

Big Data Management in Education Sector: an Overview

Ramatu Muhammad Nda¹, Rosmaini Bin Tasmin¹

¹ *Universiti Tun Hussein Onn Malaysia*

Parit Raja, Batu Pahat, Johor, 86400, Malaysia

DOI: [10.22178/pos.47-6](https://doi.org/10.22178/pos.47-6)

LCC Subject Category: [LB5-3640](#)

Received 20.04.2019

Accepted 27.05.2019

Published online 30.06.2019

Corresponding Author:

Ramatu Muhammad Nda

ramatunda@gmail.com

© 2019 The Authors. This article is licensed under a [Creative Commons Attribution 4.0 License](#)



Abstract. The advancement in technological innovation has given rise to a new trend known as Big Data today. Given the soaring popularity of big data technology, organisations are profoundly attracted to and interested in it to transform their organisation by improving their businesses. Big data is enabling organisations to outpace their competitors and save cost. Similarly, the application of Big Data management in Universities is an essential aspect to institutions that have Big Data to manage; as the use of Big Data in the higher education sector is increasing day by day. Many studies have been carried out on big data and analytics with little interest in its management. Big Data management is a reality that represents a set of challenges involving Big Data modeling, storage, and retrieval, analysis, and visualization for several areas in organizations. This paper introduces and contributes to the conceptual and theoretical understanding of Big Data management within higher education as it outlines its relevance to higher education institutions. It describes the opportunities this growing research area brings to higher education as well as major challenges associated with it.

Keywords: Big Data; Big Data Management; Big Data benefits; Higher Education.

INTRODUCTION

In recent years, a flood of data is created every day by the interactions of people using devices such as smart/cell phones, computers, Global Positioning System (GPS) devices, and medical equipment. According to a statement from International Business Machine (IBM), 2.5 quintillion bytes of data are mined every day, and 90 percent of the data in the world today were generated within the 21st century [1, 2]. This flood of generated data emerged the confinement of big data. The era of big data is fast-evolving and is likely to be subject to improvements and modifications in the future. For example, the author [3] states that companies like Amazon and Google are experts at analyzing big data. Authors [4] term big data as the “new Oil” that can be deployed, managed and used like never due to its high-performance tools such as MapReduce, NoSQL, and Hive among others. Therefore, making big data the fuel for the current revolution. The capacity for big data is predicted to influence all sectors, from ICT firm to healthcare, from the public to private sector, from media to Telecom and entertainment, from energy to retail, etc. [5]. Researchers, academicians, and policymakers are

beginning to realize the potential for channeling these surges of data into actionable information that can be used to identify needs, predict and prevent crises, and provide services for the advantage of organizations [6]. Big data can be defined as datasets characterized by such a high-volume, high-velocity and high-variety that cannot be collected, processed and analyzed by conventional techniques. Subsequently, big data requires specific technology and analytical approaches for its revolution into value. Big data cannot be mentioned without its hallmark characteristics; these include- volume, velocity, variety, veracity and value-known as the five (5) V's [7]. The higher education sector has started using the technology of big data to effectively manage data and extract insights which are now seen as a critical competitive advantage [8]. Big data is becoming more of an essential part in the way Higher Education Institutions are leveraging volume data at the right speed to solve data problems [9]. Thus, to be effective, universities of higher education will require to be able to correlate the outcomes of big data analysis with the existed data within the institution, as big data is all about high velocity, high volume, and high data variety, to get big value [10]. Also, the uni-

versities can use the big data resulting knowledge to gain a competitive advantage over others. Consequently, many institutions have made known that analytics can help meaningfully advance their universities in strategic fields as proper means of allocation & utilization, student success, and finance is taken into consideration [11]. With the popularity and use of the Internet, the operation of a higher education institution is increasingly complex and competitive [12]. Therefore, making the institution of higher education under pressure to respond to global economic, social change and political challenge, especially in the areas of student admission, student retention, and performance [13] leading institutions to embrace big data to gain insight using tools and techniques of big data.

