Characteristics of big data produced by the Technical University of Kenya and Strathmore University

Lucy Wachera Kibe*, Tom Kwanya, Ashah Owano

The Technical University of Kenya

**kibelucy871@gmail.com*

Abstract

This study investigated the characteristics of big data produced by the Technical University of Kenya (a public university) and Strathmore University (a private university) in Kenya. The two universities provided contextual insights into the differences and similarities between the characteristics of big data from the perspectives of private and public universities in Kenya. The study adopted convergent parallel mixed methods research design. Quantitative and qualitative data was collected using questionnaires and key informant interviews. The target population for the study was 22,050 respondents consisting of clients (students) as well as ICT staff, directors and managers from both TUK and SU. Information-oriented purposive sampling was used to select information-rich subjects. This gave TUK a sample size of 580 and 114 for SU. Quantitative data was analysed using Statistical Package for Social Sciences (SPSS) while the qualitative data was analysed using thematic analysis. It was established that both institutions generate big data which can be described in terms of Volume, Variety and Velocity (3Vs) of big data. The volume of big data is produced in terms of Gigabytes, Terabytes, Megabytes and Kilobytes. The velocity of processing this big data was using real time, periodic, batch and near real-time approaches. The institutions had different varieties of big data ranging from email-based data, photos, video, audio, social media data, MS Office data, cell phone data, financial data, web-log data, and gaming related data. The results of the study can be used by academic institutions to leverage on the data they produce through analytics to improve their performance. This study is original in terms of its subject matter, scope and application.

Keywords: big data, big data characteristics, Technical University of Kenya,Strathmore University, Kenya.

Introduction

Organisations are daily producing unstructured and multi-structured data. This has brought about an increase in the volume and diversity of the data produced as the years go by (Savvas, 2011). This growing data has been described as big data. Whereas Jacobs (2009) asserts that big data is data whose size and density force manipulators to use advanced technologies to gain insight from it, Russom (2011) states that it is the quantity of information resources an organisation generates or acquires over time, and that the continuous accumulation of big data may lead to an information explosion. Laney (2012) views big data in terms of velocity, volume and variety, commonly known as “3Vs”. This means that these datasets are not only voluminous; they are also generated fast,and are found in diverse formats. Therefore, big data prevails in different sizes and formats. The description of big data differs from organisation to organisation subject to the existing procedural, organisational and infrastructural capacity to manage the data an organisation produces, conveys or collects. In most organisations, big data is dispersed in different locations and with diverse personnel. The existence of big data is dictated by the rising ability of people and organisations to create and share content easily on the Internet and other Information and Communication Technology (ICT) platforms. Similarly, the increase in the advancement and ubiquity of technologies such as mobile devices has enabled users to connect to the Internet in real-time leading to the production of large amounts of data.

Literature review

According to Mayer-Schoönberger and Cukier (2013), advancements in digital technologies and the ubiquity of smart devices have created big data that traditional tools cannot manage efficiently. Besides big data, the present-day society experiences infobesity[1] occasioned by increased information generation and sharing (Demirkan & Delen, 2013). The rapid development of information technologies has made the generation of large amounts of data easier, hence the information explosion

1. Infobesity is the occurrence in information overload acquired when vast amount of information is input into a system exceeding its processing capacity.

(McAfee, Brynjolfsson, Davenport, Patil & Barton, 2012). Frizzo-Barker, Chow-White, Mozafari and Ha (2016) explain that big data is believed to change many aspects of an organisation. Organisations which embrace big data tend to shift towards thinking about data and its infrastructure, big data analytics and business intelligence. Such organisations develop strategies to help them to understand big data as a technological phenomenon. As explained earlier, Laney (2012) articulates the meaning of big data in terms of the "3Vs". He also places the "Vs" in a framework seeking to determine how big data is and the technologies required to process each of the dimensions as presented in Figure 1. The levels determine the corresponding region for each characteristic of big data generated by organisations which are between zero and three. The higher the level, the broader the characteristics and technology required for analysing the data. The sum of the levels in the digital magnitude index is generated in order to identify the technology required for big data analytics. Once the sum is generated, and the value is situated between one and three, the traditional technologies can then be used for analysis. However, if the value is between four and six the advanced technologies like NoSQl are appropriate for analysis.

