Big Data in Retail Sector – An Evolution that Turned to a Revolution

Ramandeep Kaur[1], Dr. Gagandeep Jagdev[2]

[1]*Research Scholar (Ph.D), Dept. of Comp. Appl., Guru Kashi University, Talwandi Sabo (PB)*

[2]*Dept. of Comp. Science, Punjabi University Guru Kashi College, Damdama Sahib (PB), India*

*Corresponding Author: Dr. Gagandeep Jagdev, Dept. of Comp. Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). India

Abstract: *The enormous amount of data generated from different sources in different fields and sectors and in different formats, which poses a challenge to existing data mining techniques, is referred to as big data. Today only big giants in different sectors like retail, medical, sports, banking etc. can think of handling big data because of huge cost and expenses attached to it. Big data relevant to customer purchase behavior is voluminous, volatile, and the veracity attached with big data has forced big retail industries like Walmart ($473 billion), Tesco ($130 billion), and Target ($73 billion), scrambling to bring up new data mining techniques and shifting to cheaper open source software to precisely store and examine data in real time. To be precise in knowing the interest ofcustomers, companies need to regularly mine the data related to each individual customerand gain insight of his or her shopping pattern like which goods he or she buys, customersbudget, customers likes and dislikes etc.The research paper concentrates on highlighting the changes brought in the retail sector by big data. The paper discusses the working of MapReduce technology with an appropriate example.*

Keywords: *Big Data; Map Reduce; retail industry.*

1. INTRODUCTION

It was a prominent belief among people that good marketing is an art, but of late, a scientific element has been introduced to marketing campaigns by the big data [1]. Marketers are relying on data to test, devise, and inform strategies than ever before. No doubt, there is no match for the creative minds in best marketing campaigns, but data and analytics can definitely provide the marketers with tools to upgrade performance. With the advent of universal digital technology, retailers make well-informed decisions utilizing online data in order to discover, associate, and buy products anytime from anywhere. Information acts as a game changer for retailers and is collected from different sources as shown in Fig. 1. Retail data analytics enables companies to stay well-informed of the latest shopping trends by implementing customer diagnostics to expose, understand, and act on meaningful data insights. The retailers are accepting the data-first strategy to have a deep insight of customers buying patterns, mapping them to products, and forming strategies to increase the sale in order to register increased profits. Retailers utilize the structured and unstructured information available about their consumer's behavior to utmost potential. Future of retail is totally dependent on big data analytics as it is capable of separating the wheat from the chaff. Retail is adopting data- centric technology for boosting sales. Retailers have achieved 73% increase in sales. Retailers are mining customers analytics to increase profits, increase growth and to be in competition, whether it is in-store or online. Data analytics is not a new concept in the retail sector. The retail market has always been interested in finding and analyzing new buying trends in customer's behavior. This allows retailers to associate products like shampoo and conditioner, toothpaste and toothbrushes, as the data patterns show that the people who buy shampoo will invariably also buy a conditioner. Placing related products together influences the buying behavior of the customer.

Fig1. *The figure represents the iconic presentation of different sources responsible for generation of big data*

All the data being generated in the retail sector is totally diverse consisting of raw, structured, semi-structured and even unstructured data which is difficult to be handled by the existing traditional analytic systems. Mismatched data formats and data structures represent significant challenges that can lead to analytic collapse. Fig. 2 shows the percentage of the data generated in different formats [2, 3].