RESULTS AND DISCUSSION

Big Data Definition. According to the report by R. Rossi & K. Hirama [14], big data has been in use since the completion of the 1880 Data Census in the United States. Without technology or advanced techniques for data collection and organization then, the massive data took seven years to mine before finally show outcomes. Authors [15] reason that 'big' is not precisely limited to the volume, but also refers to theme directing variety, velocity, veracity, and value which make up the big data 5V's characteristics (volume, variety, velocity, veracity, and value) as illustrated in Figure 1. Likewise, in computing, the name has been utilized, while researchers in further areas are producing results since the year 2000. In the opinion of [8] in a survey using the keywords 'business intelligence', 'business analytics' and 'big data,' the evolution of the big data is relatively new as in 2001 it was found that only one research referencing the term big data, and in 2011, 95 were found using the exact phrase big data. This shows the emerging of big data research is still scanty. Researchers [6] reported that the history of big data can be generally divided into stages as megabyte to gigabyte (the 1970s and 1980s), that is, the historically business data introduced the initial term of big data; gigabyte to terabyte (late 1980s), that is, the spread of digital technology instigated the expansion of data volumes to several gigabytes or even a terabyte, which was a challenge due to insufficient storage and/or processing capabilities of a single large computer system to analyze the data. Therefore, distributed data system was projected

to extend storage competences; terabyte to petabyte: (the late 1990s), the development of Web 1.0 was quickly introduced, and this directed the world into the Internet era, along with the coming of massive semi-structured or unstructured webpages releasing terabytes or petabytes of data. Finally, petabyte to exabyte which is under current trends, data stored and analyzed by big companies will certainly reach the size of exabyte soon.

However, existing technology still handles terabyte to petabyte data. On the other hand, authors [16] present the historical term of big data way back to the 1970s. The word 'big' at that time, referred to megabytes, and 'big' over time derived to mean gigabytes, growing to terabytes. At present, the authors indicate that this word correlated to the term big data which refers to petabytes and exabytes of data. Similarly, R. Rossi & K. Hirama [14] reported that the word big data can be viewed as a large volume of data both in an individualized concept and corporate or large organization concept; or a large volume of data at a given surveyed time. Hence, it can be concluded that big data has been around for a while now only that the techniques of processing, analyzing and managing them has advanced.

The growth and increase of data generated globally in recent times have given birth to the phrase "big data." Big data is a considerable boost to emerging technology today as it is the heart of nearly every technological and digital transformation currently taken place. Authors [17] consider big data to a concept that the size of data that is complex and cannot be handled, processed and analyzed conventionally, but require much more robust technologies, systems and people with the skills for managing these huge datasets. Researchers [5] defined big data as datasets whose size goes beyond typical data and therefore cannot be treated, processed and analyzed by conventional database but instead, databases that can be created, stored, managed and analyzed by existing tools. Also, they consider the necessity of producing technologies for managing big data. The UNGlobal Pulse [18] identified the term Big Data as an umbrella for the eruption in the amount and multiplicity of high-frequency digital data which hold the potential—yet mostly untapped—to allow decision-makers to track expansion development, improve, and comprehend where present policies and programs require modification and change. Also, H. Bhosale & D. Gadekar [19] termed Big Data as datasets

whose size (volume), complexity (variety), and rate of growth (velocity) make them complicated to be captured, managed, processed and analyzed by conventional technologies and devices, within the time frame required to make them useful. Authors [9] defined Big Data an emerging

term which explains the immense volume of structured, semi-structured and unstructured data that has the potential to be analyzed for information with key characteristics known as the 5V's as shown in Figure 1.

