International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 5, Issue 3, 2018, PP 1-8
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
DOI: http://dx.doi.org/10.20431/2349-4859.0503001
www.arcjournals.org

Augmenting Revenue Growth in Retail Segment via Data Mining

Dr. Gagandeep Jagdev

Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB)

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

Abstract: *Any data which poses a challenge for currently existing database technologies is termed as Big Data. This research paper is based on highlighting how retail industry makes use of big data in targeting customers. In today's busy life no one loves to receive unnecessary junk mails or messages. It is because of this that companies want to be crystal clear about each of its customer's interest and his or her pattern of shopping and expenditure. To be precise in knowing the interest of customers, companies needs to regularly mine the data related to each individual customer and gain insight of their shopping habits like which goods customer prefer and buy, customer's budget, customer's likes and dislikes etc. Future of retail is totally dependent on big data analytics as it is capable of separating wheat from the chaff. Retail is adopting data centric technology for boosting sales. Retailers have achieved 73% increase in sales with assistance of Big Data. Retailers are mining customers analytics to increase profits, increase growth and to be in competition, whether it is in-store or on-line. For handling such a huge database of millions of customers, companies make use of powerful framework like Apache Hadoop. Apache Hadoop is one such framework which is capable of handling huge databases via its several components. The retail market has always been interested in finding and analyzing new buying trends in customer's behavior. The central theme of this research paper is to highlight the entire process which is conducted by big retailers in different fields to attract customers and bond healthy relationships with them as satisfied customers are company's life time asset. The paper also elaborates how Apache Hadoop framework handles the enormous databases involved in retail industry with prime concern of increasing revenue growth.*

Keywords: *Apache Hadoop, Big Data, customers, economic growth, Map-Reduce, retail Industry*

1. INTRODUCTION

Different people have different understanding of Big Data depending upon their surroundings, which are prominently professional. Big data is the junction of more data from more sources than we have ever seen, it has also resulted in a new mode via which retailers associate themselves with the customers. This bottom-line impact of big data is what's making it business domineering and retailers have made their minds for using it to transform their processes, their organizations and, soon, the entire industry. Future of retail is totally dependent on big data analytics as it is capable of separating 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 on-line.

Fig. 1 below shows that data can come from different sources and can be of different types like retailer data, supplier data, market data, and shopper data. All this data flows in high volume and undergoes transformation and integration. This data is forwarded to analytical engine for further processing. This enhances retailer and supplier interaction and collaboration. All this gives results in positive aspects mentioned below.

- On shelf availability
- Budget and planning
- Category wise planning
- Shopper sensitivity
- Promotional effectiveness

- Introduction of new items
- Competitive awareness
- Holistic business planning

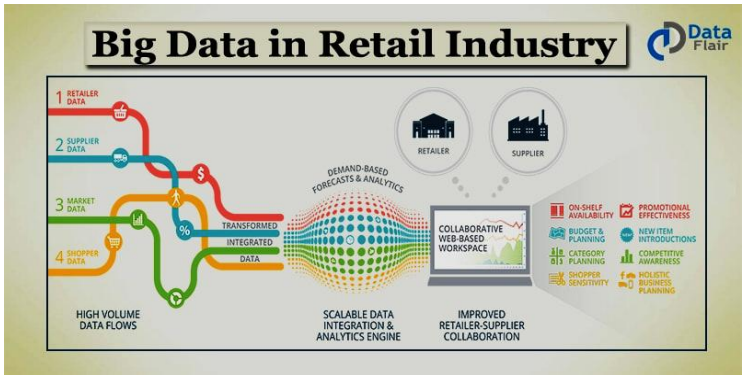

Fig1. *The figure depicts the process followed in retail industry*

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

- Amazon has brilliant and concentrated program to detect and prevent credit card frauds. This statement can be supported bythe 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 makes use of retail analytics to detect movement of goods inside the stores to provide on time information to concerned store personnel and customers for their convenience.
- Amazon respond 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 makes 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 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%.