Machine learning takes big data it gathers from computer users and is utilized to predict important insight about customers spending patterns and how these patterns undergo changes every day of the week, season, and time of the year. For example, during summers, the retail stores sell bathing suits, pool supplies, and summer dresses. Around Christmas, crafts and hobbies stores increase their revenue by selling holiday decorations and displays. Big data is indirectly intelligent to directly affect sales by assembling data about consumer's exact spending habits, through various human communications: retail point-of-sales systems in malls, and stores, restaurants eat-ins, Facebook and other social media sites, Pinterest, and the internet.
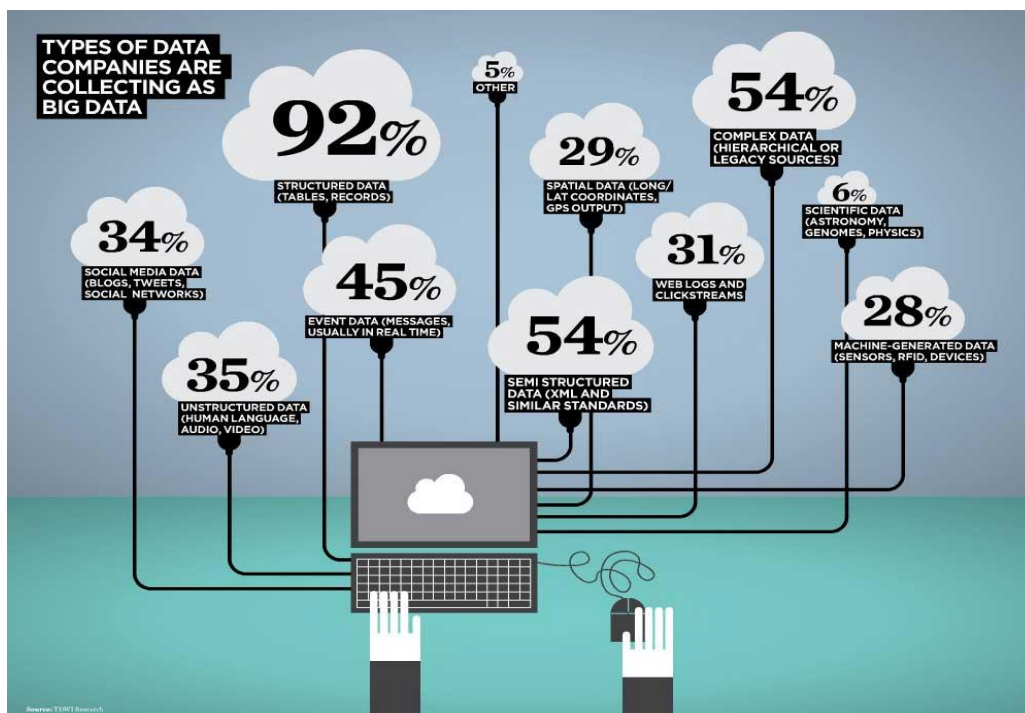

Fig2. *The figure depicts the percentage of different types of data collected by companies as big data*

2. EXAMPLES TO UNDERSTAND ROLE OF BIG DATA IN RETAIL

Examples of Positive effects of using Big Data in retail sector via major giants are as under

- Amazon has a brilliant and concentrated program to detect and prevent credit card frauds. This statement can be supported by the fact that there has been 50% reduction in frauds within first 6 months.

- Amazon has developed tools meant for fraud detection that makes use of scoring approach in predictive analysis.

- Amazon has prepared personalization strategy by using product based collaborative retail analytics.

- Amazon provides customers with intensive data-driven recommendations based on their previous purchase history, wish lists and browser cookies.

- Metro Group retailers make use of retail analytics to detect movement of goods inside the stores to provide in time information to concerned store personnel and customers for their convenience.

- Amazon responds to the competitive market rapidly because of its analytical platform which enables dynamic pricing by changing the prices of its product every 2 minutes (if required). In comparison, other retailers make this change approximately every 3 months.

- Staples, a US-based supply chain store makes use of Hadoop and Big Data technologies to predict sales by processing approximately 10 million data transactions every week and forecasts the sales on daily and weekly basis across 1100 retail outlets in the US. These predictions are used to target market promotions based on geographical area. Via using retail analytics, Staples was able to decrease its promotion costs by 25%.

3. BIG DATA REVOLUTION IN RETAIL MARKET

Big data is responsible for bringing a huge revolution in the retail sector. Some important factors in this regard are discussed as under [4, 5].

