An Investigation on Analysis of Big Data Challenges and Analytical Methods

Sarala R, Mca, M.Phil., (Ph.D.),
Assistant Professor Dept of computer science
KG College of Arts and Science, Coimbatore, Tamil Nadu, India

Abstract- *Big Data, with their latent to establish valued insights for improved decision-making process, have newly attracted significant interest from academics and practitioners. Big Data Analytics is more and more becoming a trending practice that many organizations are adopting with the purpose of constructing precious information from Big Data. A tools, is seen by organizations as a tool to improve equipped efficiency though it has planned potential, drive new returns streams and increase competitive advantages over business rivals. Though, there are various types of analytic applications to consider. Thus, prior to rapid use and buying costly Big Data tools, there is a need for organizations to first understand the Big Data Analytics landscape.*

This paper presents a up to date review that presents a holistic view of the Big Data challenges and Big Data Analytics methods. In doing so, methodically analyzing and synthesizing the existing research published on Big Data and Big Data area. A research area, evaluating contributions, summarizing knowledge, thereby identifying restrictions, implications and possible further research avenues to support the academic community in exploring research patterns. Thus, to hint the execution of Big Data strategies, a profiling method is in use to analyze articles extracted from the database. The analysis presented in this paper has acknowledged relevant Big Data research studies that have contributed both theoretically and empirically to the development and enhance of academic wealth to the Big Data.

Keywords- Big Data, Big Data Analytics, Challenges

I. INTRODUCTION

The degree of data generated and shared by businesses, public administrations numerous industrial and not-to-profit sectors, and scientific research, has enlarged infinitely. These data include textual content (i.e. structured, semi-structured as well as unstructured), to multimedia content on a multiplicity of platforms. For example machine-to-machine communications, social media sites, sensors networks, cyber-physical systems, and Internet of Things report that every day the world produces around 2.5 quintillion bytes of data, with 90% of these data generated in the world being unstructured. Gantz and Reinsel (2012) assert that by 2020, over 40 Zettabytes (or 40 trillion gigabytes) of data will have been generated, imitated, and consumed.

With this enticing amount of intricate and various data pouring from any-where, anytime, and any device, there is indisputably an era of Big Data– a discernible fact also referred to as the Data cascade. The impending of Big Data is evident as it has been incorporated in Gartner's Top 10 premeditated technology trends for 2013 and Top 10 Critical Tech Trends for the Next Five Years. It is as essential as nanotechnology and quantum computing in the present era. In essence, Big Data is the relic of human individual as well as combined intelligence generated and communal mainly through the technological environment, where almost anything and everything can be documented, measured, and captured digitally.

In line with the endorsement concept and ever increasing technological advancements, advocates assert that in the future a mainstream of data will be generated and shared throughout machines. In spite of where Big Data is generated from and shared to, with the certainty of Big Data comes the challenge of analyzing it in a way that brings Big Value. With so much value residing inside, Big Data has been regarded as today's Digital Oil with the New Raw Material of the 21st century. Suitable data processing and management could expose new knowledge, and aid in responding to promising opportunities and challenges in a timely manner. On the other hand, the expansion of data in volumes in the digital world seems to out-speed the advance of the many extant computing infrastructures.

Established data processing technologies, for example database and data warehouse, are becoming insufficient given the amount of data the world is current generating. The huge amount of data requests to be analyzed in an iterative, as well as in a time sensitive manner. With the availability of advanced Big Data analyzing, insights can be enhanced attained to facilitate in improving business strategies and the decision-making process in critical sectors such as healthcare, economic productivity, energy futures, and predicting natural catastrophe, to name but a few.

Big Data and Big Data Analytics as a research discipline are still growing and not yet conventional, thus, a lucid understanding of the occurrence, its definition and arrangement is yet to be fully established. The existing progress made in Big Data and Big Data Analytics not only exposed a lack of management research in the field but a different lack of theoretical constructs and academic rigidity – perhaps a function of an fundamental methodological rather than academic challenge. At large, there has also been a lack of research studies that comprehensively addresses the key challenges of Big Data, or which investigate opportunities for new theories or promising practices. Thus, there exists the need to conclude the Big Data challenges and associated Big Data Analytics methods to allow signposting to take place.

