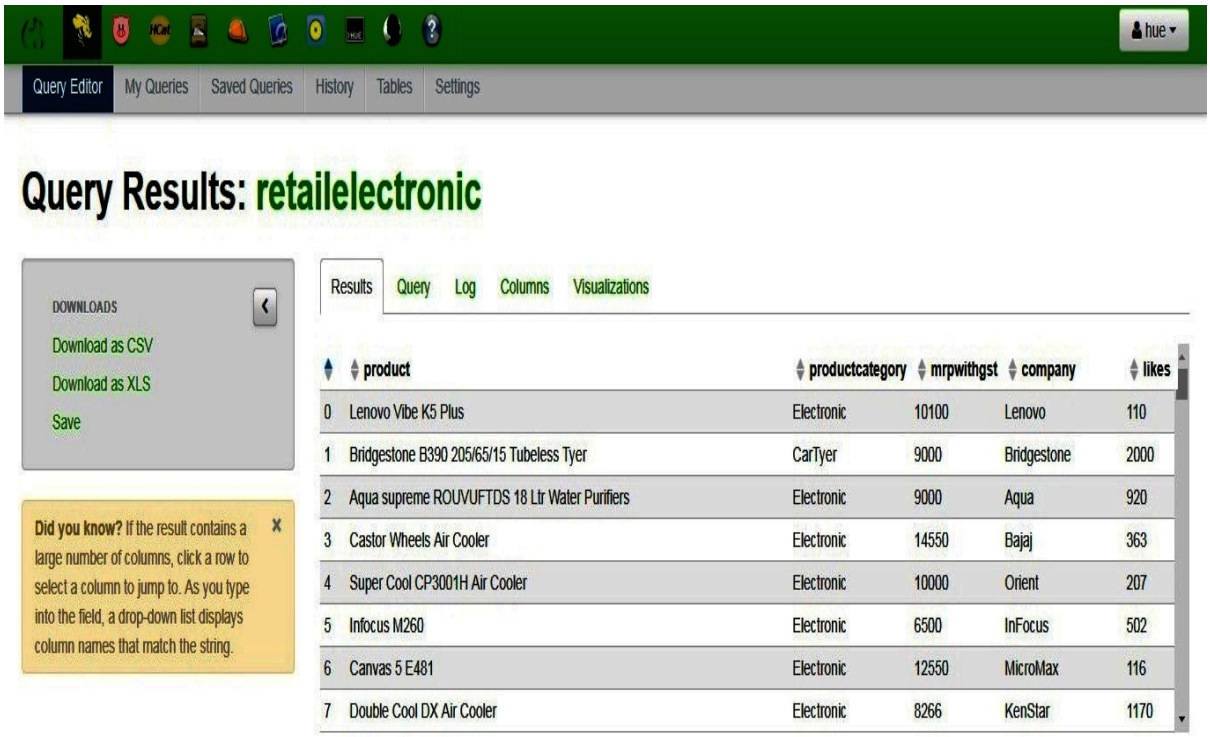

Fig. 7:Figure shows the contents of the table "retailelectronic"

Fig. 8 shows the visualization results of the table "retailelectronic". The bars in the green color shows "Mrpwithgst" and the bars in the blue shows "Likes".

