Innovative Approach to Big Data Analytics Usability

R. M. Nda
Department of Technology Management, Faculty of Technology Management and Business,
Universiti Tun Hussein Onn Malaysia.
Midland School of Business and Finance, Abuja, Nigeria
ramatunda@gmail.com

R. Tasmin
Department of Technology Management, Faculty of Technology Management and Business,
Universiti Tun Hussein Onn Malaysia.
Johor, Malaysia
rosmaini@uthm.edu.my

A. Abdul Hamid
Department of Production and Operations Management, Faculty of Technology Management
and Business, Universiti Tun Hussein Onn Malaysia.
Johor, Malaysia
aziati@uthm.edu.my

Abstract

Today, we live in an on-command, on-demand with data use multiplying at a very high pace by both individuals and organizations. Due to the rise in use of scientific applications, the scale of data is on the increase and the development of certain discipline rely heavily on the analysis of data which are; structured, semi-structured and unstructured. The conventional system of data management storage and analysis lack the requisite resources for processing the data due to its heterogeneity, volume, and speed of data generated. Therefore, this article presents an overview of big data analytics by discusses its importance, applicability, and challenges.

Keywords: Big data analytics, analytic tools, big data management, big data application

1. Introduction

Currently, there are many changes and some disturbances in the IT environment, which consist of integrating current and evolving IT technologies with quicker, next-generation attributes, and functionalities. As a result, there are great scores of new possibilities and opportunities openings for businesses and as well as individuals of an excellent eruption of entirely new IT features and solutions. There is a range of revolutionary and disruptive innovation that emanates and develops as attested by some of the leading industry analysts and analyst firms such as Gartner firm which regularly lists the ten top trends in technology every year (Gartner, 2014). These innovations have intrinsic verve to bring about various intelligent and concise transitions for both business and ordinary people. Hence, it is within this concept that gave birth to big data analytics.

Today more and more information is being collected by organizations and that information (data) is increasing daily (as shown in figure 1) which consumes storage space and becomes more difficult to handle and therefore, the term big data (Jagadish, 2015). The reason differs regarding the need to record such massive amounts of data which sometimes is centered on the adherence involvement of the organizations. For instance, this adherence could be because of regulatory laws, sometimes the need to maintain transactions and mostly for contingency planning. The concept behind big data is nothing new as it has been in existence for several decades and businesses have been using software such as business intelligence, and scientists have been analyzing to discover the hidden patterns of these datasets. Nevertheless, the data collection size is changing, thus the more data you possess, the more information could be extrapolated from them (Narayan Das et al. 2018). Therefore, the goal is to find the meaning of the data and survey the data sources evolve in more practical and significant ways in order to develop new means of capable for establishing knowledge and relationship between the existing and newly generated data, solving issues relating to productivity, profit, and quality (Fan et al. 2014).