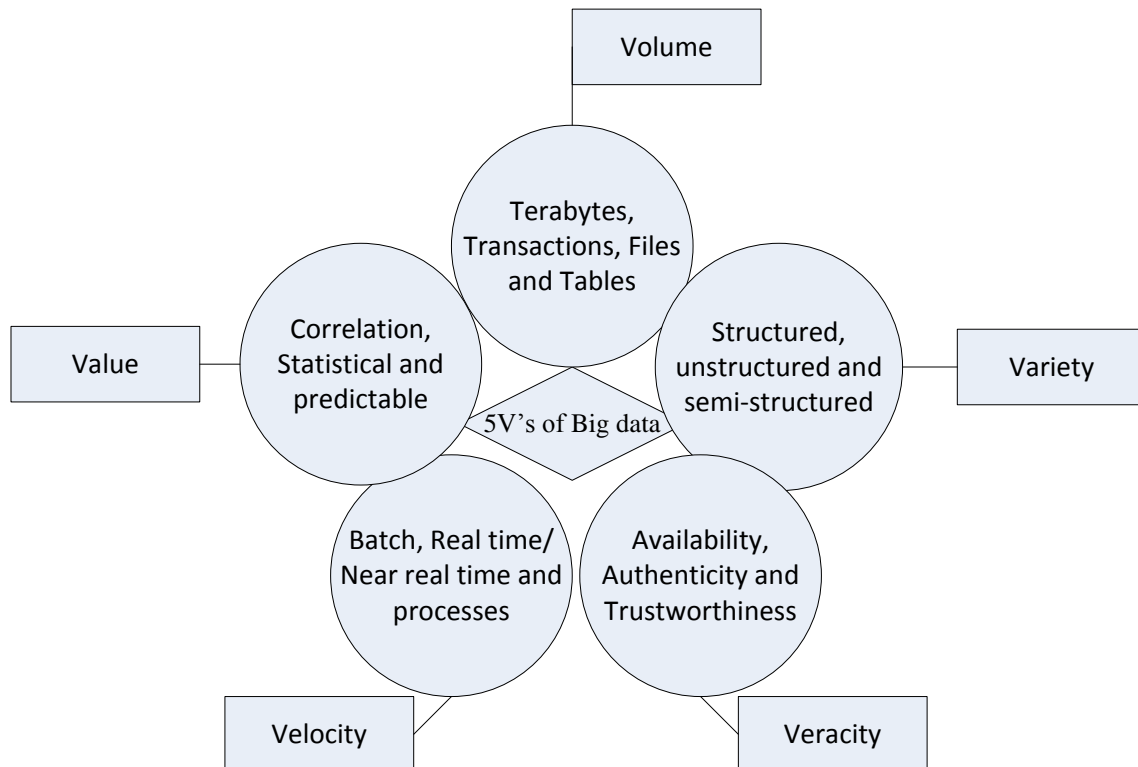


Figure 1 – The characteristics of Big Data (5V's)

These characteristics include:

1. Volume: referring to the tremendous amount of data which is often challenging to store, process, analyze and present.
2. Variety: referring to data in a diverse format both structured and unstructured.
3. Velocity: referring to the increasing rate at which information flows within an organization (e.g., streaming- real-time and near real-time).
4. Veracity: relating to the biases, noise, and abnormality in data. It refers to how data is being stored and implicitly mined to the problem being analyzed. It also covers the issues of trust, security, and uncertainty.
5. Value: most significantly refers to the data been utilized to generate the value of the insights and benefits within an organization.

Benefits and Goals of Big Data Management in Higher Education. Big data management involves

the ability to proficiently manage the big data based on its characteristics to a satisfactory result from the data and as well as maintaining the process. The education system does not only need big data to manage data but also need big data management to help in keeping the data, to improve the areas of enrollment, student's performance, progress and retention and institutional budget and finance [13]. According to [5] the following influences are considered as relevant to extract significant outcomes from big data management for any given organization: (a) data policies definition (b) specific technology and techniques (c) talents and organizational change (d) data access and (e) infrastructure. Big data have the tendency that can encourage rapid operation, storage and retrieval of distributed different, virtually raw data. G. Siemens & Ph. Long [20] denoted that Big Data represents the most dramatic context today, as it brings opportunities as well as challenges to higher institutions and it