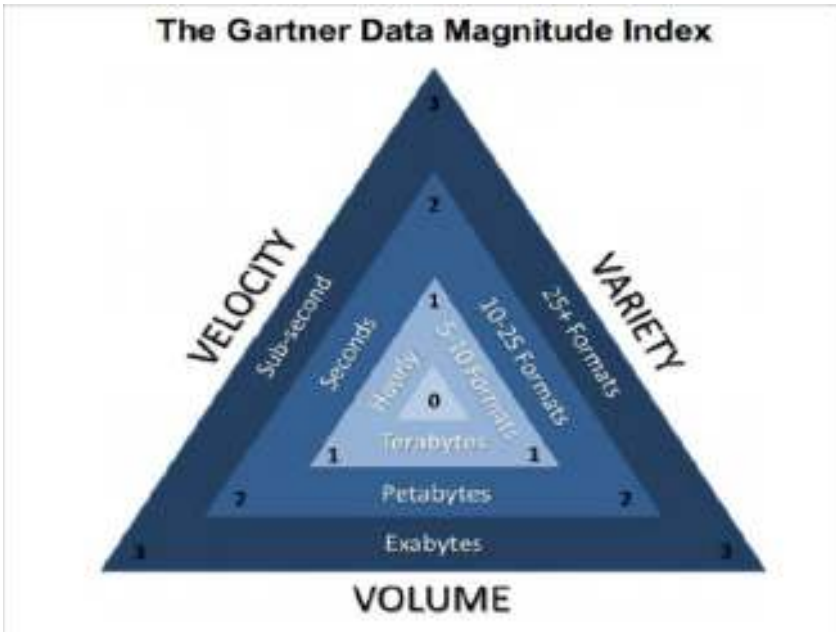

Figure 1: Data magnitude index
Source: Laney (2012)

According to Rich (2012), big data comes from various channels such as posts on social networks, photos, audio-visual and mobile phones, satellites and Global Positioning Systems indicators. Besides Laney (2012), a number of other scholars (Russom 2011; Kwon & Sim, 2013; Sagiroglu & Sinanc, 2013) also view big data in terms of volume, velocity and variety as discussed below.

Volume

According to Sharma (2016) the quantity of data has grown explosively and it is expected to continue growing as years go by, and Sharma asserts that it grows at the rate of about 2.5 Exabyte[1] each day. This results in data which is too large to manage using the traditional technologies. The situation is exacerbated by the fact that the volume of big data grows at an alarming rate. The large volume of big data can be attributed to the accessibility of devices like smart-phones and machines. Similarly, there have been an expansion and use of social network platforms which have facilitated the sharing of information more widely than before (Assunção, Calheiros, Bianchi, Netto & Buyya, 2015).

Academic institutions generate vast volumes of data. The data is generated by students and staff conducting research and from the day to day processes of the institutions. In the late 1980s and early 1990s, academic institutions invested in student information management systems which led to the accumulation of large volumes of data comprising various records of students including financial and academic records (Karani & Moturi, 2013). Enhanced data creation capacity has enabled academic institutions to be data-rich and to generate and use enormous volumes of data each day. However, most of these institutions have not exploited the opportunities brought about by the massive data they produce. Therefore, this data does accord them a competitive advantage. Unless academic institutions manage their large volumes of data by availing resources and employees with skills to manage the data, the institutions may not be able to meet the expectations of their clients. The “missing marks syndrome” is a direct result of an institutional lack of

1. It is a multiple of the unit byte for digital information in international system unit (SI). Exa indicates multiplication by sixth power of 1000(10^{18}).

capacity to manage big data which continues to plague many academic institutions in Kenya.

Velocity

According to Frizzo-Barker *et al*. (2016), velocity is the swiftness with which the examination of big data is conducted. It specifies the speed of data creation and processing thereby enhancing the availability of information when needed by clients. Real-time information processing can enhance an organisation's competitive advantage. Velocity also relates to the speed at which data seekers are able to search and retrieve the data they require (Porche, Wilson, Johnson, Tierney & Saltzman, 2014). Goes (2014) asserts that the concept of big data enables an advanced control speed of systems leading to new information termed as real-time information. Several researchers are of the view that the speed at which data is created is more important than the volume of data because the speed of data creation and retrieval gives an organisation the competitive advantage (Davenport & Harris, 2007). The speed of big data is based on the advancement of processors which permits the dispensation of real-time information.