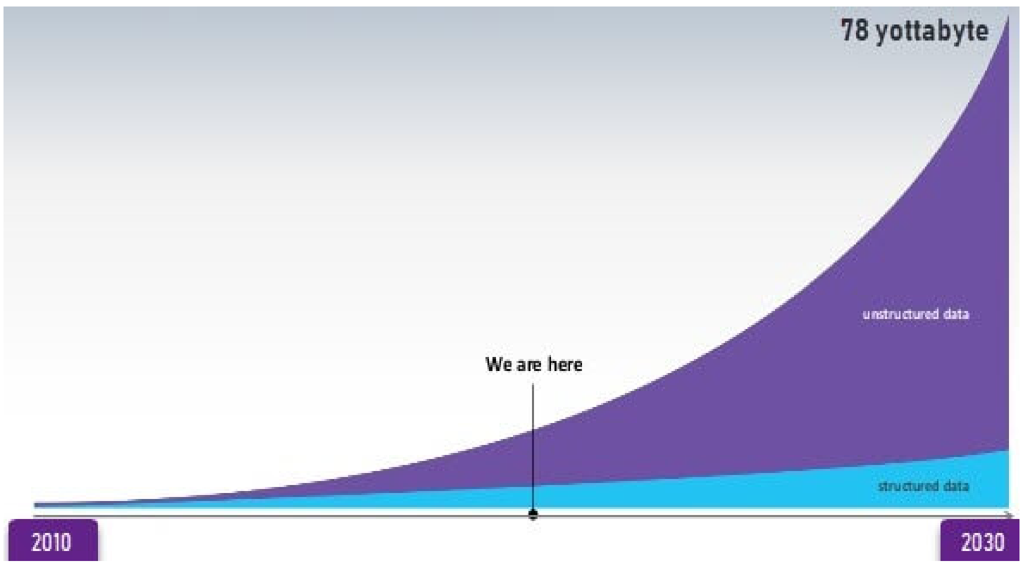

Figure 1. Evolution of data

2. Influencing evolution of big data analytics

Understanding the meaning of big data analytics and understanding its potentials are two things. This is because knowing big data analytics, its value from the information can nonetheless be challenging to visualize. Big data is defined in multidimensional characteristics including Volume- which denote the size of the data; Variety- extending to include unstructured data of all sorts outside structured data such as log files, text, video, audio, transaction history files and clickstreams among others; Velocity- the speed at which the data is generated which is sometimes sensitive, therefore, the need of big data to stream data in order to optimize its value; Veracity- misinterpretation and errors of the collected data that need to be cleansed in order to get value (Li et al. 2015). Addressing these characteristics, big data analytics have the potential to analyzed and manage the increasingly generating data using its analytics. According to Bihl et al. (2016), big data is distinguished not only by the immense of data but also by the complexity and heterogeneity of the data, and the pace of its production. This is a result of the product from new experimental techniques, quantitative executional studies, scientific mechanisms such as internet operations, extensive use of sensors and telescope and particle accelerators (Muhammad et al., 2020). Big data analytics is processed by numerous technologies. These technologies or ideas are not recent, however, have come under the umbrella of big data analytics which are categorically best described as follows:

- *Business Intelligence (BI):* its composition comprises comprehensive classification of applications and technologies, storing, analyzing, and rendering data availability. Similarly, business intelligence works by analyzing in detail the business data in-depth which is provided by data sources such as application data and databases, etc (Muhammad, 2020). Business intelligence provides actionable information, which uses fact-based support systems to help organizations make more effective decisions (Hirsimäki, 2017). Besides, business intelligence can also provide historical, current, and predictive data potentials of organizational operations for effective decision-making purposes (Ashraf, 2017).

❖ *Statistical applications:* these applications use statistical-based algorithms to analyze data which usually focus on datasets concerning surveys, censuses, and other static datasets. Preferably, the statistical applications provide sample observation which can be exploited to study datasets of a specific population for purposes such as testing, estimations, and predictive analysis (Blum et al. 2013). For instance, empirical data such as experiments, and survey data are the key sources of data for analysis in statistical applications (Fearnhead & Prangle, 2012).

❖ *Data mining:* mostly, data mining is mainly used with archival or repository data. Data mining is used in evaluating numerous viewpoints of datasets which is then transformed into summarized data and is considered as useful (Olson & Lauhoff, 2019). The techniques of data mining concentrate on modeling and discovering information that is rather predictive than strictly descriptive which indicates an excellent approach for uncovering new trends from vast amounts of data (Lynch et al. 2019).

❖ *Data modeling:* is a computational analytical application whereby multiple "what-if" setups could be applied to multiple datasets through algorithms (Moral-Benito et al. 2019). Normally in data modeling, the modeled information changes following the available information provided for the algorithm which offers value to the results on the changed datasets (Petrova et al. 2019). Modeling datasets operates together with data visualization, thereafter discovers information that can support a specific business initiative (Flanders & Jannidis, 2015).

❖ *Predictive analysis:* is a subdivision of statistical applications where datasets are assessed in order to create predictions based on patterns and database knowledge (Falkowski, 2019). Predictive analysis is mostly utilized in the scientific concepts and financial environments, where patterns continue to drive forecasts when additional external components are combined with the collected dataset. One of the significant and key objectives of predictive analysis is that it recognizes risks and identifies opportunities for corporate, industrial, and market processes (Kurosaki et al. 2010).