will efficiently exploit the vast array of data that would ultimately shape the future of higher education. In other words, big data management means to clean data to be reliable, and to analyze data of different format coming from various sources collectively and to encrypt data for security and privacy goal. Likewise, authors [21] described big data management as a process that ensures adequate access to multiple distributed systems and means to store big data. E. Brynjolfsson & A. McAfee [22] stated that big data management is responsible for striving to collect intelligence from information and interpreting that information into business gain. Hence, the goal of big data management is to certify data reliability to make the data accessible, adequately stored, manageable, and secured without difficulties. Higher Education Universities have extensive data which can be managed and used for improving teaching and learning for teachers and students [23]. Additionally, it could also be used for better decision making by the executive in the management level to enhance the quality of education [9]. Big data management is an essential asset as it efficiently manages Big Data and enables the mining of reliable insight as well as save cost. Therefore, big data management is justifiably worth executing.

Challenges of Big Data Management in Higher Education. The capability of institutions of higher education to manage big data (high volume of data with the extensive variety of data types) should be considered. So as to provide rapid responses of reality for the management that must handle the challenges of big data management. Institutions of higher education have access to the vast volume of data in a variety of format such as learning management system, registration data, student information system data, assessment/performance data, quality assurance survey data, graduate data, employees' data, complaint/appeal data and social media communication data, etc. Thus, this makes it difficult to collect and integrate data from distributed locations with scalability [6]. B. Daniel [24] observed that the lack of expertise in the field of big data is a global challenge, as there is still a division among experts as to those who know what data are available and how to extract significant data, and those who know which data are essential and how it would best be used. Ph. Russom [25] observes that there are some primary difficulties

for managing big data which include: (a) lack of adequate skills from technological fields of big data (b) insufficient infrastructures for data management and (c) treatment of undeveloped types of data from different sources (semi-structured or unstructured data). Authors [26] stated that, while some early accomplished successes have already been recorded from the use of big data management, there are also some critical challenges regarding big data management. For example challenges such as (1) to cope data cleaning with short "time to benefit" (in spite of the volume of data in universities of higher education and its production rate); (2) to choose the right data analytics for a given dataset in view of data characteristics, the type of manipulation and processing processes employed to the data, and the performance of the functions given the option of connecting simple or complex indexing formations to the collections of data.

Furthermore, according to D. Laney [3] surveys that, there are some top challenges in managing big data. These include: (1) determining how to get value from big data; (2) defining strategy; (3) obtaining skills and capabilities (4) integrating multiple data sources; (5) risk and governance issues (i.e. security, privacy and quality of data); and (6) top management leadership issues. Therefore, there is need to train those in charge of data management or employing and collaborating with professionals in the big data field.

CONCLUSION

This paper has explored the need for and relevance of big data management in the higher education sector. It has also presented opportunities and challenges of big data management to enable the higher education systems policymakers and appropriate expert to make informed preferences when considering to adopt and implement Big Data in their institutions. Big data management is an essential approach that can help higher education systems reduce difficulties related to analyzing data. Therefore, it could be argued that big data management has the potential to turn information into usable data that can help universities of education sector in decision makings, which would add value to educational outcomes.