This paper attempts to expand the scope of their reviews by additional investigating and assessing the different types of Big Data challenges and the analytical methods employed to triumph over the challenges. Even though these research studies afford worthy indulgent on some aspects of Big Data and Big Data Analytics area, there seems to be a lack of ample and precise approaches to understand the phenomenon of Big Data more particularly the types of Big Data Analytics methods thus an aide chronicle will act as a right frame of reference.

Moreover, explicitly in respect of the conclusions offered by these offered review articles, this research particularly aims to: analyze, synthesize and present a state-of-the-art structured analysis of the normative literature on Big Data and Big Data analytics to support the signposting of future research instructions.

This research aims to evaluate the obtainable research published on Big Data and Big Data Analytics by employing an renowned profiling approach and to scrutinize and analyze various Big Data challenges and Big Data Analytics technologies, techniques, methods and or approaches.

The concept of big is challenging to pinpoint, not least because a dataset that appears to be enormous today will almost definitely appear small in the near future. Adding to the complication of the Big Data itself, some practitioners dispute that massive datasets are not always complex and small data sets are always simple, therefore highlighting that the intricacy of a dataset is a major factor in determining whether it is big. This is potential when a selection of analytical methods is used to extort sense from the data, such as:

- descriptive analytics scrutinizes data and information to define the existing state of a business condition in a way that developments, patterns and exceptions turn into evident, in the form of producing standard reports, ad hoc reports, and alerts;
- intrusive analytics is about questioning data to certify or reject business propositions, for example, analytical drill downs into data, statistical analysis, factor analysis;
- predictive analytics is concerned with forecasting and statistical modeling to determine the future possibilities;
- prescriptive analytics is about optimization and randomized testing to evaluate how businesses improve their service levels while decreasing; and
- preventive analytics is about having the capability to take precautionary actions on events that may unfortunately influence the organizational performance, for example, identifying the possible perils and recommending mitigating strategies far ahead in time.

II. RESEARCH METHODOLOGY

In an attempt to better identify with and provide additional detailed insights to the phenomenon of Big Data and Big Data Analytics, the authors respond to the special issue call on Big Data and Analytics in Technology and Organizational Resource Management through a SLR methodology as opposed to narrative or descriptive reviews. The meta-based-approaches can be used to conducting a literature review. The scope and applicability of Big Data and Big Data Analytics obviously indicates that this area has the possible to support organizations, for instance, at the deliberate, organizational, operational as well as technological level.

III. TYPES OF BIG DATA CHALLENGES

Among the many Big Data challenges, the hefty datasets and the ability to process huge amount of data remains a significant challenge for outdated data processing applications and, relational database management systems. According to TDWI Predictive Analytic Study, there are several Big Data challenges pretense a peril to organizations – among these are, integrating intricate and large datasets, getting started with the right Big Data project, increasing and implementing infrastructure for managing and exemption Big Data and a lack of skilled personnel or staff with analytics skills to formulate sense of Big Data.

Data challenges are the group of the challenge associated to the characteristics of the data itself. Different researchers have different understandings towards the data characteristics – such as some say 3Vs [volume, velocity and variety] of data, others reported 4Vs [volume, velocity, variety, and variability] of data and 6Vs [volume, velocity, variety, veracity, variability, and value] of data.

Process challenges are the group of challenges encountered while dealing out and analyzing the data that is from capturing the data to interpreting and presenting the end results. As large datasets are frequently non-relational or unstructured, accordingly processing such semi-structured data sets at scale poses a important challenge; possibly more so than managing Big Data.

Management challenges correlated to Big Data are a group of challenges encountered, for example though accessing, managing and governing the data. Data warehouses store enormous amounts of sensitive data such as financial transactions, medical procedures, insurance claims, diagnosis codes, personal data, etc. Organizations and businesses require to guarantee that they have a strong security infrastructure that enables employees and staff of each division to only view relevant data for their department. Moreover, there must be several standard privacy laws that may preside over the use of such personal information and strict observance to these privacy regulations must be applied in the data warehouse

.

Big Data comprising of hefty raw data set on its own does not offer a lot of value in its unprocessed form. If its possible value is to be unlocked, businesses need capable processes and methods to turn high volumes of structured and unstructured data to analyze these raw datasets. Analytics in this circumstance refers to the methods used to analyze and gain intelligence from Big Data. Descriptive analytics are the simplest form of Big Data Analytics method, and involves the summarization and description of knowledge patterns using simple statistical methods, such as mean, median, mode, standard deviation, variance, and frequency measurement of specific events in Big Data streams.