- **Generation of Recommendations**

On the basis of customer purchase history, retailers can forecast what the customer is more likely to purchase in coming future. Machine learning models are accomplished in handling historical data to assist the retailer in generating precise recommendations.

- **Forming strategic decisions**

There is a need to unite data by companies which would assist in reaching informed business decisions making use of a reliable and single source of information regarding customers and products. Retail dashboards would provide with a high-level overview of performance metrics comprising catalog movements and pricing promotion.

- **Predicting trends**

By making use of demographic data and economic indicators, retailers can better know about market demands and predict trends accurately.

- **Bringing Market Basket Analysis to practice**

The Market Basket analysis has helped in figuring out what products customers are most likely to purchase together. By using Hadoop, the retailers can perform better analysis even further.

- **Enhancing pricing**

The big giants like Walmart are investing big in real-time merchandising systems. Currently, Walmart is concentrating on the development of private cloud that would be able to track millions of transactions every day. The important factors like competitors, inventory levels, and demand can be tracked and changes in the market can be answered automatically [6].

- **Snooping to social media**

It is very important for retail industries to listen to what customers say on social media. Hadoop is capable of handling such enormous amount of unstructured data. Natural language processing is used to mine information from social media websites. This gives the retailer an advantage over competitors.

- **Predicting trends**

Today sentiment analysis is quite popular among marketers. This can be accomplished by developing sophisticated machine learning algorithms. The gathered data can be effectively utilized to come out with top-selling products in any particular category very accurately.

- **Expansion in cross-channel data volumes**

The increase in the number of mobile, social media and tablets has accelerated the boost in the available customer data. A usual retailer today knows the demographic information about a customer, his/her purchase pattern, likes and dislikes, and his/her call center interaction, social interaction and much more. The retailer shave extraordinary information about customers today and they can use this information to full potential to build a healthy relationship with the customer and ensure long-term benefits [9, 10].

- **Boosting investment in technology**

Storage today has become so cheap that companies have no more issues related to the cost of storing enormous data. Retail giants have invested in building a centralized database and concentrating on data analytics in order to gain insight into their customer's preferences and dislikes which was not possible few years back. Companies are willing to spend billions in this regard and ensure their economic growth [8, 10].

- **Solving the omnichannel puzzle with data**

Retailers having data-centric attitude are gathering the unbelievable amount of data related to customer behavior to have a deep insight of how the customer is researching and buying products. Retailers are trying to tackle the problems like foot traffic and in-store checkout wait times. New features like in-store cabins, free Wi-Fi and equipping the sales staff with mobile devices helps to better serve the tech-savvy customers on spot. The retailer needs to concentrate on every possible channel. Marketers should not ignore one channel at the cost of other [3].

- **Improving personalization**

Retailers need to build a personal database of their customers. Customers are willing to share their personal information if it earns some sort of benefit to them. The main aim of the retailers is to reach to the customer son right time, at right place and via right channel. This is possible only if company conducts periodic personalization of their customers.

- **Filtering out the most valuable customers**

Mining Big Data is a gigantic undertaking and the payoff lies in searching the gainful customers. For achieving success, it is essential to give high priority to such customers because it costs more to build relations with new customers than to keep the best customers. For brands competing in an industry with slim margins, mining the right data and conducting smart analysis will lead to better engagement, more faithful customers and a competitive benefit [4, 7].

4. WORKING OF MAP REDUCE ALGORITHM

Algorithm for Map-Reduce

The working of Map Reduce algorithm is mentioned as under [8, 11].

- The input data can be divided into n number of chunks depending upon the amount of data and processing capacity of the individual unit.

- Next, it is passed to the mapper functions. All the chunks are processed simultaneously at the same time, which embraces the parallel processing of data.

- After that, shuffling happens which leads to aggregation of similar patterns.

- Finally, reducers combine them all to get a consolidated output as per the logic.

This algorithm embraces scalability as depending on the size of the input data, one can keep increasingthe number of the parallel processing units [7, 8].

