The Characteristics of Big Data in Successful of Software Development Life Cycle for Mobile Application

Zairina Ibrahim, Md Gapar Md Johar, Che Ku Nurul Hasmaria Che Ku Yahaya, Normy Rafida Abdul Rahman

Abstract: ***This paper aims to software development life cycle for big data, software development and mobile application for the new environment of the real world. The big data is a new design in the world for the Internet devices to connected life in new environment. The paper reviews 6Vs such as (Volume, Variety, Velocity, Variability, Veracity and Value) for big data an environment application used in the mobile application for students and instructors in university. In fact, for the digital content is to efficiency and effectively for accessed the contents with anytime and everywhere to capture the speeding and capacity data. Current research on software development on application for students and instructors is limited to readiness and awareness. This research extends the need for empirical findings from system analyst and top management of digital contents for big data approach.***

Index Terms*: *Big Data, 6 Vs, Software Development Life Cycle, Digital Content.

I. INTRODUCTION

A Big Data is a terminology that is now becoming a buzz in the world of research data science and data analytics. In general, Big Data is a phenomenon of abundant amounts of data (in digital form) that can now be stored and processed into valuable information for various purposes. The data is abundant on the World Wide Web (WWW) in the form of text on Facebook, images on Flickr, videos on Youtube, data on our location on Foursquare or Google Map, temperature data on the thermostat of the refrigerator or AC, to the data from CCTV camera sensors, all now can be captured and processed into information that is very useful for human life. Because the data taken from various tools comes from billions of people, then you can imagine how much data is collected (De Cnudde & Martens, 2015).

Therefore, in 2012, the tweets were more than 400 million a day with a growth of 800% a year. In the same year, at least we "like" on Facebook more than three billion times a day. That's just a day, imagine the amount if the data is collected a month, a year, or even years Chen et al. (2015). The Google is currently processing around 24 petabytes (Petabytes = 1 million gigabytes) of data every day. One of them is our search query that is hundreds of millions a day. This amount of data is astonishing because it is roughly equivalent to thousands of times the total number of printed documents that the Library of Congress has had to date Chan (2016).

The Volume for data capacity related to the size of the data storage media that is very large or may be unlimited to units of petabytes or zettabytes. The Variety related to the type or type of data that can be processed from structured data to unstructured data. In fact, while Velocity is related to the speed of processing data generated from various sources, ranging from batch data to real time, while the characteristics of Veracity (truth) and Value (value) are related to data uncertainty and the value of benefits from the information produced Sukumar & Ferrell (2013).

The benefits of Big Data have been felt, especially in the education sector, including to knowing the community's response to the products issued through sentiment analysis on social media. In fact, to help education lines make decisions more precisely and accurately based on data. In addition, to being used for business analysis, there is great hope that Big Data technology will also be widely used in government. Some opportunities for Big Data utilization in the public sector include getting community feedback and response as a basis for policy formulation and improving public services Brown (2014).

Table I: Big Data is defined in terms of either the first 3, or all 6, Vs.

Vs	Description
Volume	To define the size of Big Data that is stored and managed by an organization. Its evaluates the massive amount of data in data stores and concerns related to its scalability, accessibility and manageability.
Velocity	To define the speed of increase in Big Data volume and its relative growth of their big data and how quickly that data reaches sourcing users, applications and systems.

Revised Manuscript Received on September 22, 2019.

Zairina Ibrahim, Information Technology and Innovation Centre, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia.

Md Gapar Md Johar, Information Technology and Innovation Centre, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia.

Che Ku Nurul Hasmaria Che Ku Yahaya, Information Technology and Innovation Centre, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia.

Normy Rafida Abdul Rahman, Faculty Business Management and Professional Studies, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia.

Valence	Its how Big Data can bons with each other, forming connections between otherwise disparate datasets.
Veracity	To the quality or trustworthiness of the data, which can vary greatly.
Variety	To the ever an increasing different forms that data can come in.
Value	To the worth of the data being extracted and processing Big Data must bring about value from insights gained.

Source: Techopedia and Jenn Cano (March 2011)

Nevertheless, according to Naismith et al. (2004) century 21th, big data is more popular and interested for the mobile applications to more scalability, accessibility and manageability models for using and developing mobile applications for learning are somewhat lacking. There is a need to formulate appropriate pedagogical models and to develop innovative strategies to integrate mobile applications in learning and teaching. Our project proposes to address this need by designing learning materials and applications that take advantage of the unique features of the cellular phone. In particular, our educational design takes into consideration the socio-cultural, situated learning paradigm. Therefore, according to Low and O'Connell (2006), the data connectivity and communication aspects of mobile devices support social interaction, collaboration, and the construction of learning, and may enhance interpersonal communication Taylor et al. (2005). Thus, being able to exchange work and applications through MMS and SMS, students and instructors can create a community in which they can work together, share knowledge, inspire each other, and interact socially (Tu and Corry, 2003; Reynolds et al., 2001). However, mobile devices also offer opportunities to gain access to learning experiences while being immersed in a learning context of real world (Low and O'Connell, 2006). According to Naismith et al. (2004), embedding the learner in a realistic context at the same time as offering access to supporting tools can enhance the active construction of personal knowledge.