Therefore, the endless quest of value-driven by sustainable competitive advantage has been encouraging organizations to turn to large corporate and external data repositories to discover new patterns, statistics platforms, and other actionable applications to assist to boost and manage their data usage. Furthermore, these platforms have helped in the emergence and popularity of big data among top executives and technologists along with its related platforms, tools, and analytics. This is because organizations tend to collect and store data in a wide variety of formats which can include either or both structured, semi-structured, or/and unstructured data and tend to have numerous storage and management requirements. Additionally, the open sources have brought forth some of the tools of big data (Muhammad, 2020). These tools include:

Hadoop- the most prominent and used tool in big data perspective is Apache Hadoop; an open-source framework that runs on commodity hardware with its enormous capacity to process data on a large-scale (Chawda & Thakur, 2016). Hadoop works in parts and includes the Hadoop Distributed File System (HDFS), MapReduce, YARN, and libraries. The next big data hype tool is Apache Spark. Significantly, Apache Spark is an open-source hat that fills the processing gaps of Apache Hadoop as both batch data and real-time data can be managed (Wang et al. 2015). Equally, Apache Spark processes in-memory data which makes the processing of data much quicker than processing data using traditional disks. This is also an added advantage for analysts to process other forms of data to achieve quickest results (Bharill et al. 2016). Apache Spark is versatile while working with HDFS and other data storage, as it is quite easy and convenient to run Spark on a single local system making the development and testing easier (Wang, 2015). Like Apache Hadoop, Apache Spark facilitates Input/output (I/O) functionality, distributed task transmission, and scheduling (Bharill, 2016). Thus, Apache Spark is an excellent alternative to MapReduce of Hadoop because Spark can execute tasks 100 times more quickly than MapReduce (Yan et al. 2015).

Another known big data tool is the Apache Storm. The storm is a real-time distributed platform designed to consistently process the limitless data stream and suitable for any programming language (Salloum et al. 2016). The Storm framework has unique features which include; "fail quick, auto-reboot" approach, vast scalability, and fault tolerance(Ta et al. 2016). RapidMiner on the other hand is among the valuable big data tools that aid several machine learning phases such as preparation of data, predictive analytics, optimization, visualization, statistical model, validation, and deployment (Massaro et al. 2018). Clouds, complex event processing are relatively newer tools of big data analytics, while data visualization and predictive analytics have been around for quite some time now but are recently experiencing widespread adoption (Russom, 2011). Statistical analysis applications and hand-coded SQL are both older and have been rather established *(Slavakis* et al. 2014). However, applications such as Hortonworks, Cloudera, and MapR are selling its applications of big data which makes them easier to use and as well manage (Wang, 2015). Other tools include Cassandra, MongoDB, and R programming tools among others (Dwivedi et al. 2016). Also, the increasing yield for on-demand of big data services from cloud providers and the decision-making circles such as the executives and managers become more challenging for organizations (Bendre & Thool, 2016). Therefore, organizations must invest in exploration and carryout due diligence in choosing the appropriate framework and approach for executing business plan to be successful.

2. Applicability of big data analytics

The data flow comes from both unstructured data such as social networking messages, blogs, web pages, and the likes; and structured data such as corporate databases which makes the volume of data obtain increases exponentially and similarly the evolution of data processing technologies. For instance, there are presently limitless digital sensors in automobiles, industrial types of equipment, shipping cabinets, and electrical meters globally. These sensors are capable of measuring and communicating with the setting, movement, temperature, changes in chemical air, and vibration (Dey et al. 2018). Big establishments today consider data like a shield. Big companies like Amazon and Walmart utilize data to analyze and process the analysis regarding product price, demographic, economic, and weather data fits product choices in different stores and timing of price markdowns (Watson, 2014; Hashem et al. 2015). Equally, the comprehensively managing and maintaining of transactional data from databases and retrieving data cautiously from the warehouses, organizations collect indescribable quantities of log data from servers, forms machine-generated data, and other data sources can be managed by big data analytics (Watson, 2014).

