

A Review on Big Data Analytics in Business

Manbir Sandhu*, Purnima, Anuradha Saini

Department of Computer Science, Sri Guru Gobind Singh College, Panjab University, Chandigarh, U.T., India

ABSTRACT

Big data is a fast-growing technology that has the scope to mine huge amount of data to be used in various analytic applications. With large amount of data streaming in from a myriad of sources: social media, online transactions and ubiquity of smart devices, Big Data is practically garnering attention across all stakeholders from academics, banking, government, health care, manufacturing and retail. Big Data refers to an enormous amount of data generated from disparate sources along with data analytic techniques to examine this voluminous data for predictive trends and patterns, to exploit new growth opportunities, to gain insight, to make informed decisions and optimize processes. Data-driven decision making is the essence of business establishments. The explosive growth of data is steering the business units to tap the potential of Big Data to achieve fueling growth and to achieve a cutting edge over their competitors. The overwhelming generation of data brings with it, its share of concerns. This paper discusses the concept of Big Data, its characteristics, the tools and techniques deployed by organizations to harness the power of Big Data and the daunting issues that hinder the adoption of Business Intelligence in Big Data strategies in organizations.

Keywords : Big data, Business Intelligence, Data Analytics.

I. INTRODUCTION

Digital revolution, Internet of Things (IoT), consumerism and the ubiquity of connected smart devices generates prodigious data, referred to as *Big Data*. Data is an asset for any business organization. It is crucial for the strategic decision making of an organization :to identify challenges, to capitalize on opportunities and to predict future trends and behavior of the customers This gargantuan data is automatically generated at various customer interaction points from e-commerce ,social platforms like Facebook and Twitter, audio/video calls and search engines, websites etc **The seven properties that well describe the data are :Volume, Variety, Velocity, Value, Variability, Veracity and Visualization. Owing to data's substantial size (Volume), the traditional database systems have

limited usage for its processing. The complex data can be represented (Visualization) by graphs or charts rather than through the conventional spreadsheets for better understanding [11]. The data generated from disparate sources (Variety) comes in diverse formats (Variability) ranging from unstructured, structured to semi-structured; meaningful or meaningless (Value); accurate or inaccurate (Veracity). The Velocity of data generation can vary considerably across numerous communication channels of an organization from seconds, minutes, hours, days or weeks to months. To make informed choices, the data has to be structured, reliable, timely available and stringently analyzed .The prime challenge for the organizations is the management of the exponentially growing data, its meaningful analysis, deploying low cost processing tools and practices while minimizing the potential risks relating to safety, inconsistency,

redundancy and privacy. Consequently, Big Data and its analytics is gaining central ground in all business establishments for the efficient utilization of resources: storage and time and efficient decision making to exploit new avenues and maximize profits.

II. TOOLS AND TECHNIQUES

There are various tools which can be used for data management from data acquisition, data storage to data visualization. In business applications, almost all the things rely on the decisions of the officials which are further dependent on the data mining techniques, algorithms and frameworks. With the accumulation of the data mining techniques with the computational models and frameworks, one can acquire the capability of making perfect choices in the business. [7]

Every business necessitates several tools to examine and assemble dissimilar data types as mentioned in the Figure 1 from diverse sources. If they acquire command to analyze data of every size, type then end result will provide most reliable information about business trends, values and patterns. This section describes those tools which are used for different purpose are described below:

R Tool: This is the most commonly used data visualization tool for graphic and statistical computing in the field of data science, which has the deposit of modeling facilities and also, it offers a superior environment for the composition of data models [3]. This is supported by the R Foundation for Statistical Computing [5] and is widely used among statistical area and data miners for data analysis.

No SQL: It is the category of storage engines that store data in non-relational format in the area of data science and retrieves the data modeled in means other than tabular relations used in relational databases.[5] They are sometimes called “Not Only SQL” to emphasize that their scope is not limited to SQL like

query languages. There exists different kinds of No SQL databases like key-value store, graph db etc.

