



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: I Month of publication: January 2020

DOI: <http://doi.org/10.22214/ijraset.2020.1031>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Management Challenges in Big Data – A Study

G. Sandra.

Assistant Professor, Sree Narayana Guru College of Advanced Studies, Cherthala, Alappuzha, Kerala. 688582

Abstract: At every moment huge amount of data is generated in this digital world, which is called Big Data. Its volume is increasingly exponentially. This big data is the treasure house of opportunities, from which valuable insights and knowledge discoveries are made. They are used by different sections of society for their prosperity. In spite of the benefits conferred by big data, the modern information systems are confronting with many challenges also from it. Scholars have broadly classified them into data challenges, process challenges and management challenges. In this study an attempt is made to survey them by giving greater emphasis to management challenges.

Keywords: Big Data Challenges, Data Challenges, Process Challenges, Management Challenges.

I. INTRODUCTION.

The proliferation of information technology has resulted in the spontaneous growth of digital data. An astounding growth of digital data can be seen in sectors of business, manufacturing, science, personal lives etc. The digitalization from the use of internet, smart phones, social networks, streaming technologies, etc, has resulted in the huge volume of digital content. The magnitude of this type of data generated and shared by the stakeholders has increased immeasurably [1]. Organization can profitably use these data for their prosperity through processing and analyzing them in an efficient and befitting way. But this task is too tedious. The traditional tools are inefficient to process them optimally and to gain new insights and discovery of knowledge. This necessitates the need for developing efficient modern technology for processing and analyzing the ever growing datasets whose size is too big and beyond the managerial efficiency of the traditional techniques. This exponential growth of digital content is called Big Data.

Scholars have defined the concept of Big Data differently. Popularly it is believed as the data that should include the datasets with huge size beyond the ability of commonly used computer system to process and manage within a reasonable time limit. Big data is the treasure trove of valued insights, which are extensively used by the corporate for business decision making, to improve their operational efficiency, explore new revenue sources and gain competitive advantages over their rivals. The velocity of this treasure trove is more or less doubling with in every two years [2]. It is reported that there produces around 2.5 quintillion bytes of data every day in this world [3]. It is expected that by 2020, the digital data generation would be 40 zettabytes [4]. In short, a data deluge is created from different sources, which contain structured, semi-structured and unstructured data. They are of textual content and multimedia content on a multiplicity of platforms. It is believed that big values reside within big data, which scholars considered as today's digital oil [5]. Big data is offering many opportunities and challenges. The opportunities include value generation, advanced business intelligence for decision making, market researches, rational consumer behavior analysis, optimal resource allocation, etc. [6]. Like opportunities, the challenges are also equally significant. The prominent among them are data complexities, data integration, data security, privacy issues, lack of resources, lack of skilled persons, data mining and cleansing, data warehousing, data aggregation, data interpretation, data analysis, data modeling, data ownership, data governance, etc. This study is made an earnest effort to survey the widely discussed big data challenges.

The study is organized in the following way. The section II summarizes the classification of big data challenges. Section III deals with a comprehensive review of the management challenges confronted by the data scientists. The section IV concludes the study with suggestions for future researches.

II. CHALLENGES OF BIG DATA.

The accumulation of big data has been pervasive. From different domains, it has been accumulated, like retail business, public administration, health, scientific researches, web-based applications, internet text, social computing, IoT, etc. The big data from these domains offers many insights and knowledge, but the opportunities provided by the big data followed untold challenges also. Researchers have examined them in many ways and a broad classification of them is: data challenges, Process challenges and management challenges [7].