In the context of big data analytics, data must be processable in parallel through multiple servers given the volumes of data being analyzed. According to Huang et al. (2016) noted that big data is the new transformative currency in science, commerce, education, engineering, and as well as finance. Big data have the innovative potentials and opportunities resulting from a heterogeneous dataset collected, stored, and visualized whilst taking into consideration the significant privacy and security implication of the data. big data has proved its usefulness for the business market. For example, businesses like Amazon, Facebook, and Google have come to depend on big data analytics as their key marketing structures to better serve their customers (Varia & Mathew, 2014). Amazon for one has made effective use of big data analytics to create an extremely accurate depiction of which products should a customer purchase (Chen et al. 2016). Amazon does this by saving queries of each customer acquisition and almost every other available information and thereby using algorithms to the gathered information to compare customers' information (Chen, 2016). The findings are accurate and tangible and also provide a realistic value to the client *(*Han et al. 2015). Hence, making big data analytics an essential asset to all aspects of Information Technology (IT) ecosystem discovery and innovation that could be applied to all facets such as governmental settings, health sector, academics, investment environments, industrial and entrepreneurial landscape (Oostveen, 2016). For example, big data analytics can be used by the health sector in a numerous way:

3.1 Big data analytics in the healthcare sector

In health institutions' perspective- evidentiary medicine and support for decision making are some of the challenging issues. Hence, the use of big data analytics can ease the processes of decision-making and which can likewise provide solid evidence to authenticate a specific direction for decisions (Groves et al. 2016). And for instance, this can be achieved by observing and predict that when a patient suffers from chronic illness, there is a strong likelihood that anything can happen to the patient due to his/her background history (Sukumar et al. 2015). The possible progression or result can then be brought up at the outset of the treatment process and the physician in charge should be immediately notified and information like this comes from processes of big data analytics (Kalid et al. 2018).

From the perspective of the patient, big data analytics can help the patient to decide which nearest hospital is best regarding the treatment of his/her illness condition. There are several options available today for patients to choose such as doctor's recommendations and requirements of insurance among other aspects (Groves, 2016). Therefore, by using big data analytics reports can be derived through analytics as the technique can produce these reports so as to unlock all the detailed information and enforce regulations and reports (Sukumar, 2015). However, to attain this, different IT infrastructures are required to be put in place. These technologies include business intelligence, analytics and dashboard, clinical intelligence, and institutional revenue-cycle management intelligence to extract unstructured data (Kaul et al. 2015). Unfortunately, the health system restricts information from a patient's viewpoint (Dwivedi et al. 2019). Most patients have no insight about what exactly is going about, not until at least the doctors' arrival. Similarly, most patients are uneasy talking to doctors (Kruse et al. 2016). This is an intellectual barrier for both the patients and practitioners which creates a condition in which it is more complicated for both patients and doctors to make choices. Also, big data have the ability and potentials to resolve the difficulty as big data can breakdown the flow of information making it easier for physicians to handle and for patients to access (Mehta & Pandit, 2018).

In a cost and a quality-of-care perspective- there are a variety of areas that big data can help to improve. Big data have the ability which allows a proactive strategy plan for treatment that reduces injuries or other challenges that can hinder the quality of the healthcare system (Sukumar, 2015). Thus, big data can offer substantial savings by avoiding these sorts of incidents and issues. The patient-physician relationship perspective; with the use of smartphone apps (social media), big data technologies make it easier to identify health problems and also improve communication and contact between physicians and patients. According to Groves, (2016), noted that to detect health pattern, unstructured, and stored data can be analyzed and processed against social media data. health systems can then make use of these data to help keep the patients safe. The transformation emphasis of the healthcare system is to improve the safety of patients by living healthier and at the same time reducing the cost (Mehta, 2018). Numerous studies have shown that the health care sector is becoming to see the potentials and opportunities in big data analytics which is likely to advance soon.