Example 1

Suppose we have a file with size about 200 MB, suppose content as follows

-----------file.txt-----------

_____**File(200 MB)**_____

hi how are you

how is your job (64 MB) 1-Split

_____

how is your family

how is your brother (64 MB) 2-Split

_____

how is your sister

what is the time now (64 MB) 3-Split

_____

what is the strength of hadoop (8 MB) 4-Split

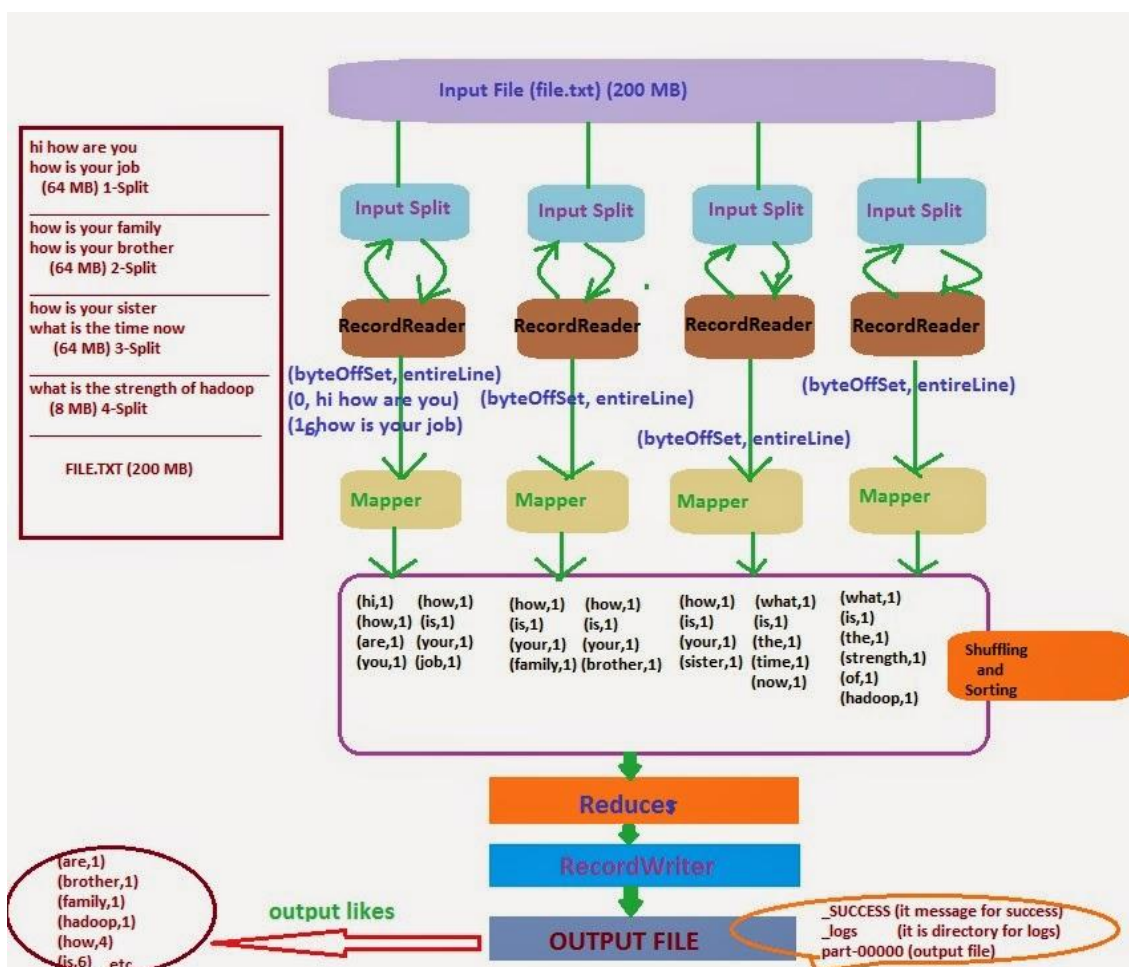The working of MapReduce technique in case of above Example 1 is demonstrated in Fig. 3.