The data challenges are emanating from the characteristics of big data itself. From time to time various dimensions of big data are proposed by scholars such as volume, velocity, variety, veracity, value, variability, visualization etc [8]. The large datasets generated from different domains consisting of different big sizes of terabytes, petabytes, zettabytes or more than that. They are the source of different challenges. This sheer volume of data is heterogeneous, ubiquitous and dynamic in nature. To make them

determining, retrieving, processing and integrating is a herculean task [9]. The accumulated big data are diverse and dissimilar in forms. They are collected from diverse sources; like messages of text, e-mail, blogs tweets, transactional data, web data, user generated content, scientific data, etc. They are diverse in nature and a big challenge to comprehend and manage [10]. Since the collected large volume of datasets is complex in structure, anonymous, imprecise and inconsistent, scholars have expressed doubt about its data quality and accuracy, otherwise called, the veracity of the accumulated data [11]. To deal with this type of inaccurate and ambiguous data and managing and mining them with appropriate tools is always a great challenge. The high rate of inflow of data or velocity of big data with non-homogeneous structure creates many hardships to data scientists. They are forced to develop real-time analytic tools to processes and analyze the complex data. For gaining insights and discover knowledge, the accumulated data has to put in data mining process, but every time the data offers different meaning. The data while used for mining process whose meaning is constantly changing is called variability of data. For example, in the same tweets, a word can have different meaning. For proper processing and analysis of big data, this is very challenging [12]. To make sense out of big data and develop knowledge, it is essential to present data in a readable manner. This process is known as visualization. For this different visual formats have to be developed. It is essential that big data visualization tools must be capable of transforming large and complex datasets into natural depictions. But at present the big data visualization tools have only poor performances in functionalities and scalability [6]. Extracting insights and knowledge or value from the unstructured and complex accumulated data without loss is a major challenge of the data scientists. They are also facing the problems related to store, manage and extract high valued data in a cost effective manner.

In big data challenges are developed not only from the dimensions of data but from the processing and analyzing it also. Since the datasets of big data are mainly unstructured in nature, processing them is really a significant challenge, which is more difficult than managing it. Scholars have identified them and grouped it into data acquisition challenges, warehousing challenges, data mining and cleansing challenges, data integration and aggregation challenges, data analysis challenges, data modeling challenges and data interpretation challenges [13]. The innate complexity of big data and its exponential growth lead to unforeseen problems in big data analytic like the data acquisition and storage of it for value generation. Management of big data is very complex. Organizations have to deal with high degree of inaccurate and disparate data. While undertaking challenges from the sheer amount of data generated; the lack of scientific processing systems, data management systems and inadequate data strategies and fragmented data ownership are the crucial issues associated with it. The other prominent barriers to big data analysis are the unawareness about the data ownership, the data scaling, the capability of the system to analyze the rapidly growing data and the resources to store them for future use. This will affect the efforts to derive valuable information from the data analysis. To capture the useful information, data scientists have to eliminate the unwanted data and inconsistencies of the collected data. This challenge is related to data mining and cleansing of the large scale unstructured data collected. This effort is very challenging because of the diverse and unreliable features of the big data [14]. So to use the huge data collected in a fruitful way, scientific extraction methods have to be developed, which is again a most challenging task. While the unstructured data is mined, then we have to aggregate and integrate the cleaned data, which is very much problems ridden. Since the unstructured data is diverse, vibrant and inter-related in nature, they lack meaningful information. The aggregation of such data is beyond the ability of the current data integration technique [15]. Once data is mined, cleaned and integrated, there comes the challenge of analysis and modeling of big data. Since the collected data is too complex, heterogeneous, dynamic and unreliable, the traditional techniques of data modeling cannot be applied effectively because to process huge data requires unprecedented storage capacity and computing efficiency [16]. To make the data understandable for users, the analyzed data should be presented to the decision makers in a legible way to interpret them for knowledge discovery and extracting insights. For this, appropriate technology and trained people with sufficient analytical skills and experience are the needs of the time, which is also too challenging. Having discussed the data challenges and process challenges of big data in brief, we are striving to discuss the management challenges of big data in a comprehensive manner in the next section.

III. MANAGEMENT CHALLENGES.

The group of challenges regarding accessing, managing and governing are referred as management challenges. It comprises challenges related to security and privacy, information sharing, cost efficiency, data ownership and data governance.

Security of data aims to restricting the data use only to the authorized users. The task of preserving sensitive information is a great challenge to big data management. The big data is subjected to high security risks [17]. Strict security in data warehouse can be offered by restricting the data to unauthorized access, disclosure, disruption, inspection, modification and destruction. If data use is not restricted, then it is very easily to be attacked. Data privacy is a major security issue. While in collecting and storing, the data privacy can be jeopardized. If a malicious node gains access to the data collected, then there is the chance to steal or alter the data.