Variety

Variety denotes the different forms of data available. This can be expressed in terms of structured, unstructured or semi-structured data (Frizzo-Barker *et al*., 2016). Big data is manifested in diverse formats, such as social media where data takes different forms such as notifications, messages and status updates (Sharma, 2016). Academic institutions produce big data from various activities which yield different varieties of big data such as videos, images and text documents. In academic institutions the variety of data can be categorised as follows:

- Students and staff data;
- Social media data;
- Institutional marketing data;
- Website browsing patterns data;
- Research data gathered by all staff; and
- Institutions' process data such as financial and admissions data, among others

Vesset, Woo, Morris, Villars, Little, Bozman and Eastwood (2012) added a fourth "V", which stands for *value*, to the notion of big data. Value relates to the benefits and usefulness of data in an organisation. Quality data enables decision making in an organisation to be effective and efficient (Zhou, Fu & Yang, 2016). In addition, White (2012) recommended a fifth "V", for *veracity*, to the big data concept. Veracity entails an assessment of the quality of data and the level of trust of various data sources. The quality of data is important because it determines the accuracy of information generated from it. For instance, an incorrect link can lead to an institution having inappropriate analysis of an organisation's opportunities. According to Assunção *et al*. (2015), veracity is linked to reliability and the quality of data. Failure to use reliable or quality data leads to limited value or negative influence on an organisation's performance.

According to Quitzau (2013), big data brought into being the concept of open data, and introduced another "V", for *visibility* of open data, as regards the issues of privacy and security. Open data is data provided with no cost or license constraints. Organisations have to decide what to make open and what remains as closed data based on their privacy, commercial and security concerns. Quitzau (2013) observed that the concept of big data can be explained as illustrated by Figure 2 below.

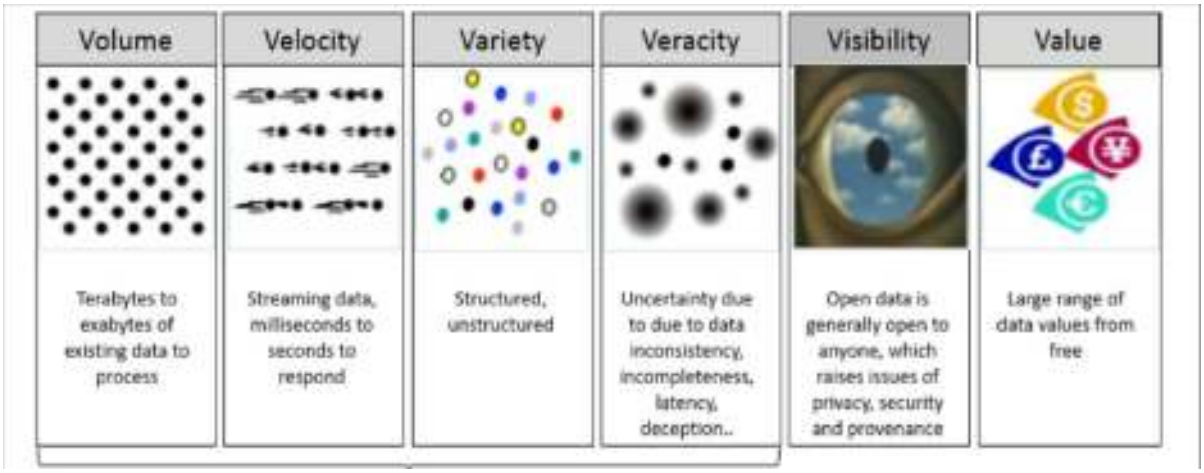

Figure 2: The six V's of big data
(Adapted from Quitzau, 2013, slide 24)