3.2 Big data analytics in the higher education sector

Digital transformation in higher education systems has contributed to more data collection than earlier. As a result, the academic institutions' leaders are finding ways to incorporate theoretical strategies as big data analytics is seen to be improving other organizations that use them. Although, unlike the health system, the education sector is still at the preliminary stage of big data usage. Like every sector, the education system is faced with challenges in which the technology of big data can improve. These challenges include rising pressure to address economic political social changes, cultural issues, and also the cut in financial allocation to institutions (Charan & Murty, 2018). Therefore, there is a need to tackle these problems in which big data analytics can be applied to resolve. Equally, with the increased use of digital platforms, higher education institutions collect large amounts of data and this has generated an avenue for big data analytics in academic institutions (Dede et al. 2016). Big data analytics can be applied in educational institutions in various ways including:

Teaching- Learning perspective; digital learning is increasingly progressing in higher education systems that can lead to vast analytical data collection (Picciano, 2014). This can contribute to both teachings and learning with the use of big data analytics. For example, fine-grained data can be collected by digital learning due to the capacity of big data analytics and in turn, improving students' performances and learning courses can be obtained in a significantly detailed approach. Also, this can support teachers to provide guided instructions which will have a significant positive academic impact on students and other universities can replicate (Sivarajah et al. 2017). Similarly, the use of modern computers and smartphones can provide new methods to collect, process, and analyze data and to aid interactive teaching-learning such as intellectual tutoring approaches, simulation, and games in order to find trends and patterns to enhance teaching-learning orientation (Thilagaraj & Sengottaiyan, 2017).

Strategic management perspective; university systems work in a complex and competitive environment and executives need to be at their top game to sustain competitive advantage. Big data mechanism can support and aid the academic executives and strategic managers in making effective decision-making with the use of a data-driven approach for analyzing institutional data, and this can be achieved using the predictive modeling (Harmokivi-Saloranta & Parjanen, 2018). Therefore, for institutional executives, big data analytics can be used in numerous institutional concepts such as financial planning and allocations, recruitment and processes of student registration, monitoring of student performance, and prediction to recognize potential difficulties in time for immediate interventions (Nalchigar & Yu, 2018). In general, According to Williamson, (2018), higher education institutions can profit from big data analytics in areas such as a) creating a culture of achievements, b) Reform of education establishment, c) Optimization of procedures, processes and non-core facilities, and d) Restructure of key support amenities such as human resources, academic resources and fund to strategically create valuable information. Therefore, big data analytics can be used in almost all aspects of life.

4. Challenges of big data analytics

Fundamentally, recorded data is generated from data-generating sources, and collecting data is an analog to identifying and analyzing the world. However, most of the collected data have little value which can be cleaned and constricted by many scales and this creates more problems; that is, identifying filters that do not dispense valuable data. thus, these criteria and other factors need to be discussed before the data filtration (Najafabadi et al. 2015). This may lead to needing new methodologies and techniques to logically processed primal data and distribute the data in a manageable portion to get value from the data. Additionally, filtering problems arise from in-house data processing, where data are streaming, and one does not have the option to first save data and then to minimize them (Sivarajah, 2017). Mengke et al. (2016) stated that the automated development of the correct metadata to explain what data is documented and analyzed is another challenge of big data analytics. For instance, a significant amount of information might be needed in scientific experiments concerning particular procedures and experimental settings for appropriate interpretation of the outcomes (Mengke, 2016). Thus, such metadata must be collected using observer data. However, when properly implemented, metadata acquisition systems will reduce the need for manual processing and thereby substantially reducing the human duty of metadata storage (Boubiche et al. 2018). Another challenge of big data is that, it is much more difficult for data analysis to locate, identify, understand, and quote data (Anagnostopoulos et al. 2016). All of these are required to be done entirely automatically for efficient large-scale analysis. This includes variations in the structure and semantic to be presented to both machine-readable and computer-resolvable ways (Anagnostopoulos, 2016). And a lot of work will be needed to achieve an automated error-free resolution.