REFERENCES

1. Worster, A., Weirich, T. R., & Andera, F. (2014). Big Data: Gaining a Competitive Edge. *Journal of Corporate Accounting & Finance*, 25(5), 35–39. doi: 10.1002/jcaf.21970
2. Wardman, D. (2014). *Bringing Big Data to the Enterprise – Gaining new insight with Big Data capabilities*. Retrieved from ftp://ftp.software.ibm.com/software/os/systemz/pdf/09_-_Dan_Wardman_-_Bring_Big_Data_to_the_Enterprise_.pdf
3. Laney, D. (2012). *Big Data Means Big Business*. Retrieved from <http://media.ft.com/cms/4b9c7960-2ba1-11e3-bfe2-00144feab7de.pdf>
4. Cavanillas, J. M., Curry, E., & Wahlster, W. (Eds.). (2016). *New Horizons for a Data-Driven Economy*. doi: 10.1007/978-3-319-21569-3
5. Manyika, J., Chui, M., Brown B., Bughin, J., Dobbs, R., Roxburgh, Ch., & Byers, A. (2011, May). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf
6. Shekhar, H., Sharma, M. (2014). *A Framework for Big Data Analytics as a Scalable Systems*. Retrieved from <https://www.ijana.in/Special%20Issue/C14.pdf>
7. Sagioglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*. doi: 10.1109/cts.2013.6567202
8. Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
9. Deshmukh, D., & More, A. (2017). *Applying Big Data in Higher Education*. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2), 1302–1309.
10. Hurwitz, J., Nugent, A., Halper, F., Kaufman, M. (2013). *Big Data for Dummies*. Retrieved from <https://eecs.wsu.edu/~yinghui/mat/courses/fall%202015/resources/Big%20data%20for%20Dummies.pdf>
11. Dede, C., Ho, A. & Mitros, P. (2016). Big Data Analysis in Higher Education: Promises and Pitfalls. *EDUCAUSE review*, 9, 22–34.
12. Wielki, J. (2013). Implementation of the Big Data concept in organizations – Possibilities, impediments and challenges. In *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems* (pp. 985–989). Retrieved from https://www.researchgate.net/publication/257341462_Implementation_of_the_Big_Data_concept_in_organizations_-_Possibilities_impediments_and_challenges
13. Patel, M. R., & Desai, T. (2016). Big Data Analytics in Optimizing the Quality of Education: Challenges. *International Journal for Innovative Research in Science & Technology*, 3(6), 165–167.
14. Rossi, R., & Hiram, K. (2015). Characterizing Big Data Management. *Proceedings of the 2015 InSITE Conference*. doi: 10.28945/2192
15. Demchenko, Y., Grosso, P., de Laat, C., & Membrey, P. (2013). Addressing big data issues in Scientific Data Infrastructure. *2013 International Conference on Collaboration Technologies and Systems (CTS)*. doi: 10.1109/cts.2013.6567203
16. Borkar, V., Carey, M. J., & Li, C. (2012). *Inside “Big Data Management”: Ogres, Onions, or Parfaits?* Retrieved from <https://www.ics.uci.edu/~chenli/pub/edbt2012-asterix.pdf>
17. Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *Interactions*, 19, 50–59.
18. Global Pulse. (2012, May). *Big Data for Development: Challenges & Opportunities*. Retrieved from <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobaIPulseJune2012.pdf>

19. Bhosale, H. S., & Gadekar, D. P. (2014). *A Review Paper on Big Data and Hadoop*. *International Journal of Scientific and Research Publications*, 4, 1–7.
20. Siemens, G., & Long, Ph. (2011). *Penetrating the Fog: Analytics in Learning and Education*. *EDUCAUSE Review*, 46, 30–40.
21. Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4), 431–448. doi: [10.1016/j.jksuci.2017.06.001](https://doi.org/10.1016/j.jksuci.2017.06.001)
22. McAfee, A., & Brynjolfsson, E. (2012, October). *Big Data: The Management Revolution*. Retrieved from <https://hbr.org/2012/10/big-data-the-management-revolution>
23. Murumba, J., & Micheni, E. (2017). Big Data Analytics in Higher Education: A Review. *The International Journal of Engineering and Science*, 06(06), 14–21. doi: [10.9790/1813-0606021421](https://doi.org/10.9790/1813-0606021421)
24. Daniel, B. (2014). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920. doi: [10.1111/bjet.12230](https://doi.org/10.1111/bjet.12230)
25. Russom, Ph. (2013). *Managing Big Data*. Retrieved from <http://www.datascienceassn.org/sites/default/files/Managing%20Big%20Data%202013.pdf>
26. Adiba, M., Castrejón, J. C., Espinosa-Oviedo, J. A., Vargas-Solar, G., & Zechinelli-Martini, J.-L. (2016). *Big Data Management Challenges, Approaches, Tools and their limitations*. In Yu, S., Lin, X., Mišić, J., Shen, X. (Eds.), *Networking for big data* (pp. 1–22). Boca Raton: Chapman & Hall/CRC.