Apposite Utilization of Big Data in Retail using Hadoop Ensures Economic Growth

Gagandeep Jagdev[1] and Ramandeep Kaur[2]

[1]Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB).
Email: drgagan137@pbi.ac.in
[2]Research Scholar (Ph.D.), Dept. of Computer Applications, Guru Kashi University
Talwandi Sabo, Punjab

Abstract—**Big Data is no longer an unpopular word anymore. Any data which poses a challenge for currently existing database technologies is termed as Big Data. Today only big giants in the different fields like retail, medical, stock exchange etc. can think of handling it because of expensive and huge infrastructure involved in handling it. This research paper is based on highlighting how retail and cellular industry makes its use 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 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 his or her shopping pattern like which goods he or she buys, customers budget, customers likes and dislikes etc. 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. It makes use of Map Reduce technology. Map Reduce technology is the same technology which is used by Facebook in handling its popular section "People you may know". The central theme of this research paper is elaborating the working of Map Reduce technology and Apache Hadoop framework in handling the enormous databases involved in retail industry.**

Index Terms— **Apache Hadoop; Big Data; Map Reduce; retail industry.**

I. INTRODUCTION

"Big data"– which admittedly means many things to many people – is no longer confined to the realm of technology. While big data is the convergence of more data from more sources than we have ever seen, it also represents a cultural shift in the way retailers connect with consumers in a meaningful way. This bottom-line impact of big data is what makes it a business imperative and why retailers around the world are leveraging it to transform their processes, their organizations and, soon, the entire industry. Much more data can be derived as businesses and consumers move into the realm of pervasive connectivity that defines the Internet of Everything (IoE) world. IoE combines people, processes, data, and things to enable the transfer of information and create new potential for business innovation. Sensors and devices offer up data from

previously unconnected processes and their components, expanding the role of data in decision making across the entire retail enterprise [4].

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.

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.

II. CHALLENGES RELATED WITH BIG DATA

Data Volume – It refers to the enormous amount of data that is been created each second, each minute and each hour of the day. 571 websites are created in a single minute. Total of 625000 GB of data is transferred from one end to another in single internet minute, may be terms of mails, pictures, posts etc. If we burn the amount of data present on planet earth today on DVDs and pile them in the form of a stack one upon another, the pile will be such huge that one can climb it and touch the moon, come back to earth and again repeat this process once [10].

Data Velocity – Data is being created at such high velocity that companies are finding it difficult to cope up with such high speed. They have to establish their infrastructure in such a manner that it is capable of handling such generated data Social media, E-Commerce has rapidly increased the speed and richness of data used for different business transactions [8].

Data Variety - All the data being generated 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 [10].

Data Value –There is a huge gap in between the business leaders and the IT professionals. The main concern of business leaders is to just add value to their business and to maximize their profit. On the other hand, IT leaders deal with technicalities of the storage and processing.

Data Complexity – The biggest complexity faced while running big data using relational databases is that they require parallel software running on hundreds of servers and data scientists have to match and transform data across systems coming from various sources [10].

Data Veracity - Veracity refers to the preciseness of data or how much faith one can have on data. The data on internet is not always accurate or precise. For example, if some male pretends himself as a female on his facebook profile, there is no authenticity check in such cases. Similarly twitter makes use of abbreviations and hash tags, but big data enables us to work with even this type of imprecise data [1, 6, 9].

III. 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 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 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%.

A. Shoppers Stop Ltd.

Shoppers Stop is among the major fashion chains in India. Three years back when they initiated their Big Data programme, 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 (about Rs.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.

IV. FIVE WAYS BIG DATA IS REVOLUTIONIZING RETAIL MARKETING

A. 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 retailers have extraordinary information about customers today and they can use this information to full potential to build healthy relationship with customer and ensure long term benefits [9, 10].

B. Boosting investment in technology

Storage today has become so cheap that companies have no more issues related to cost of storing enormous data. Retail giants have invested in building 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].

C. Solving the omni-channel puzzle with data

Retailers having data centric attitude are gathering unbelievable amount of data related to customer behavior to have a deep insight of how 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. Retailer needs to concentrate on every possible channel. Marketers should not ignore one channel at the cost of other [3].

D. Improving personalization

Retailers need to build 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 customers on right time, at right place and via right channel. This is possible only if company conducts periodic personalization of their customers.

E. 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].

V. APACHE HADOOP FRAMEWORK

Hadoop is a java based framework that is efficient for processing large data sets in a distributed computing environment. 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 is a framework which comprised of six components [4]. Every component is assigned a particular job to be performed.