Therefore, the big data analysis benefit can only be understood if it can be vigorously implemented under certain daunting conditions. The information built from this data can, however, be used to correct errors and eliminate uncertainties. The next wave of interactive analysis of data can be seen to provide real-time solutions using automation (Han, 2015). This signifies that machine intelligence can be utilized in the future to direct queries generated automatically towards analytics of big data- a core capacity to expand the value of data for automated websites content production, to fill in recommendations and as well as providing an ad-hoc review of the value of dataset to determine either to retain or discard it (Sivarajah, 2017). This is because advances are on-going daily so what is difficult today will be resolved soon as computing capacity rises and data becomes more coherent (Mengke, 2016). In addition, analysts are hindered by a tedious process of disseminating data from the database, executing non-SQL procedures and retrieving the data (Najafabadi, 2015). Achieving these goals will require learning how to better obtain, store, and analyze attribution in combination with the use of techniques to collect sufficient metadata. Also, to build an infrastructure that will allow analysts to interpret the analytical findings and replicate the analysis with different criteria, assumptions, or datasets. Ultimately, realizing the enormous potentials of big data analytics requires a long-time, systematic strategy, ambitious and sustainability not only for IT practitioners but also for government and research institutions across the globe.

5. Conclusion

Critical effects on the economy include observations and forecasts from large and complex data collection as access to information is turning conventional businesses which generates opportunities into new markets. Big data facilitates the development of innovative IT and data analytics related products and services and improves the efficiency of businesses to make strategic decisions and recognizes emerging market trends. Big data progress is critical to accelerate the pace of discovery in almost every discipline and which can create new strategies for effective data-driven discovery. Therefore, big data analytics has the potential to solve several of the world pressing challenges with huge societal gain in areas such as healthcare sectors, commerce, environment, education, medicine, cyber and national security, and businesses. Thereby, laying a foundation for sustainable competitiveness for decades to come.

Acknowledgments

- The authors highly appreciate Universiti Tun Hussein Onn Malaysia (UTHM) for funding this research work under the Fundamental Research Grant Scheme FRGS vot K20

References

Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: research challenges and future directions. *The Journal of Supercomputing*, *72*, 1494–1516.
Ashraf, T. (2017). *Organizational development and big data: Factors that impact successful big data implementations*. Benedictine University.
Bendre, M. R., & Thool, V. R. (2016). Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics*, *3*, 206–239.
Bharill, N., Tiwari, A., & Malviya, A. (2016). Fuzzy based scalable clustering algorithms for handling big data using apache spark. *IEEE Transactions on Big Data*, *2*, 339–352.
Bihl, T. J., Young II, W. A., & Weckman, G. R. (2016). Defining, understanding, and addressing big data. *International Journal of Business Analytics (IJBAN)*, *3*, 1–32.
Blum, M. G. B., Nunes, M. A., Prangle, D., & Sisson, S. A. (2013). A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science*, *28*, 189–208.
Boubiche, S., Boubiche, D. E., Bilami, A., & Toral-Cruz, H. (2018). Big data challenges and data aggregation strategies in wireless sensor networks. *IEEE Access*, *6*, 20558–20571.
Charan, P., & Murty, L. S. (2018). Institutional pressure and the implementation of corporate environment practices: examining the mediating role of absorptive capacity. *Journal of Knowledge Management*.
Chawda, R. K., & Thakur, G. (2016). Big data and advanced analytics tools. In *2016 symposium on colossal data analysis and networking (CDAN)* (pp. 1–8). IEEE.
Chen, H.-M., Schütz, R., Kazman, R., & Matthes, F. (2016). Amazon in the air: Innovating with big data at Lufthansa. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5096–5105). IEEE.
Dede, C. J., Ho, A. D., & Mitros, P. (2016). Big data analysis in higher education: Promises and pitfalls. *EDUCAUSE Review*.
Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A., & Satapathy, S. C. (2018). *Internet of things and big data analytics toward next-generation intelligence*. Springer.
Dwivedi, A. D., Srivastava, G., Dhar, S., & Singh, R. (2019). A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors*, *19*, 326.
Dwivedi, S., Kasliwal, P., & Soni, S. (2016). Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)* (pp. 1–8). IEEE.
Falkowski, L. (2019). *Presidents, secretaries of state, and crises in US foreign relations: A model and predictive analysis*. Routledge.
Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*, 293–314.
Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*, 419–474.
Flanders, J., & Jannidis, F. (2015). Data modeling. *A New Companion to Digital Humanities*, 229–237.
Gartner, W. B. (2014). Notes towards a theory of entrepreneurial possibility. In *Handbook of research on small business and entrepreneurship*. Edward Elgar Publishing.
Groves, P., Kayyali, B., Knott, D., & Kuiken, S. Van. (2016). The'big data'revolution in healthcare: Accelerating value and innovation.
Han, Z., Bennis, M., Wang, D., Kwon, T., & Cui, S. (2015). Special issue on big data networking-challenges and applications. *Journal of Communications and Networks*, *17*, 545–548.
Harmokivi-Saloranta, P., & Parjanen, S. (2018). The knowledge-creating pattern in user-driven innovation. In *EAPRIL 2018 Conference Proceedings* (Vol. 5, pp. 37–51).
Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115.
Hirsimäki, R. (2017). Critical success factors for business intelligence system implementation.
Huang, Y., Zhang, Y., Youtie, J., Porter, A. L., & Wang, X. (2016). How does national scientific funding support emerging interdisciplinary research: A comparison study of big data research in the US and China. *PloS One*, *11*, e0154509.
Jagadish, H. V. (2015). Big data and science: Myths and reality. *Big Data Research*, *2*, 49–52.