Data privacy is related to how to secure to use the data, what information is stored, where and how and the way in which data is handled. The security of big data is mainly concerned with the confidentiality, integrity and availability of data. The most crucial security challenges of big data are; vulnerability to fake data generation, presence of un-trusted mappers, weakness of cryptographic protection, possibility of sensitive information mining, data provenance difficulties and lack of security audit [18]. The security challenges of big data have caused to develop many security measures for access controls, authentication, differential privacy, identity based anonymization, encryption methods based on attribute, homomorphism, storage path and hybrid clouds. But still the security issues and concerns of big data are quite crucial. Hence attempts have to be made to develop multi-level security and privacy preserved data models.

Sharing data and information is a critical issue to big data management. It is an act making the same data and information owned by one entity available to another voluntarily. Both analog data and digital data are shared in this way. It encourages more connections and collaborations between users. It has greater impact on others and remains an important component of information behavior. It is an unavoidable activity of all collaborative works and helps to bind groups and communities together. When information is not effectively shared, the collaborative group work fails [19]. For sharing data, it has to be organized in a useful way. The problem of copy rights and licensing come across in the sharing of data. It is not easy to know which repository to use the data. There is the lack of time to deposit data and also the problem of cost sharing of data. There is also the challenges of confidentiality and trust of sharing data. People voluntarily give data but have a deep concern about its misuse. The challenge of privacy also remains always with the sharing of data and information.

Managing big data operational cost efficiently is a critical task. When the data is constantly rising, there is greater demand for big data processing for knowledge discovery. This has led to the increase in the operational cost. The cost escalation comes from different data intensive operations such as acquisition of data, warehousing of data, mining and cleansing of data, aggregation and integration, data processing and interpretation. Hence the data scientists confront with operational challenges from different angles. To them the data quality remains an issue.

They must know how to deal with data, how to organize its storage, how fast it is likely to grow and change and how long the data is relevant and valuable. To overcome these operational challenges, organizations have to introduce automation than manual efforts. Hence developing cost efficient techniques is an emergent challenge [20]. The cost minimization of operational expenditures of the data processing centers is a sensitive issue to various organizations. To leverage the benefits of data explosion, they need to train IT experts properly.

To take right decisions regarding better management and governance of data is a major challenge to any organization. This challenge is wholly related to data ownership. Data owners are those who are structured into the organization through their roles like a system owner, manager of running operations, project manager or processes leader. But most of the owners are not knowledge workers and have no operational awareness. They have not much skill about how data are being applied for operational value. They are not enabled to take decisions. Data owners enforce policy and manage risks related to their data. There are many challenges data owners may face while sharing real time data, especially in social media, like who owns the data, for example, in Facebook or Twitter. Data ownership is a much crucial issue. It is believed that both the user and the social media provider own the data. But this issue is not yet settled and this dichotomy is still persisting [14], which requires more research. To orchestrate data governance activities, the data owners require knowledge about data management and governance, processes and techniques, which remain a crucial challenge all concerned. Along with this, it is also a challenge for them in leveraging metadata management and data quality services.

To ensure the quality of mined data, effective big data governance is absolutely essential [21]. It is the process and management of data that enables an organization to leverage data as an enterprise asset. Raw data has no value. Only when it is refined, it can be an asset to the organizations. It includes all the arrangements starting from storing of data to secure it from any mishap. The data governance is the mechanism to generate critical insights with which the business decision can be improved. But this task is big challenge in big data due to the huge amount of data, issues around the data quality, structured and semi-structured nature of data. The data governance provides the users the mechanism to turn the raw data into real intelligence. But in this effort they have to overcome issues of accessibility, usability and quality of data.

They have to manage the entire data supply chain without compromising its quality. Users do not know whether the data can be trusted, accurate, consistent and reliable. They do not have the knowledge where the data come from, what it is and who owns the data. So every organization has to develop comprehensive data governance programs to solve the above challenges so that the organizational data will be reliable, understandable and usable.