Fig2. *Shows the growth of data from Megabytes to Petabytes as it parallel increases from ERP (Enterprise resource planning) to Big data via CRM (Customer relationship management) and web.*

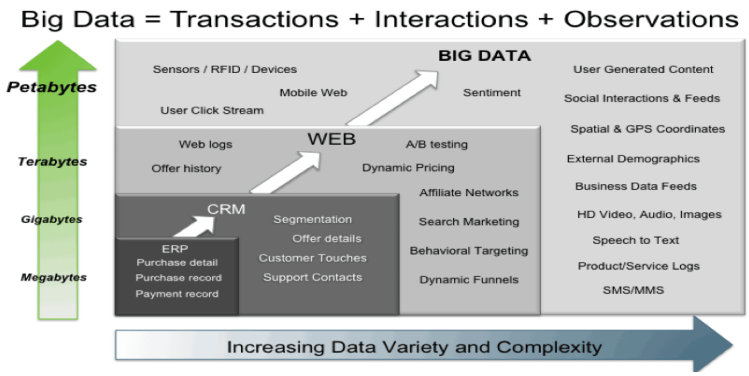

Fig2. *The figure depicts the growth of data*

Data analytics is not a new concept in 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 shows that the people who buy shampoo will invariably also buy a conditioner. Placing related products together influences the buying behavior of the customer [1, 2, 12].

2. EXAMPLES TO UNDERSTAND ROLE OF BIG DATA IN RETAIL

Shoppers Stop Ltd.

Shoppers Stop is among the major fashion chains in India. Three years back when they initiated their Big Data program, they were totally unaware about the gains it would bring them. It named this as First Citizen. Shoppers Stop started concentrating on targeted promotions for sale of trousers. This resulted in INR 10 crore additional sales within a period of three weeks. After studying it First Citizen database, company observed that not all those buy shirts also buy trousers. But those who buy both men's shirts and trousers spend 60% more a year on average than those who buy only shirts, and thrice as much as those who don't buy men's shirts at all. The company then came out with the list of 9 lakhs people for targeting them for sales of trousers. These 9 lakhs people were further divided into three groups.

The first group included customers showing interest in new brands in non-trouser categories. They were sent information regarding launch of new trousers.

The second group included those who exhibited multiple buying patterns in other categories. They were sent attractive deals if they bought two or more trousers.

Finally, the third was a "control group" to measure success or failure of the promotions. Control Group is a kind of a practice that is done by the company for all our analytics insights. The targeted communication exercise led to a lift of 30% in sales (aboutRs.10 crore) when compared with the response received from the control group.

Today Big Data analytics is a crucial part of the company's strategy. Big Data is engaged in bringing a major change in the dynamics of retail industry. This would lead to more intensive offers, promotions, store merchandise and layout as retailers spend more time to understand shoppers buying patterns. Retailers plan entire store activity around the shoppers rather than experimenting uselessly. Retailers have started using business intelligence to have more focused offerings for their clients rather than general promos.

3. CHALLENGES IN BIG DATA

The five major challenges [17, 18] faced by big data are mentioned as under and shown in Fig. 3.

- Volume –Volume refers to huge amount of data that is generated from variety of sources.
- Velocity – Velocity refers the speed at which the data is generated.
- Variety – Variety refers to the kind of data that is to be handled (structured, semi-structured, and unstructured).
- Veracity – Veracity refers to the preciseness of data.
- Value – Value refers to the worth that data carries.

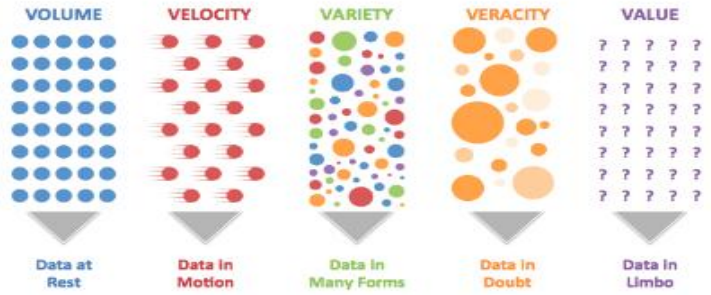

Fig3. *The figure shows the five challenges faced by big data*

4. BIG DATA ANALYTICS IN RETAILING

Important aspect of big data is more data you collect act upon it, you get more benefit. Data in retail sector is growing exponentially. Retail industry generate a flood of both structured and unstructured data, which when analysed helps the retailers to solve a lots of problems and led them better results like increase sale volume, reduce spoilage, improve launch success of product, helps in maintaining sale-demand chain [15, 16].