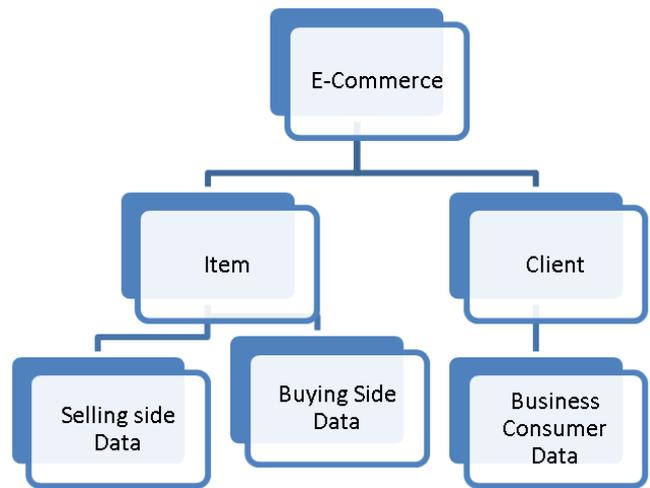


Figure 1 : Data Types for Decision support

Tableau: It is the tool which emphasizes on the visualization of end product in the form of charts, maps, graphs, diagrams and many other graphical representations.[5] The implementation of this tool is explained by a desktop application that employs visual analytics.

Map Reduce: The Map Reduce programming environment authorizes the execution of the bigger jobs scalability and its implementation is divided into two main parts; map task and reduce task. The Map task changes the input dataset into a dissimilar set of values whereas the Reduce task combines numerous outputs of the Map task to form compact rows. [5]

Cloud Vista: Clustering algorithm is an eloquent job where there is a need to categorize a fixed set of groups or clusters to illustrate the data [8] and is closely related to the task of probability density estimation. [5]The most popular clustering tool is Cloud Vista which is used in cloud computing to implement the clustering process in parallel. BIRCH and other clustering methods are used in Cloud Vista to show that can be handle very large scale data.

Regression: A technique for evaluating statistical relationship between a dependent variable and a series of independent or changing variables and plots a data item to a real-valued prediction variable which is explained by building a model to predict the probability of fraud transactions on a given a data set of credit card transactions.[8]

III. ISSUES AND CHALLENGES

Outlay: In small business, usually the outlay is limited. In small- and medium-sized enterprises, the expense of expressing the business intelligence service is an indispensable concern. These enterprises are incapacitated by the restraining cost of the programming and tools needed for the implementation of this service. Similarly, the skilled experts also involve the expenditure else the inhibited means of identifying the well-equipped professionals. New Entrepreneurs should make a wise choice of selecting the business intelligence service with flexible understanding. Nowadays, the self-service business intelligence plays an important role to provide considerable amount of money for meeting the needs of the companies. In recent years, it has been observed that the cost of business intelligence is no more a challenge but considered as a help after the introduction of the stages of Self-Service Business Intelligence. [4]

Illustration: Sometimes the dataset has definite levels of dissimilarity in structure, semantics, type, organization, granularity and accessibility, the aim to represent the data properly for data analytics and user analysis is very significant [1]. Any improper data representation may reduce the value of the data originality and even disturbs effective data analysis process [5]. Businesses may put resources into enormous data investigation but can't finish the assignments in time. Entrepreneurs are swinging to imaginative BI devices to address this business intelligence issue. They may result in individuals investing hours in cleaning and organizing the data

first and after that, utilizing the BI service. These instruments can undoubtedly blend distinctive datasets on-the-fly without the need to rebuild databases or set up a data distribution center. A Business Intelligence service could be stacked with programmed ETL abilities to process datasets that should be rebuilt. This permits small businesses to interface all their data sources, see past the numbers, find new connections, and recognize patterns to remove the mystery from critical business choices.

Privacy: It is another big concern in this area as the service owners and owners of the data could not maintain and investigate such huge datasets successfully [5] [9]. At times, they require the help of skillful people or other tools to examine such data resulting in the increase of the chances of probable security risks. Thus, there is a need to create the business intelligence service and computation more secure and safe. The major challenges are secure data storage, data integrity, data availability and data backup.