IV. CONCLUSION.

The objective of this survey is to give an overview of the management challenges of big data within the broad perspective of big data challenges. The paper highlights the major management challenges of big data such as security and privacy issues in data profiles, the information sharing issues, cost effectiveness and data ownership challenges and data governance difficulties to be faced in solving big data challenges. The study reviewed the efforts on the part of the researchers in proposing solutions. But various issues related to big data challenges are still moving forward; hence greater efforts are required to fill the research gaps.

REFERENCES.

- [1] R. Agarwal and V. Dhar, "Big data, data science and analytics: the opportunities and challenges", information system research, 2014, 25(3), pp. 443-448.
- [2] IDC, "Big Data in 2020", 2012, Retrieved from <http://www.emc.com/leadership/digital-universe/2012-view/big-data-2020.htm>.
- [3] C. Dobre and F. Xhafa, "Intelligent service for Big Data science", Future generation Computer System, 2014, 37, pp. 276-281.
- [4] J. Gantz and D. Rensel, "The Digital universe in 2020: Big data, bigger digital shadows and biggest growth in Far East", IDC-EMC Corporation, 2012.
- [5] X. Yi, F. Liu and H. Jin, "Building a network highway for big data: architecture and challenge", IEEE Network, 2014, 28(4), pp.5-13.
- [6] C.L.P Chen and C.Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on big data", Information Sciences, 2014, 275, pp. 314-347.
- [7] Uthayasankar Sivarajah, Muhammad Mustafa Kamal, Zahir Irani and vishany Weerakkody, "Critical analysis of Big data challenges and Analytical methods", Journal of Business research, 2017, vol. 70, pp. 236-286.
- [8] A.Gandomi and M. Haider, "Beyond the hype: Big Data concepts, Methods and Analytics", International journal of Business management, 2015, 35(2), pp. 137-144.
- [9] P. Baranghi, Sheth and C. Henson, "From data to actionable knowledge: Big Data Challenges in Web of things", IEEE Intelligent System, 2013, 26(^) pp. 6-11.
- [10] A.Labrinidis and H. V. Jagadish, "Challenges and opportunities with Big Data", Proceedings of the VLDB Endowment, 2012, 5(12), pp. 2013-2033. [11].
- [11] R. Akerkar, "Big data Computing", CRC Press, Taylor & Francis Group, Florida, USA, 2014.
- [12] X. Zhang Y.Hu, K. Xie and W. Sarle, "Massively parallel feature selection: an approach based on variance preservation" Machine learning, 2013, 92(10), pp. 195-220.
- [13] S. Kaisler, F. Armour, J.A Espinosa and W. Money, "Big data, issues and challenges moving forward", 46th Hawaii International Conference on system sciences (HICSS), 2013, pp. 995-1004.
- [14] J. Chen Y. Chen, X. Du, C. Li, J. Lu, S. Zhou and X. Zhou, "Big Data challenges: A Data management Perspective", Frontiers of Computer Science, 2013, 7(2), pp. 157-164.
- [15] N. Karacapilidis, m. Tzagarakis and S. Christodoulou, "On a meaningful exploitation of machine and human reasoning to tackle data intensive decision making", Intelligent Decision Technologies, 2013, 7(3), pp. 225-236.
- [16] E. Barbierato, M. Gribaudo and M. Lacono, "Performance evaluation of NoSQL Big Data Applications using multi-formalism models", Future generation Computer Systems, 2014, 37, pp. 345-353.
- [17] H. Zhu Z. Xu and Y. Huang, "Research on security technology of big data information", International conference on Information technology and Management Innovation, 2015, pp. 1041-1044.
- [18] Alex Bakker, "Buried under Big Data: Security issues, Challenges, Concerns", Science Soft, 2018 (April 4th).
- [19] E. Davenport and H. Hall, "Organization Knowledge and Communities of Practice", annual reviews of Information science and Technology, 2002, 36, pp. 171-222.
- [20] Z. Irani, "Investment evaluation within Project management: An Information System perspective" Journal of the Operational Research society, 2010, 61(6), pp. 917-928.
- [21] I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani and S. U.Khan, "The rise of Big data on Cloud Computing: Review and Open Research issues", Information Systems, 2015, 47, PP. 98-115.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)