Retailer get data from vast number of sources some of them are following:

- Point-of-sale (POS) devices
- Websites
- Mobile commerce solutions
- Social media sites
- Video surveillance systems
- RFID and UPC readers
- Employee devices
- Other Sensors like Near Field Communication (NFC)

The data collected by different sources is combined and then analysed by using big data analytic tools. For example, if a famous doctor tweets that Aloe Vera juice is best for weight reduction, this lead the customers to run for juice and soon the stock will be finished. This problem can be solved by using big data analytic for sentimental analysis of social sites. The refined information by data analytic can be used for making better decision for the organisation which intends help in revenue growth.

5. APACHE HADOOP FRAMEWORK

Hadoop is a java based framework that is efficient for processing large data sets in a distributed computing environment [14]. Hadoop is sponsored by Apache Software Foundation. The creator of Hadoop was Doug Cutting and he named the framework after his child's stuffed toy elephant. Applications are made run on systems with thousands of nodes making use of thousands of terabytes via Hadoop. Distributed file system in Hadoop facilitates fast data transfer among nodes and allows continuous operations of the system even if node failure occurs. This concept lowers the risk of disastrous system failure even if multiple nodes become inoperative. The inspiration behind working of Hadoop is Google's Map Reduce which is a software framework in which application under consideration is broken down into number of small parts [5, 6, 10].

Hadoop framework is shown in Fig. 4.

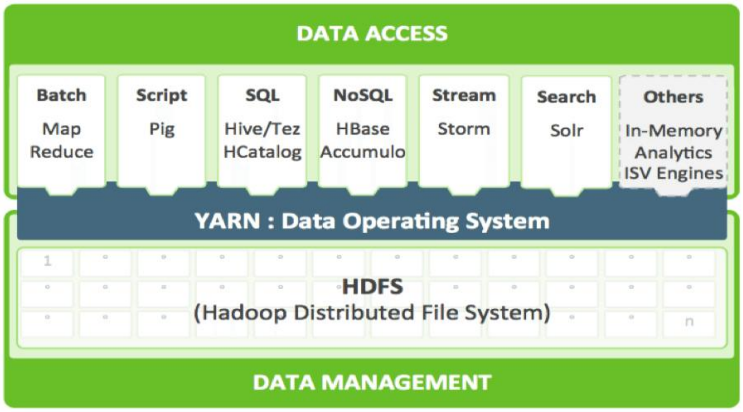

Fig4. *Hadoop Framework*

6. WORKING OF MAP REDUCE TECHNOLOGY

The flowchart depicting the working of Map-Reduce technology [9, 11, 13] is shown in Fig. 5.