Fig3. *The figure shows the working of MapReduce technique*

Example 2

Let's consider another example to understand the concept of working of MapReduce technique. Suppose we have four different categories (A, B, C, and D) having names of mobile manufacturing companies and price of most expensive mobile model manufactured under different categories by them.Each category further consists of four different records. Our aim here is to find out the company manufacturing the most expensive mobile from different categories in the database under consideration.

Category A:

Sony, 26000

Karbonn, 15000

Samsung, 25000

HTC, 10000

Category B:

Gionee, 50000

Karbonn, 25000

Samsung, 15000

HTC, 10000

Category C:

Sony, 26000

Oppo, 45000

Samsung, 15000

HTC, 1000

Category D:

Sony, 26000

Karbonn, 25000

Redmi, 45000

HTC, 10000

The Map phase – The first job to be performed is to form key-value pairs which is as simple as just enclosing the each record in the form <k, v> : <company name, price of mobile>.
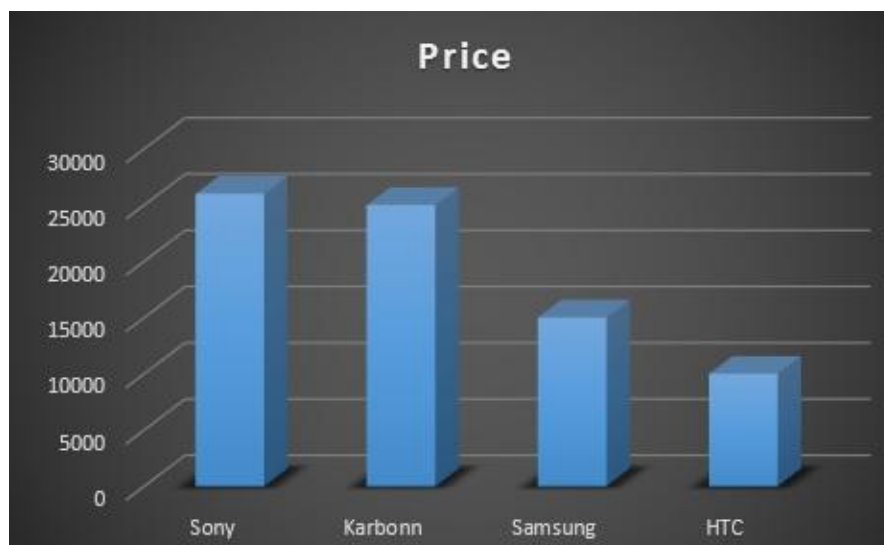
Category A:

<Sony, 26000>

<Karbonn, 25000>

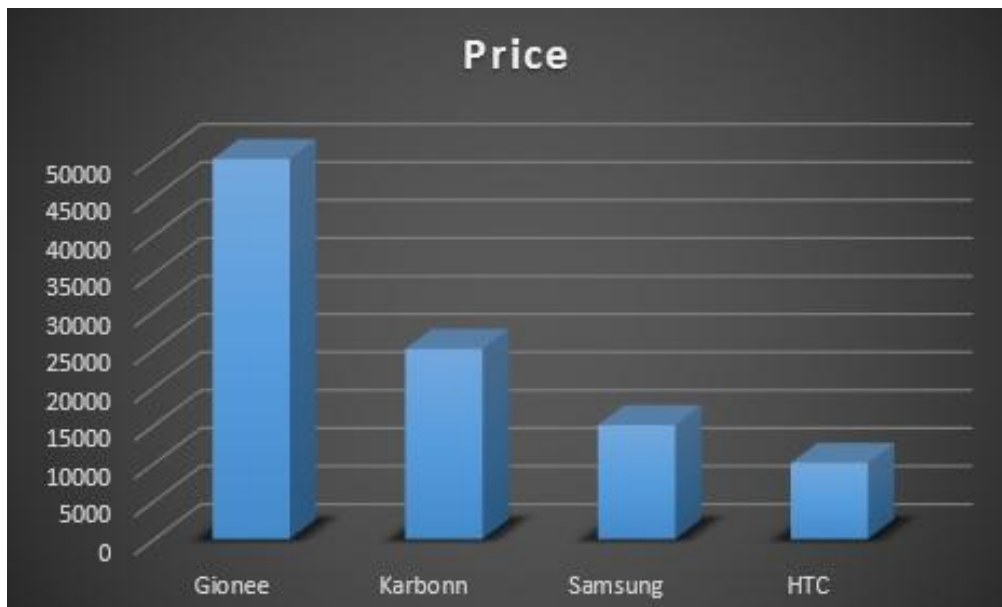<Samsung, 15000>

<HTC, 10000>

Fig4. *The figure depicts the graph plotted from data of Category A*

Category B:

<Gionee, 50000>

<Karbonn, 25000>

<Samsung, 15000>

<HTC, 10000>

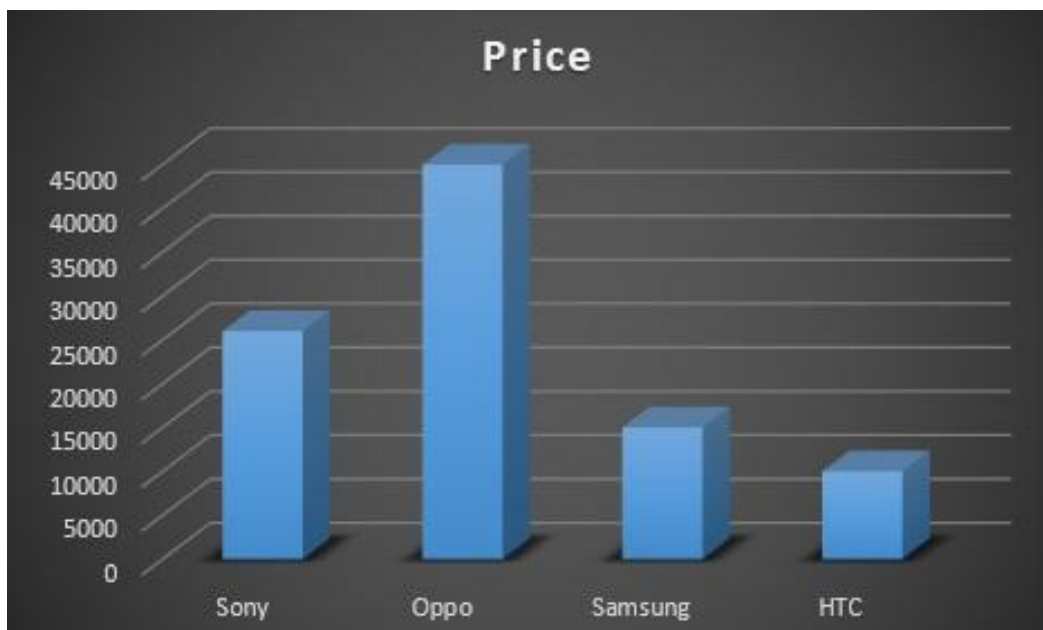**Fig5.** *The figure depicts the graph plotted from data of Category B*

Category C:

<Sony, 26000>

<Oppo, 45000>

<Samsung, 15000>

<HTC, 10000>

Fig6. *The figure depicts the graph plotted from data of Category C*

Category D:

<Sony, 26000>

<Karbonn, 25000>

<Redmi, 45000>

<HTC, 10000>

Fig7. *The figure depicts the graph plotted from data of Category D*

The combiner phase (searching technique) - Here one can write a code to find out the most expensive mobile from each file as follows:

<k: companyname, v: price of mobile>

Max= the cost of a first company mobile, treated as most expensive

if(v(second company).price> Max)

{

  Max = v(price);

}

else

{

  Continue checking;

}

The expected result is as follows –

A - <Samsung, 26000>

B- <Gionee, 50000>

C- <Oppo, 45000>

D- <Redmi, 45000>

Fig8. *The figure depicts the graph showing final result*

Reducer phase − Form each file, you will find the most expensive mobile. The same algorithm is used in between the four <k, v> pairs. The final output should be as follows –

<Gionee, 50000>.