- HDFS – HDFS are distributed cages where all animals live i.e. where data resides in a distributed format.
- Apache HBase – It is a smart and large database.
- Zookeeper- Zookeeper is the person responsible for managing animals play.
- Pig – Pig allows to play with data from HDFS cages.
- Hive- Hive allows data analysts play with HDFS and makes use of SQL.
- HCatalog helps to upload the database file and automatically create table for the user.

VI. WORKING OF MAP-REDUCE TECHNOLOGY

The flowchart depicting the working of Map-Reduce technology [7, 8, 9] is shown in Fig. 1.
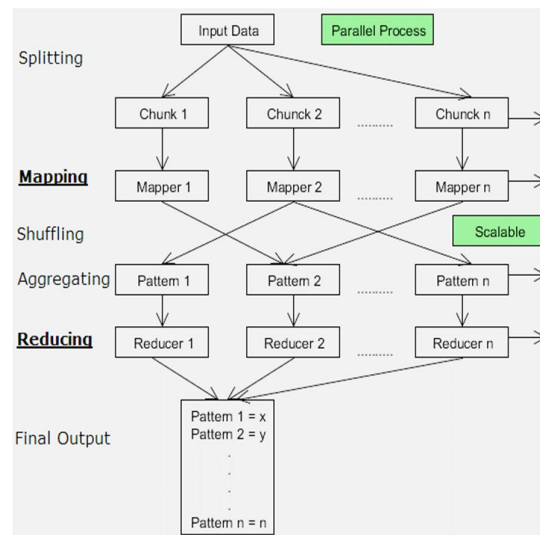

Fig. 1. Flowchart showing working of Map-Reduce technology

Algorithm for Map-Reduce

- The input data can be divided into n number of chunks depending upon the amount of data and processing capacity of 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 increasing the number of the parallel processing units.
  Example (a)

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
_____

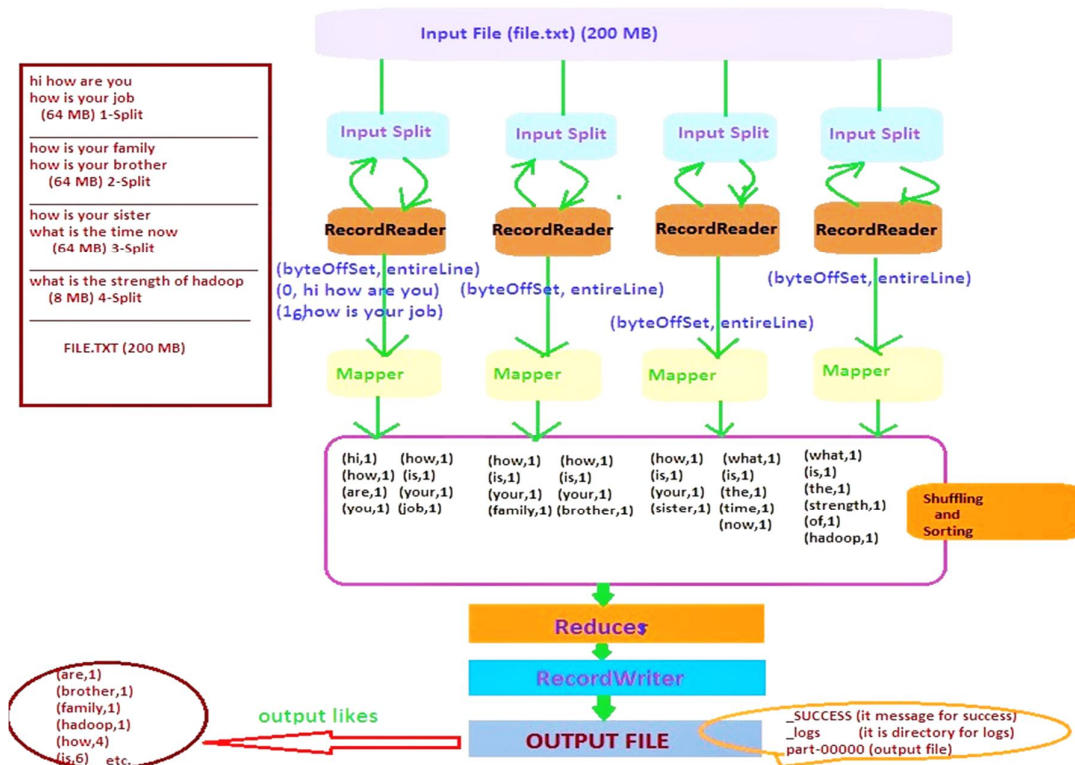what is the strength of hadoop (8 MB) 4-Split