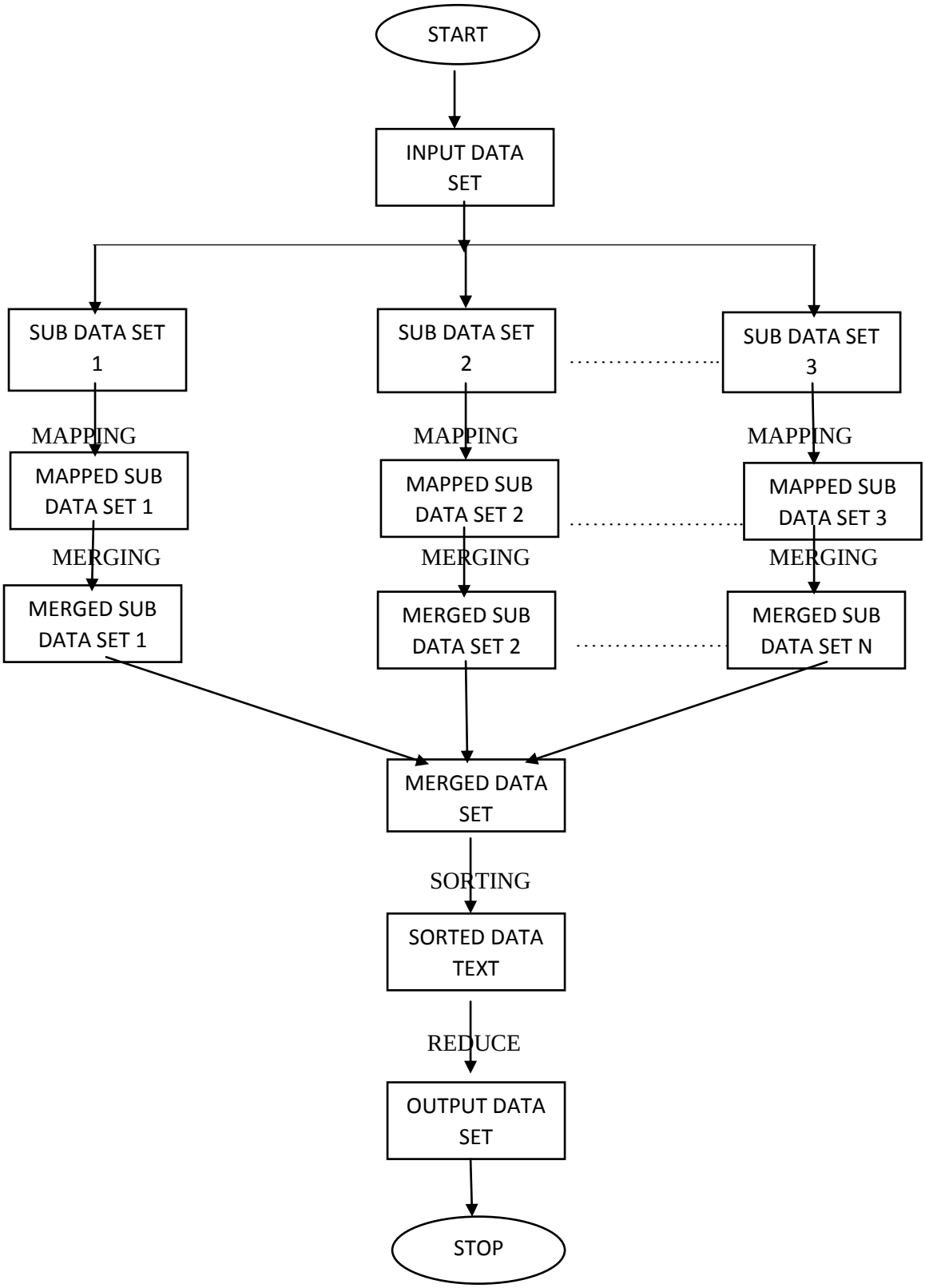

Fig5. *Flowchart showing working of Map-Reduce technology.*

Example

Let's consider an example to understand the concept of working of Map-Reduce technique. Suppose we have cell phones of four different mobile manufacturing companies along with their cost. Each file further consists of four different records. Our aim here is to find out the most expensive cell phone from the database under consideration.

File A:

Sony, 26000 Samsung, 25000 Micromax, 15000 Lava, 10000

File B:

HTC, 50000 Samsung, 25000 Micromax, 15000 Lava, 10000

File C:

Sony, 26000 Nexus, 45000 Micromax, 15000 Lava, 1000

File D:

Sony, 26000 Samsung, 25000 BlackBerry, 45000 Lava, 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> : <manufacturer, cost>.

File A:

<Sony, 26000> <Samsung, 25000> <Micromax, 15000> <Lava, 10000>

File B:

<HTC, 50000> <Samsung, 25000> <Micromax, 15000> <Lava, 10000>

File C:

<Sony, 26000> <Nexus, 45000> <Micromax, 15000> <Lava, 10000>

File D:

<Sony, 26000> <Samsung, 25000> <BlackBerry, 45000> <Lava, 10000>

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

<k: manufacturer, v: cost>

Max= the cost of first cell phone. Treated as most expensive

```
if(v(second manufacturer).cost > Max)
{
  Max = v(cost);
}
else
{
  Continue checking;
}
```

The expected result is as follows –

A - <Sony, 26000>

B- <HTC, 50000>

C- <Nexus, 45000>

D- <BlackBerry, 45000>

Reducer phase − Form each file, you will find the most expensive cell phone. The same algorithm is again used in between the four <k, v> pairs. The final output should be as follows <HTC, 50000>.