Background of the study

In academic institutions, big data comes from various academic, managerial and operational processes. On 25th July 2015[1] International Business Machines (IBM) invested six billion Kenya shillings (USD 60 million) in a skills venture platform in academic institutions (universities and polytechnics) in East Africa. The fund was used to develop cloud systems for over fifty institutions of higher learning in Kenya to enable the development and management of study material. The programme was undertaken in partnership with Kenya Education Network (KENET) (Ochieng', 2015). The skills imparted included cyber-security, mobile education and business analytics. The training began on September 2015[2] with technology and engineering students to prepare them for the job market and to arrest the situation where some students do not get jobs due to non-recognition of their degrees by the Engineers Board of Kenya arising from non-registration of the engineering courses offered by their universities. In December 2016, Dell Egan, Marino and Curly (EMC) held a meeting with Information Technology (IT) faculty from Kenyatta University, Riara University, Moi University, Multimedia University, Jomo Kenyatta University of Agriculture and Technology, Kenya College of Accountancy University, Strathmore University, United States International University – Africa and Africa Nazarene University. The faculty were trained on data science and big data analytics with the aim of empowering them to deal with the large amount of content their universities were generating and to improve the quality of service offered by the universities to their clientele (Sang, 2017).

These initiatives demonstrate that universities in Kenya are not new to the concept of big data and big data analytics. Nonetheless, it is also evident from the foregoing that universities in Kenya are commencing to experiment with big data, including its characteristics, tools and procedures. This study was conducted with that background, and specifically, in the contexts of private and public universities in Kenya. Two universities, that is, Strathmore University and The Technical University

1. *http://mobile.nation.co.ke/business/IBM-to-spend-Sh6bn-on-training-in-universities/1950106-2808272-format-xhtml-ooogskz/index.html*
2. *http://www.nation.co.ke/lifestyle/smartcompany/IBM-Sh6bn-plan-to-train-tech-students-takes-off-/1226-2890260-kwht3c/index.html*

of Kenya, were used to provide the contexts. These two institutions were used to facilitate an understanding of the differences and/or similarities between characteristics of big data from the perspectives of private and public universities in Kenya.

Rationale of the study

Academic institutions have not been left behind in the production of big data in terms of student and staff records, research output and innovations, as well as administrative, logistics, financial and procurement records. These records are produced fast, in vast volumes and diverse formats. Mayer-Schonberger and Cukier (2013) affirmed that big data is the new petroleum that will power the future knowledge economy. Academic institutions are operating in increasingly complex and competitive environments and need to respond to the changing world around them. The institutions need to exploit the data they generate to respond better to the changes around them. To do this, they need to overcome challenges associated with big data management. Some of these challenges revolve around difficulty to analyse, capture, curate, search, share, storage, transfer, visualise, and protect big data. The value of big data to organisational performance can only be unleashed through effective analytics. To analyse the data, new approaches must be used to process it. However, before analysing it, there is a need for organisations to identify the different characteristics of the big data they produce or own. This study investigated the characteristics of big data produced by two academic institutions in Kenya.

Methodology of the study

This study was designed as a mixed methods research project. According to Creswell and Plano-Clark (2015), mixed method design enables the understanding of a problem of study by gaining different corresponding data and enhancing their validation. This study adopted a convergent parallel design. Both qualitative and quantitative data was collected and analysed. The authors merged the results from both sets of data for comparison and validation to enhance the interpretation of similar and dissimilar concepts. Primary data was collected through structured questionnaires and interviews from the Technical University of Kenya (TUK) and

Strathmore University (SU). The study population was 22,050. It comprised 15,020 students and information communication technology (ICT) staff from TUK and 7,030 from SU. Information-oriented purposive sampling technique was used in the selection of information-rich subjects. Thus, class representatives in all academic programmes in both universities were selected to participate in the study, while all the ICT staff, directors and managers were chosen. The sample size for the study was 694 respondents, of which TUK was 580, while SU was 114. Questionnaires were distributed to all the students using a drop and pick technique. Face to face interviews was used to collect data from all ICT staff in both universities as shown in Table 1. The staff acted as key informants in the study.

Table 1: Sample size

Institution	Respondents	Population	Sample size
Technical University of Kenya	ICT Staff	20	20
	Students	15,000	560
Strathmore University	ICT Staff	30	30
	Students	7,000	84
Total		22,050	694

Quantitative data was analysed using statistical analysis by the help of SPSS and presented using descriptive statistics while qualitative data was analysed using thematic analysis.