Fig. 2. Pictorial representation of Example (a)

66

VII. CONTRIBUTION AND IMPLEMENTATION

A day will come when the drones will be zooming through the skies to deliver us with the commodities that we have not even ordered yet. Although we are not at the stage of drone delivery and this future is still a long way to come. But if we talk about current state, then it can be stated that retailers are constantly engaged in finding out innovative methods and ways to have a deep insight of ever-increasing structured and unstructured information about customers behavior and buying patterns. So this will be how Big Data and analytics will be changing the retail landscape in the coming future. Today company's tries to predict what will be the popular products in coming future, forecasting the locations where the demand for the product would be maximum, optimizing their product costs to survive market competition, targeting customers more probable to be interested in their products and finding out the best way to approach them. Smart and intelligent retailers are already reaping dividends. In our research work, we have created a structured database having 20 attributes which are mentioned as under with their details.

*Product *Type *Category *Start-Range *End-Range *MRP *Season
*Month_of_max_sale *Mfg_year *Exp_year *Guarantee_in_years *Warranty
*Company *Territory *Offers *COD *Delivery_charges *State/Province
*Deliver_accept *Likes

The glimpse of the snapshot of the structured database are shown below (Fig. 3).

	A	B	C	D	E	F	G	H	I
1	PRODUCT	TYPE	CATEGORY	PRICE_RANGE_RUPEES	MRP	SEASON	MONTH_OF_MAXSALE	MFG_YEAR	EXP_YEAR
2	Flow Excel Bluetooh Multimedia Speaker Sytem	NonConsumeable	Electronic	2000_5000	5200	Any	January	2016	2018
3	Lenovo Vibe K5 Plus	NonConsumeable	Electronic	7300_9999	10000	Any	March	2016	2017
4	Insect and Mosquito Killer with Night Lamp	NonConsumeable	Electronic	149_449	503	SUMMER	June	2015	2019
5	Bridgestone B390 205/65/15 Tubeless Tyer	NonConsumeable	Car_Tyer	5300_8400	9000	Any	May	2016	2026
6	FAB M 70W 1200mm Ceiling Fan	NonConsumeable	Electronic	999_1700	2000	Summer	March	2016	2018
7	EF03 3 Blade Table Fan	NonConsumeable	Electronic	2500_3500	3900	Summer	March	2016	2017
8	Trans Air 4 Blade Exhaust Fan	NonConsumeable	Electronic	1400_1700	1999	Any	May	2016	2017
9	Aura Metallic 47" 1200 cm Ceiling Fan	NonConsumeable	Electronic	2000_2300	2500	Summer	June	2016	2018
10	Formula 1 Nutritional Shake Mix	Consumeable	Health_Drink	1200_1800	2300	Any	May	2016	2017
11	Lazer U11300In All Purpose Home Blower Tower Fan	NonConsumeable	Electronic	2800_3800	4100	Any	All	2015	2017
12	Aqua supreme RO_UV_UF_TDS 18 Ltr Water Purifiers	NonConsumeable	Electronic	4000_8000	10000	Any	June	2016	2017
13	GC1010 Steam Iron	NonConsumeable	Electronic	1000_1400	1700	Any	All	2016	2018
14	16 litre Unbreakable Non-Electric Water Purifier	NonConsumeable	Non_Electric	700_1200	1544	Any	June	2016	2017
15	HI114 Dry Iron	NonConsumeable	Electronic	800_995	1156	Any	All	2016	2018
16	Castor Wheels Air Cooler	NonConsumeable	Electronic	9000_13555	19000	Summer	May	2016	2017
17	Super Cool CP3001H Air Cooler	NonConsumeable	Electronic	6600_9000	11000	Summer	May	2016	2016

Fig. 3. Snapshot of constructed database

VIII. POINTS TO BE CONSIDERED BY RETAILERS

- 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 their product based on what products have been bought by similar customers.
- 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.
- 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.

IX. CONCLUSION

In this paper, we have introduced a vision of analytics as a new guiding principle for operating in today's tumultuous retail environment. We've discussed 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. We've also covered how data can reveal exposures as well as opportunities for the retailer. 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/
[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] Dr.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] Dr. 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] Dr. 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.