Often, large volumes of chronological data is used in descriptive analytics to identify patterns and create management reports that is anxious with modeling past behavior. Watson (2014) asserts that descriptive analytics, such as reporting, dashboards, scorecards, and data visualization, have been extensively used for some time, and are the core applications of traditional business intelligence.
Another form of descriptive analysis, pointed out by Banerjee, Bandyopadhyay, and Acharya (2013) is the use of dashboard sort of application when a business routinely generates different metrics including data to monitor a process or multiple processes across times.

The predictive analytics is anxious with forecasting and statistical modeling to determine the future possibilities based on supervised, unsupervised, and semi-supervised learning models. Gandomi and Haider (2015) asserts the need to extend new solutions for predictive analytics for structured Big Data. Predictive analytics are predominantly based on statistical methods and seeks to uncover patterns and confine relationships in data. Gandomi and Haider (2015) categorized predictive analysis into two groups – regression techniques and machine learning.

The prescriptive analytics is performed to establish the cause-effect relationship among analytic results and business process optimization policies. Thus, for prescriptive analytics, organizations optimize their business process models based on the feedback provided by predictive analytic models. Although thorny to deploy, prescriptive analytics contribute to managing the information shift and the continuous evolution of business process models.

IV. CONCLUSION

This paper presented a view of Big Data practices and application of Big Data Analytics methods as offered in a normative slice of literature. Based on the findings from existing research studies, the presented research has required to analyze, synthesize and present a complete structured analysis on Big Data and Big Data Analytics to support the signposting of future research directions.

REFERENCES

[1] Agarwal, R., & Dhar, V. (2014). Editorial – big data, data science, and analytics: the oppor- tunity and challenge for is research. Information Systems Research, 25(3), 443–448.
[2] Gantz, J., & Reinsel, D. (2012). The Digital Universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC – EMC Corporation.
[3] Savitz, E. (2012a). Gartner: Top 10 strategic technology trends for 2013. Online Available at http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-rends-for-2013/ (Accessed on 3rd March 2016).
[4] Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. IEEE Network, 28(4), 5–13.
[5] Jukić, N., Sharma, A., Nestorov, S., & Jukić, B. (2015). Augmenting data warehouses with Big Data. Information Systems Management, 32(3), 200–209.

[6] George, G., Haas, M. R., & Pentland, A. (2014). Big Data and management. Academy of Management Journal, 57(2), 321–326.

[7] MIT Technology Review (2013). The Big Data Conundrum: How to define it? Available Online at https://www.technologyreview.com /s/519851/the-big-data-conundrum-how-to-define-it/ (Accessed 19th May 2016).

[8] Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. IT Professional, 15(6), 43–48.

[9] Bihani, P., & Patil, S. T. (2014). A comparative study of data analysis techniques. International Journal of Emerging Trends & Technology in Computer Science, 3(2), 95–101.

[10] Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2), 77–84.

[11] Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. IT Professional, 15(6), 43–48.

[12] Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media. 6th IEEE international conference on digital ecosystems technologies (DEST) (pp. 1–6).

[13] Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence informed management knowledge by means of systematic review. British Journal of Management, 14(3), 207–222.

[14] Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic review process research in software engineering. Online Available at http://www.citeulike. org/group/14013/article/7874938 (Accessed on 19th December 2015).

[15] Jiang, H., Chen, Y., Qiao, Z., Weng, T. H., & Li, K. C. (2015). Scaling up MapReduce-based big data processing on multi-GPU systems. Cluster Computing, 18(1), 369–383.

[16] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. 46th Hawaii International Conference on System Sciences (HICSS) (pp. 995–1004).

[17] Uthayasankar Sivarajah , Muhammad Mustafa Kamal, Zahir Irani, Vishanth Weerakkody (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research 70 (2017), 263-286.

[18] Assunçao, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: trends and future directions. Journal of Parallel and Distributed Computing, 79, 3–15.

[19] Watson, H. J. (2014). Tutorial: big data analytics: Concepts, technologies, and applications. Communications of the Association for Information Systems, 34(1), 1247–1268.

[20] Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: hyped up aspirations or true potential. Vikalpa. The Journal for Decision Makers, 38(4), 1–11.

[21] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144.

[22] Rehman, M. H., Chang, V., Batool, A., & Teh, Y. W. (2016). Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management.

[23] Sasirega.D, (2017). Comparison of Big Data and Data Mining. International Journal of Computer Science and Mobile Computing, Issue 11.