Data complexity: It refers to the specific features of data including huge size, high dimensionality, and extreme imbalance, online and real-time interaction and processing, cross-media applications, mixed sources, strong dynamics, high frequency, noise mixed with data, unclear structures, imprecise hierarchy, heterogeneous or uncertain distribution and unclear availability of specific sometimes critical data.[1] This is also an essential issue for business intelligence merchants as it involves the complex relations hidden in data. [2] There exists an elementary dilemma to develop or explain the vital attributes of the data complexity. In current scenario, the knowledge about the theory of complexity will facilitate the users to realize the necessary characteristics and creation of composite computing models, simplify its representation on big data. In order to do this, we need to arrange the internal relationship between data complexity and computational complexity. Moreover, by modeling

and exploring the intrinsic mechanisms of data complexity, we will be able to illustrate the principles and mechanisms for processing the big data into a firm founding for computation.

Computation Complexity: One of the main challenges that organizations face even after having the strong business intelligence system and an infallible machinery arrangement is the specialized abilities for the execution of the same.[9] Another factor that kept small and medium enterprises away from business intelligence is obvious necessity of comprehensive training. It is tough to analyze the data due to the unavailability of the well-equipped tools and manpower. Thus the superfluous and professional training is essential for data monitoring, maintaining and managing. The organizations must be aware about the importance and services of business intelligence. It is required to invest intelligently in the training and computational process, which will help an organization to select the best tool or technique of big data, best suitable for the current scenario. [4]

Authenticity & Power Management: In current years, some networks provide redundant data like sensor based networks which needs to be filtered and duplicity may be reduced in order to give make the data analysis or data processing more effective[8][10]. In processing of data, huge input datasets and their storage management as well as analytic processes are involved which undoubtedly consume more electric energy. [5] Hence, power consumption management mechanism may be implemented for the data and its processes to make sure that the less energy is consumed.

IV. CONCLUSION

In the age of big data, everyone needs recent and efficient equipments along with various tools, techniques and algorithms to deal with the issues and challenges of the business intelligence while working with the huge data. [6] Big data analytics is one of the

reasons for the universal success of any business organization. Big companies are merging and combining the big data analytics with traditional analytics that will affect the organization's abilities, guidance, arrangements and technologies. The organizations which are still not familiar with the techniques in big data analytics are likely to be visually and physically handicapped as they would undergo monetary losses in terms of their upcoming clients and enhanced prospects in making investments. The origin of big data reveals the inadequacies of existing data mining technologies which in turn raised new challenges. In this paper, we have presented a very brief overview of big data, its attributes, the most accepted big data processing techniques; NoSQL, Map Reduce and Tableau is given which helps researchers and data scientists to investigate the big data and discover unseen and unfamiliar prototypes as well as the issues and challenges which the organizations face while they come across the concept of business intelligence in big data. [7].

V. REFERENCES

- [1]. Longbing Cao, Data Science: Challenges and Directions, August 2017, Vol. 60, No. 8, Communications of the ACM, pp: 59-68.
- [2]. Matsudaira, K, The science of managing data science, June 2015, Vol. 58, No.6, Communications of the ACM, pp: 44-47.
- [3]. <https://datajobs.com/what-is-data-science>
- [4]. <https://dzone.com/articles/the-5-biggest-business-intelligence-challenges-fac>
- [5]. M.D. Anto Praveena, Dr. B. Bharathi, A Survey Paper on Big Data Analytics, International Conference On Information, Communication & Embedded Systems (ICICES 2017).
- [6]. Mudasir Ahmad Wani, Suraiya Jabin Big Data: Issues, Challenges and Techniques in Business Intelligence.
- [7]. Jafar Raza Alam et al, A review on the role of Big Data in Business, International Journal of

Computer Science and Mobile Computing,
Vol.3 Issue.4, April- 2014, pg. 446-453.

- [8]. Panagiotis Barlas Ivor Lanning Cathal Heavey , (2015), "A survey of open source data science tools", International Journal of Intelligent Computing and Cybernetics, Vol. 8 Issue 3 pp. 232 – 261.
- [9]. Xiaolong Jin, Benjamin W.Waha, Xueqi Chenga, Yuanzhuo Wanga, Significance and Challenges of Big Data Research, in: Big Data Research 2, Elsevier Science, 2015, pp: 59–64.
- [10]. Radwa Elshawi, Sherif Sakrb, DomenicoTaliad, PaoloTrunfiod, Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service, in: Big Data Research, Elsevier Science, 2018.
- [11]. Simoff , S., Bohlen, M.H., Mazeika, A.: Visual data mining: theory, techniques and tools for visual analytics, vol. 4404, Springer Science & Business Media (2008).