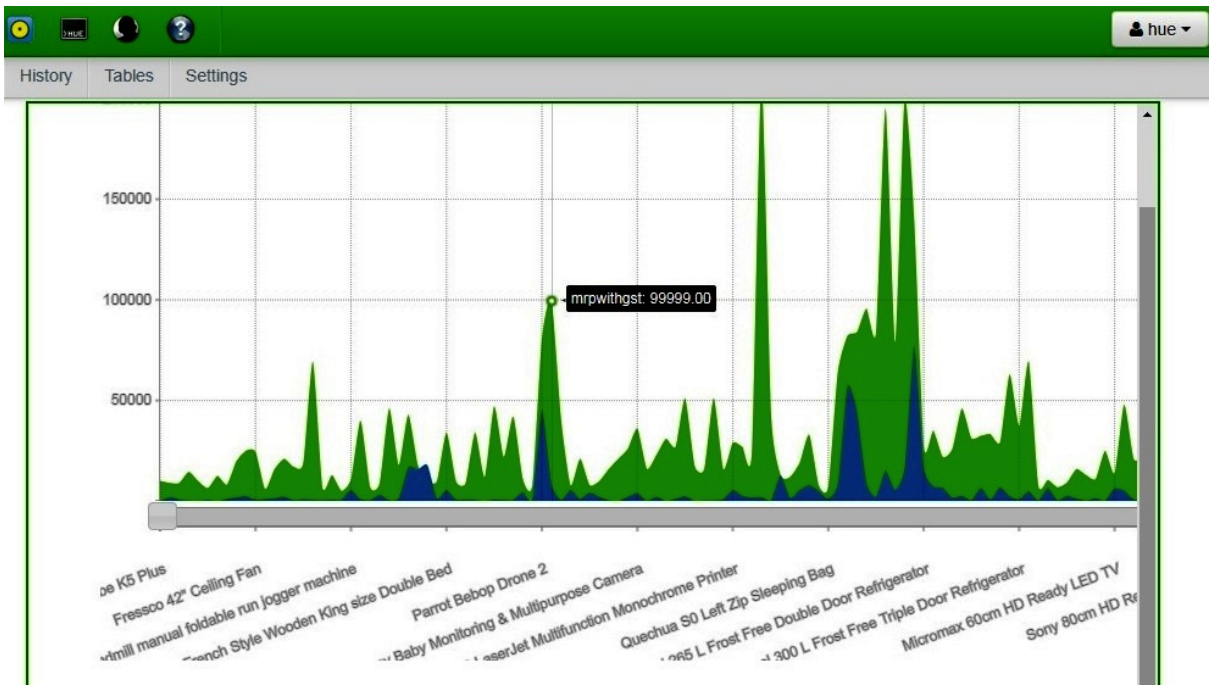

Fig. 8: Figure shows the visualization results of the table "retailelectronic"

V. CONCLUSION

The research paper elaborated the value carried by big data analytics. The technological aspect has been discussed via appropriate implementation using Hadoop based Hortonworks Sandbox 2.2.0. On the basis of the above discussion conducted in the research paper, it can be concluded that big data analytics helps one to explore the potential of consumer analytics and allows incorporation of methods in which retail industry have gained benefit from analytics like understanding examples of how consumer analytics has created value for other organizations; identifying use cases for areas that would lead to value creation; and identifying potential styles or types of analytics based upon use cases.

REFERENCES

1. Mohey El-Din Mohamed Hussein, D.: A survey on sentiment analysis challenges. J. of King Saud Univ. – Engg. Sci. 2016; 1 – 9 .
2. Jagdev, G. Sentiment Analysis and its Impact in Modeling Election Scenario. IJRSCSE, 2018; 5(2): 22-27.
3. Bermingham, A., Smeaton, A. F. Classifying sentiment in microblogs: is brevity an advantage? In Proc. of the 19th ACM int. conf. on Inform. and Know. Manag., 2010; 1833 – 1836.
4. Sharma, Y., Mangat, V., Kaur, M. A Practical Approach to Sentiment Analysis of Hindi Tweets. In Proc. of the 2015 1st Int. Conf. on Next Gen. Comp. Techno. (NGCT), 2015; 677 – 680.
5. Jagdev Gagandeep et al.A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm. IJARSE, 2017; 6(5).
6. Elghazaly, T., Mahmoud, A., Hefny, H. A. Political Sentiment Analysis Using Twitter Data. In Proc. of the Int. Conf. on Inter. of things and Cloud Comp.,2016.
7. Sharma, P., Moh, T-S. Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter. 2016 IEEE Int. Conf. on Big Data.2016; 1966 – 1971.
8. Jagdev Gagandeep et al. Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data. IJATES. 2017; 5(5).
9. Kharde, V. A., Sonawane, S.S.: Sentiment Analysis of Twitter Data: A Survey of Techniques. Int. J. of Comp. App. 2016; 139(11): 5 – 15.
10. Medhat, W., Hassan, A. Korashy, H.: Sentiment analysis algorithms and applications: A survey. A. S. Engg. J. 2014; 5: 1093–1113.

11. Kaur Amandeep,Jagdev Gagandeep. Exploring Application of Big Data in Elections – From Data to Action. IJRSCSE. 2017; 4(4): 64-71.
12. Niyogi, M., Pal, A. K. Discovering conversational topics and emotions associated with Demonetization tweets in India. Ar Xiv. 1711.04115v1 [cs.CL]. 2017: 1 – 6.
13. Vijayarani, S., Ilamathi, J., Nithya, M.: Preprocessing Techniques for Text Mining - An Overview. Int. J. of Com. Sci. & Comm. 2015; 5(1): 7-16.
14. Jagdev Gagandeep et al.Analyzing Maneuver of Hadoop Framework and Map R Algorithm Proficient in supervising Big Data. IJATES. 2017; 5(5).
15. Mohammad, S. M., Kiritchenko, S. Using Hashtags to Capture Fine Emotion Categories from Tweets. Sp. issue on Semantic Analy. in Soc. Med., Comp. Intell. 2013; 1 –22.
16. Jagdev Gagandeep et al.A Study of Clustering and Classification Techniques involved in Data Mining. IJATES. 2017; 5(5).
17. Sharma, G.: Extract the context: Sentiment Analysis and Opinion Mining. Parallel Dots. 2017.
18. Jagdev Gagandeep et al.Analyzing and Filtering Big Data concerned with elections via Hadoop Framework. IJARSE. 2017; 6(4).
19. Vu Dung Nguyen, Blesson Vaghese, “Royal Birth of 2013:Analysing and Visualising Public Sentiment in the UK using Twitter,” Research Gate, 2013.
20. Jagdev Gagandeep et al.Comparing Conventional Data Mining Algorithms with Hadoop based Map-Reduce Algorithm considering elections perspective. IJIRSE. 2017; 3(3).