5. CONCLUSION

The shopping trends have undergone a total change in the past few years as the control has now shifted in the hands of consumers. Before buying any product consumer can compare the price of that particular product at different online shopping websites and choose for the economical one. Further consumers today have a privilege to read and give their views related to the product and provider. It provides a lot of help to the buyer to learn from the experiences and reviews of the people who have already tried that particular product. It is now the time for retailers to think and employ new strategies to attract customers with themselves because of hard competition prevailing in the retail market. For this retailers make use of big data for harnessing massive volumes of new data available. The research paper illustrated the role played by big data in the retail sector and also depicted the working of Map Reduce technology via two examples. Retailers need to detect fraudulent activities in order to safeguard their margins and reputations. Big Data can provide assistance to the retailers to recognize anomalies and patterns by continuously monitoring the tactics and practices that looks unusual. This can help indicate incidents of fraud such as shrink and store associate theft and look for exceptions.

REFERENCES

[1] Dean, Jeffery, and Ghemawat Sanjay. 2004. "MapReduce: Simplified Data Processing on Large Clusters." Google.

[2] Katal, A., Wazid, M., &Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. IEEE, 404-409.

[3] Gagandeep Jagdev et al., "Comparing Conventional Data Mining Algorithms with Hadoop based Map-Reduce Algorithm considering elections perspective", *International Journal of Innovative Research in Science and Engineering (IJIRSE)*, ISSN: 2454-9665 (O), ISSN: 2455-0663(P), Volume – 3, Issue – 3, March 2017.

[4] Gagandeep Jagdevet al., "Association of Big Data with Map-Reduce Technology Augments for Economic Growth in Retail", *International Journal of Engineering Technology Science and Research (IJETSR),* ISSN: 2394 - 3386, Volume 4, Issue 2, February 2017.

[5] N. Marz and J. Warren. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications, 2013.

[6] R. Smolan and J. Erwitt. The Human Face of Big Data.Sterling Publishing Company Incorporated, 2012.

[7] Gagandeep Jagdevet al., "Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data", *International Journal of Advanced Technology in Engineering and Science (IJATES)*, ISSN – 2348-7550,Volume – 05, Issue – 05, May 2017.

[8] Gagandeep Jagdev et al., "Implementation and Applications of Big Data in Health Care Industry", *International Journal of Scientific and Technical Advancements (IJSTA)*, ISSN: 2454-1532, Volume 1, Issue 3, pp: 29 – 34.

[9] Undefined by Data: A Survey of Big Data Definitions, by Jonathan Stuart Ward and Adam Barker, School of Computer Science, University of St. Andrews, UK, 2013. Available at: http://arxiv.org/ pdf/ 1309.5821v1.pdf.

[10] Gagandeep Jagdev et al., "Big Data proposes an innovative concept for contesting elections in Indian subcontinent", *International Journal of Scientific and Technical Advancements (IJSTA),*Volume 1, Issue 3, pp. 23-28, 2015, ISSN No. 2454-1532.

[11] Gagandeep Jagdev et al., "Analyzing and Scripting Indian Election Strategies using Big Data via Apache Hadoop Framework", *IEEE Xplore*, DOI: 10.1109/WECON.2016.7993431, INSPEC Accession Number: 17061464, 27 July 2017.

AUTHOR'S BIOGRAPHY

Dr. Gagandeep Jagdev, is a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). His total teaching experience is above 11 years and has above 118 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of several international peer-reviewed journals and has been active Technical Program Committee member of several international and national conferences conducted by renowned universities and academic institutions. His field of expertise is Big Data, ANN, Biometrics, RFID, Cloud Computing, Cryptography, and VANETS.