Findings of the study

The response rate was obtained by calculating the number of respondents who successfully completed the questionnaire or participated in the interview divided by the sample size, multiplied by one hundred to get the percentage. The research involved a total sample size of 694, comprising 580 from the Technical University of Kenya and 114 from Strathmore University. The research response rate per university is shown in Table 2 below.

Table 2: Research response rate

Academic institutions	Respondents	Sample size	Number of responses	Response rate

TUK	Staff	20	15	82%
	Students	560	459	
SU	Staff	30	24	89%
	Students	84	78	
Source: Researcher (2018)				

Data analysis for this study was based on responses from questionnaires administered and interviews conducted by the researchers. TUK had a response rate of 82% while SU had 89%. The difference in response was due to the challenge of administering the questionnaires to respondents at TUK, because there was no drop-off and pick-up point while SU had a drop-off pick-up point hence the slight difference. Overall, 576 respondents participated successfully in the study, giving a response rate of 83%. Of the sample, 118 (17%) did not provide responses, that is, they did not return the data collection tools while some pulled out of the interviews. According to Mugenda and Mugenda (2012), a response rate of at least 50% is adequate for analysis; a 60% response rate is generally good while a 70% response rate is excellent. This is in agreement with Kothari (2014) who asserts that a response rate of above 70% is deemed to be very good for data analysis.

Characteristics of big data at TUK and SU

Identifying the characteristics of big data is essential to the understanding of the different forms of big data generated by institutions on a day-to-day basis. The majority of the respondents 32 (86%) from both institutions described big data as data drawn from a combination of multiple sources and consisting of multiple formats that require advanced technology to process. The following are some of the verbatim reports of the respondents:

> *I can't add much to the usual definitions of big data. However, what big data makes special to me is the combination of multiple data sources as well as multiple types of data and how it is unstructured. [SU]*

> *Big data is hard to process, and comes from even unstructured elements like social media and blogs. [TUK]*

> *All the forms of data that were difficult to process by computers, for certain reasons, but which we are now able to process. [SU]*

The study also found that big data in academic institutions comes from various units which are scattered within the institution. Some of the respondents, 28 (72%), suggested that this data should be brought together and analysed to get insights that can help to improve the institutions. They said that:

> *Big data in the academic institutions is often discussed in relation to the scattered data [TUK]*
>
> *Big data is actually like a cloud solution, which makes data more accessible. [TUK]*
>
> *Big data is not about the technology and is not about the data. At a certain point you have a goal and you want information for this goal. Many sources of data are not integrated. What you want is a data environment that integrates these data sources, which allow data sharing to be faster and more flexible. [SU]"*

Views of the respondents in the category of students on big data characteristics are shown in Table 3. The results in Table 3 indicate that most of the respondents, 358 (33%) from TUK and 56 (29.6 %) from SU, handled big data volumes in terms of Gigabytes (GB). The minority, 190 (17.5%) from TUK and 34 (18.0%) SU, dealt with Terabytes in terms of the volume of big data. For characteristics related to velocity, most of the respondents use real=time processors to enable them to gain insights from big data. This was according to 273 (42%) from TUK and 47 (42%) from SU. Most respondents from TUK, 363 (14.8%), dealt with photo, video and audio varieties of big data, while SU dealt with e-mails, 67 (15%). This showed that both the academic institutions and their clientele understand the concept of big data and its characteristics.

Table 3: Students' responses on the characteristics of big data

	Strathmore university		Technical university of Kenya	
	Responses	Percent	Responses	Percent

	Volume of big data				
	Kilobytes	44	23.3%	223	20.6%
	Megabytes	55	29.1%	314	28.9%
	Gigabytes	56	29.6%	358	33.0%
	Terabytes	34	18.0%	190	17.5%
Velocity of big data					
Real time		47	42%	273	42%
Periodic		31	28%	178	27%
Batch		28	25%	127	20%
Near real time		6	5%	71	11%
Variety of big data					
Email		67	15.40%	329	13.40%
Photo, video and audio		61	14.00%	363	14.80%
Social media data		55	12.60%	362	14.80%
Ms. Office data		54	12.40%	333	13.60%
Cell phone data		53	12.20%	327	13.30%
Financial data		44	10.10%	96	3.90%
Website content		40	9.20%	173	7.10%
Blogs		32	7.40%	219	8.90%
Gaming related data		18	4.10%	86	3.50%
Web logs data		6	1.40%	82	3.30%
Click stream data		3	0.70%	39	1.60%
Sensors data(GIS)		2	0.50%	43	1.80%