7. POINTS TO BE CONSIDERED BY RETAILERS

Certain points that retailers needs to be consider are mentioned as under.

7.1. Providing Up-Sell and Cross-Sell Recommendations

The most widely adopted use case in retail sector is up-sell and cross-sell recommendations. Retailers can promote thir product based on what products have been bought by similar customers. Big Data provides real-time capabilities that enable recommendations to be delivered at the right time and place to the right device

7.2. Social Media Analysis

Today there is need for retailers to monitor online sentiment of customers and respond in real time via offers or messages. This helps retailers to gain insights of customers purchasing behavior by analyzing their prior transactions and social network activity.

7.3. Dynamic Pricing Across Multiple Channels

It is obvious that when customers can shop across multiple channels in real time, even minor difference in price can make difference in their purchasing decision. Big data provides more refined set of indicators as compared to conventional influencers like time and availability. Some other pointers to consider are weather, location, social media presence etc.

7.4. Fraud Detection

Retailers need to detect fraudulent activities in order to safe guard 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.

8. CONCLUSION

In this paper, we the authors have introduced a vision of analytics as a new guiding principle for operating in today's tumultuous retail environment. There has been discussion on the power of becoming a data driven decision-making culture, and shown how access to accurate, scalable, and actionable data can help retailers set a roadmap to success through a better understanding of their customers and of their store operations. The concept of how data can reveal exposures as well as opportunities for the retailer have also been covered. Knowing who is not purchasing and why can be as important as understanding those who do purchase. The right insights enable a closer, stronger relationship with consumers.

REFERENCES

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

[2] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and *Challenges Moving Forward. International Conference on System Sciences (pp. 995-1004). Hawaii:IEEE Computer Soceity.*

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

[4] 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

[5] http://mike2.openmethodology.org/ , Last accessed : 14 September 2018.

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

[7] Sagiroglu, S.; Sinanc, D. ,(20-24 May 2013),"Big Data: A Review"

[8] Gagandeep Jagdev et. al.; "Affirmative aspects of Big Data assures new revolution in retail sector"; RTEST-2016 at Rayat-Bahra University; March 2016;ISBN: 978-93-84935-82-5.

[9] Gagandeep Jagdev et. al.; "Constructive aspects of Big Data in retail sector for financial growth"; CIOGEBS-2016 at Maharaja Ranjit Singh State Technical University, Bathinda; April 2016;ISBN: 978-93-5258-981-4.

[10] Gagandeep Jagdev, "Analysis and Application of Big Data in Retail sector intended for revenue growth" at National Conference on Future Patterns in Information Technology , D.A.V. College, Abohar (PB),2015.

[11] R. C. Taylor, "An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics," BMC Bioinformatics, vol. 11, no. 12, article S1, 2010.

[12] Edala, Seshu. 2012. "Big Data Analytics: Not Just for Big Business Anymore." Forbes.

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

[14] Gagandeep Jagdev et al., "APPOSITE UTILIZATION OF BIG DATA IN RETAILUSING HADOOP ENSURES ECONOMIC GROWTH", Proc. of Int. Conf. on Emerging Trends in Engineering & Technology, IETET, Grenze ID: 02.IETET.2016.5.

[15] Gagandeep Jagdev et.al, "Big Data in Retail Sector – An Evolution that Turned to a Revolution", International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), Volume 4, Issue 4, pp.43-52, DOI: http://dx.doi.org/10.20431/2349-4859.0404006.

[16] Gagandeep Jagdev et al., "A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm", in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O) – 2319-8354, ISSN (P) – 2319-8346,Volume – 06, Issue – 05, May 2017.

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

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

AUTHORS' 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 more than 12 years and has 125 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, Data Mining, Image processing in Diabetic Retinopathy, ANN, Biometrics, RFID, Cloud Computing, Cloud security, Cryptography, and VANETS.

Citation: *Jagdev, G. (2018). Augmenting Revenue Growth in Retail Segment via Data Mining. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(3), pp.1-8. http://dx.doi.org/10.20431/ 2349-4859.0503001*