Kalid, N., Zaidan, A. A., Zaidan, B. B., Salman, O. H., Hashim, M., & Muzammil, H. (2018). Based real time remote health monitoring systems: A review on patients prioritization and related" big data" using body sensors information and communication technology. *Journal of Medical Systems*, *42*, 30.

Kaul, C., Kaul, A., & Verma, S. (2015). Comparitive study on healthcare prediction systems using big data. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1–7). IEEE.

Kruse, C. S., Goswamy, R., Raval, Y. J., & Marawi, S. (2016). Challenges and opportunities of big data in health care: a systematic review. *JMIR Medical Informatics*, *4*, e38.

Kurosaki, M., Matsunaga, K., Hirayama, I., Tanaka, T., Sato, M., Yasui, Y., Tamaki, N., Hosokawa, T., Ueda, K., & Tsuchiya, K. (2010). A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatology Research*, *40*, 251–260.

Li, J., Tao, F., Cheng, Y., & Zhao, L. (2015). Big data in product lifecycle management. *The International Journal of Advanced Manufacturing Technology*, *81*, 667–684.

Lynch, C. F., Merceron, A., Desmarais, M., & Nkambou, R. (2019). Proceedings of the International Conference on Educational Data Mining (EDM)(12th, Montreal, Canada, July 2-5, 2019). *International Educational Data Mining Society*.

Massaro, A., Maritati, V., & Galiano, A. (2018). Data Mining model performance of sales predictive algorithms based on RapidMiner workflows. *International Journal of Computer Science & Information Technology (IJCSIT)*, *10*, 39–56.

Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*, *114*, 57–65.

Mengke, Y., Xiaoguang, Z., Jianqiu, Z., & Jianjian, X. (2016). Challenges and solutions of information security issues in the age of big data. *China Communications*, *13*, 193–202.

Moral-Benito, E., Allison, P., & Williams, R. (2019). Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond. *Applied Economics*, *51*, 2221–2232.

Muhammad, R. N., Tasmin, R., & Aziati, A. H. N. (2020). Sustainable Competitive Advantage of Big Data Analytics in Higher Education Sector: An Overview. In *Journal of Physics: Conference Series* (Vol. 1529, p. 42100). IOP Publishing.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*, 1.

Nalchigar, S., & Yu, E. (2018). Business-driven data analytics: a conceptual modeling framework. *Data & Knowledge Engineering*, *117*, 359–372.

Narayan Das, N., Choudhary, M., & Yadav, S. (2018). Application of Big Data Analysis in Industry. *International Journal of Advanced Studies of Scientific Research*, *3*.

Olson, D. L., & Lauhoff, G. (2019). Descriptive data mining. In *Descriptive Data Mining* (pp. 129–130). Springer.

Oostveen, M. (2016). Identifiability and the applicability of data protection to big data. *International Data Privacy Law*, *6*, 299–309.