Discussion of the findings

The objective of the study was to identify the characteristics of big data produced by TUK and SU. The main goal was to determine whether the institutions generated the "3Vs" of big data. The results of the study indicate that the two institutions generate big data in terms of volume, variety and velocity. These findings support the views of Laney (2012) that the 3Vs are the most common characteristics of big data that organisations generate. A number of scholars (Russom, 2011; Kwon & Sim, 2013; Sagiroglu & Sinanc, 2013) also concur with Laney (2012) that big data is characterised in terms of volume, velocity and variety. The results are similar to the findings of Daniel (2015) who also observed that big data in academic institutions is perceived in terms of volume, velocity and variety.

The most commonly produced volume of big data was GB (TUK 33% and SU 29.6%) and least was TB (TUK 17.5% and SU 18%). This is because they are producing data from similar clients such as students as well as teaching and non-teaching staff. The volume of big data can differ from institution to institution based on the activities and the number of clients they serve on a day-to-day basis. Thus, some institutions can generate gigabytes of data while others generate terabytes (Jakub, 2015). The

findings further indicated that TUK and SU did not have major differences in the volume of big data they generate. This can be attributed to the fact that they have clients with similar characteristics. Smolan (2013) asserts that as the amount of big data rises day-by-day, it is expected that the institutions concerned begin to face challenges associated with managing and getting meaning from the data, hence the need to undertake big data analytics.

The study found that TUK and SU produce various forms of big data. The varieties include data generated from e-mail, photos, videos and audio, social media, MS Office applications, cell phones, financial transactions, website content, blogs, gaming and related applications, web logs, click stream, and GIS utilities. The most generated formats of big data from TUK are photos, video, audio and social media data 363 (14.8%) while SU mainly generate e-mail-based data 67(15.4%). Photos, videos, audio and social media data are examples of unstructured big data. According to IBM (2017), 80% of the data organisations currently generate are unstructured and consist of diverse formats such as text, video, audio, diagrams, images and combinations of any two or more formats. The results of the study also concur with Basu (2013) who observed that most organisations today run on unstructured data. Similarly, these results compare favourably with Shacklock (2016) who averred that academic institutions generate data of different varieties coming from students and staff log-in, research and day-to-day processes.

In terms of velocity, the study found that the two universities process big data using real-time, periodic, batch and near-time approaches. Real-time processing for big data was preferably used by 42% of the respondents because they wanted to get meaning from the data as fast as possible. Real-time processing technology captures, processes, and responds to big data as the events generating that data are happening in the real world. It deals with a continuous stream of inputs and has strict deadlines for completing the tasks. The approach of processing of big data in real-time is termed as a best practice because it enables immediate retrieval and use of data (Ounacer, Talhaoui, Ardchir, Daif & Azouazi, 2017). The approach enables institutions to keep abreast with data generation speeds and to respond to the needs

of their clientele (Borkar, Carey & Li, 2012). The results were also in agreement with Frizzo-Barker *et al.* (2016) who asserted that to get meaning from big data, real-time data processing is the best processing technique. Porche *et al*. (2014) also argue that there is need for the clientele to be able to retrieve data they require as fast as possible. Davenport and Harris (2007) were of the view that the speed at which data is created and retrieved is important because speed of data creation and retrieval helps an organisation to get meaning from the data promptly. The respondents indicated that real-time approach of big data processing enables the institutions to take immediate action when responding to events, issues or scenarios. The short response time span and access to up-to-date information gives the institutions the ability to gain insight from the updated data to detect patterns of either opportunities or threats to their operations in a timely manner.

Conclusion

The study concludes that the two institutions generate big data in terms of the 3Vs (volume, variety and velocity). This is an indication that both institutions and their clientele understand the concept of big data and its characteristics.

Recommendation

Both TUK and SU produce large volumes of big data as indicated by the results of the study. These volumes come from students' information, including the enrolment, academic and disciplinary records and also from staff and research. The academic institutions should leverage on these big data through big data analytics to improve the life of students in their institutions. The universities can address student needs with customised modules, assignments, feedback and learning trees in the curriculum that will promote better and richer learning.

References

Assunção, 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.

Basu, A. T. A. N. U. (2013). Five pillars of prescriptive analytics success. *Analytics magazine*, 812

Borkar, V., Carey, M. J. & Li, C. (2012). Inside Big Data management: ogres, onions, or parfaits? *Proceedings of the 15th International Conference on Extending Database Technology*, New York, 3-14.

Creswell, J. W. & Plano, C. V. L. P. (2015). Designing and conducting mixed methods research. Sage, Thousand Oaks.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press, Massachusetts.

Demirkan, H. & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M. & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, *36*(3), 403-413.

Goes, P. B. (2014). Editor's comments: big data and IS research. *MIS Quarterly*, *38*(3), iii-viii.

International Business Machine (2017). Turn data and analytics into competitive advantage. Retrieved 25 January, 2018 *https://www.ibm.com/big-data/ke/en/big-data-and-analytics/.*

Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM* 52(8), 36-44.

Jakub K. (2015). Big data analytics for mobile networks. Czech Technical University, Prague.

Karani, M. & Moturi, C. (2013). *Knowledge based student academic advising in Institutions of higher learning in Kenya* (Master's thesis). Retrieved from University of Nairobi Institutional Repository.

Kothari, C. R. (2014). Research methodology: Methods and techniques. New Age International, New Delhi.

Kwon, O. & Sim, J. M. (2013). Effects of data set features on the performances of classification algorithms. *Expert Systems with Applications*, 40(5), 1847-1857.

Laney, D. (2012). 3D data management. Retrieved October 2, 2017 from *http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.*

Mayer-Schonberger, V. & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, London.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J. & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60-68.

Mugenda, O. M. & Mugenda, A. G. (2012). Research methods: quantitative and qualitative approaches. Acts Press, Nairobi.

Ounacer, S., Talhaoui, M. A., Ardchir, S., Daif, A. & Azouazi, M. (2017). A New Architecture for Real Time Data Stream Processing. *International Journal of Advanced Computer Science and Applications*, *8* (11), 44-51.

Porche, I. R., Wilson, B., Johnson, E.-E., Tierney, S. & Saltzman, E. (2014). *Data flood: helping the Navy address the rising tide of sensor information*. National Defense Research Institute, Santa Monica.

Quitzau, A. (2013). Transforming Energy and Utilities through Big Data & Analytics. Retrieved February 20, 2018 from *https://www.slideshare.net/ AndersQuitzauIbm/big-data-analyticsin-energy-utilities*

Rich, S. (2012). Big Data Is a 'New Natural Resource'. Retrieved January 01, 2018 from *http://www.govtech.com/policy-management/Big-Data-Is-a-New-Natural-Resource-IBM-Says.html.*

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*. TDWI Research institute, London.

Sagiroglu, S. & Sinanc, D. (2013). Big data: A review. *Proceedings of Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.

Savvas, A. (2011). IBM: Businesses unable to analyse 90 percent of their data. Computerworld. Retrieved November 10, 2017 from *https:// www.computerworlduk.com/it-vendors/ibm-businesses-unable-to-analyse-90-percent-of-their-data-3313304/*

Shacklock, X. (2016). *From bricks to clicks: the potential of data and analytics in higher education*. Higher Education Commission, London.

Sharma, S. (2016). Expanded cloud plumes hiding Big Data ecosystem. *Future Generation Computer Systems*, *59*, 63-92.

Smolan, R. (2013). The human face of big data. Radio Adelaide, Adelaide.

Vesset, D., Woo, B., Morris, H. D., Villars, R. L., Little, G., Bozman, J. S. & Eastwood, M. (2012). Worldwide big data technology and services 2012–2015 forecast. *IDC report*, *233-485*.

White, M. (2012). Digital workplaces: Vision and reality. *Business Information Review*, *29*(4), 205-214.

Zhou, K., Fu, C. & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, *56*, 215-225.