Petrova, E., Pauwels, P., Svidt, K., & Jensen, R. L. (2019). In search of sustainable design patterns: Combining data mining and semantic data modelling on disparate building data. In *Advances in Informatics and Computing in Civil and Construction Engineering* (pp. 19–26). Springer.

Picciano, A. G. (2014). Big data and learning analytics in blended learning environments: Benefits and concerns. *IJIMAI*, *2*, 35–43.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*, 1–34.

Salloum, S., Dautov, R., Chen, X., Peng, P. X., & Huang, J. Z. (2016). Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*, *1*, 145–164.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286.

Slavakis, K., Giannakis, G. B., & Mateos, G. (2014). Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge. *IEEE Signal Processing Magazine*, *31*, 18–31.

Sukumar, S. R., Natarajan, R., & Ferrell, R. K. (2015). Quality of Big Data in health care. *International Journal of Health Care Quality Assurance*.

Ta, V.-D., Liu, C.-M., & Nkabinde, G. W. (2016). Big data stream computing in healthcare real-time analytics. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 37–42). IEEE.

Thilagaraj, T., & Sengottaiyan, N. (2017). A Review of Educational Data Mining in Higher Education System. In *Proceedings of the 2nd International Conference on Research in Intelligent and Computing in Engineering,*

RICE (pp. 349–358).
Varia, J., & Mathew, S. (2014). Overview of amazon web services. *Amazon Web Services*, 1–22.
Wang, J., Ji, Z., Shi, M., Huang, F., Zhu, C., & Zhang, D. (2015). Scenario analysis and application research on big data in smart power distribution and consumption systems. *Proceedings of the CSEE*, *35*, 1829–1836.
Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*, 65.
Williamson, B. (2018). The hidden architecture of higher education: building a big data infrastructure for the 'smarter university.' *International Journal of Educational Technology in Higher Education*, *15*, 1–26.
Yan, Y., Huang, L., & Yi, L. (2015). Is apache spark scalable to seismic data analytics and computations? In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 2036–2045). IEEE.

Biographies

Ramatu Muhammad Nda is a Doctor of Philosophy scholar at Universiti Tun Hussein Onn Malaysia. She holds a MSc (Information Systems) from National Open University Nigeria, and a Bsc. (mathematics/Computer Science) from the Federal University of Technology Minna, Niger State-Nigeria. Her career objective is to tap computer-based information as an instrument for advancing business practices, supporting ICT research and improving quality service.

Dr. Rosmaini Tasmin is currently an Associate Professor at Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia (UTHM). He earned his PhD in corporate Knowledge Management (KM) from Multimedia University (MMU) Cyberjaya, Malaysia. He pursued technology management studies at master's level in Universiti Teknologi Malaysia (UTM) for which he was awarded master's in technology management. Dr. Rosmaini, studied computer science engineering during his degree Programme and was awarded BSc. (Hons.) in Computer Science and Engineering (CSE) from the University of Texas at Arlington (UTA) in the United States of America. Dr. Rosmaini has over 10 years of industrial work experience in multinational corporations including Sony Video Malaysia, Sumitomo Electric (JB) and Thomson Mechatronics (USA group). He has served in several capacities during his stint in corporate organisations from computer-aided testing engineer, senior process engineer, assistant production planning manager and his last industrial designation as the production manager at Thomson Mechatronics. He is currently the Head of Service and Operation Management (SOM) Research Focus group. He has taught many courses and supervised many research theses at both undergraduate and postgraduate levels. He has published several research papers and conferences. His research interest areas are; train system assurance (SA), big data management, cloud computing, the Internet of Things (IoT) and mobile service-oriented computing.

Nor Aziati Abdul Hamid is an associate professor and senior lecturer at the Department of Production and Operations Management, Faculty of Business Management and Technology. Obtained a PhD in system management and a bachelor's degree in computer science. She has conducted research mainly in the field of Information Systems, Knowledge Management, Strategic Information System Planning, Industrial Data System and Operation Management. Dr. Aziati is the principal researcher of Service and Operations Management Focus Group established at the Faculty of Technology Management and Business. She is also actively involved in collaborative and consultancy projects with public agencies and industries. Areas of Interest include Information, Computer and Communications Technology (ICT).