II. LITERATURE REVIEW

Since the back of this year, mobile applications are very popular among students. Therefore, in education it is very important to increase the level of education to the e-learning rank to make it easier for students to access learning outside the classroom or everywhere Fuchs et al. (2014).

A. Big Data Parking

Nowadays, the clouds are a solution for some university or company to store the data. But then, they also throw up fresh challenges. Paradoxically, their production can cause a bottleneck if data end up parked on several clouds and thus still need to be moved to be shared. In fact, a using clouds means entrusting valuable data to a distant service provider who may be subject to power outages or other disturbances. According to Hunter (2013), using the cloud services an experience for a many things, but then always keep a local copy of scientifically important data and software. The scientists experiment with different groups to suit their needs and trust levels.

However, according to Sundquist (2013), most researchers tend to download remote data to local hardware for analysis. Therefore, this method is backward. In fact, the data are so much larger than the tools, it makes no sense to be doing that. Thus, the alternative is to use the cloud for both data storage and computing.

B. Data Highway

In previous research, according to Beijing Genomics Institute, they can transfer about 1 terabyte per day to its clients. In fact, if the transferring data one genome at a time, the data should not have a problem. Thus, if the data sequence 50, that is not practical to transfer that through the internet and take about 20 days to completed Helbing et al. (2011).

C. Challenges and Opportunities

The connecting powerful computers and numerous tools for data analysis is crucial in Massive data centers filled with exabytes (one billion gigabytes) of transaction records, financial information, browsing habits, social media activity, and mobile data are impotent without software developers writing programs to facilitate the analytics process Helbing et al. (2011).

accommodation for students there are students who choose off campus accommodation to on campus due to lack of privacy sharing room or place with new students and some decide off campus accommodation for the price which is reasonable or cheaper, new housing and new accommodation with good facilities plus near to university [16]. Stating that compulsory all students must follow the rules and regulation that come from campus and any decision at university campuses "tended to work unilaterally and usually without question, and housing for these students tended to work within the same power structure with colleges instilling restriction rather than freedom and residence with appointed bounds [14].

III. CONCEPT OF BIG DATA

Big Data terminology is often associated with data science, data mining, and data processing. However, Big Data involves infrastructure and data mining techniques or data processing more than ever before. In implementing Big Data technology in an organization there are 4 important elements that are challenges, namely data, technology, processes and human resources (Aryasa, 2015).

A. Data

The basic description of the data refers to objects, events, activities, and transactions that are documented, classified, and stored but not organized to provide a specific meaning. Data that has been organized so that it can give meaning and value to the recipient, The basic description of the data refers to objects, events, activities, and transactions that are documented, classified, and stored but not organized to provide a specific meaning. Data that has been organized so that it can give meaning and value to the recipient, called information. Data availability is the initial key to Big Data technology. There are several

organizations that have a lot of data from their business processes, both structured and unstructured data, such as the telecommunications industry and banking. However, there are also organizations that need to buy or cooperate with other parties to get data Butler (2008).

B. Technology

According to Caragliu et al. (2011), this is related to infrastructure and tools in the operation of Big Data, such as computational and analytical techniques, and storage media. Usually, organizations will not experience significant obstacles in terms of technology because technology can be obtained by buying or collaborating with third parties.

C. Process

In the process of adopting Big Data technology requires changes in organizational culture, according to Caragliu et al. (2011), for example, before the existence of Big Data, a leader in running an organization, making decisions only based on 'intuition' based on their values, beliefs or assumptions. But after the Big Data technology, the leader is able to act "data-driven decision making" means to make decisions based on accurate data and relevant information.

D. Subscriber Data Management

In applying big data technology, in previous researched by Fan et al. (2015), human resources are needed with analytic and creativity skills, namely the ability / skill to determine new methods that can be done to collect, interpret and analyze data, computer programming skills, and business skills, namely understanding business objectives.

E. Big Data Analysis

Data analytics refers to the Business Intelligence and Analytics technologies that are grounded mostly in data mining and statistical analysis. According to Chaudhuri et al. (2011), most of these techniques rely on the mature commercial technologies of relational DBMS, data warehousing, Extract, Transform and Load (ETL), Online Analytical Processing (OLAP), and Business Process Management (BPM). Therefore, before do the analysis for the big data task to solve is big data integration, manipulation, quality and governance as well as a project management.

According to Wu et al. (2007), late 1980s, various data mining algorithms have been developed by researchers from the artificial intelligence, algorithm, and database communities. In the IEEE 2006 International Conference on Data Mining (ICDM), the 10 most influential data mining algorithms were identified based on expert nominations, citation counts, and a community survey. In ranked order, they are C4.5, k-means, SVM (support vector machine), Apriori, EM (expectation maximi- zation), PageRank, AdaBoost, kNN (k-nearest neighbors), and CART. In fact, according to witten et al. (2011) these algorithms cover classification, clustering, regression, association analysis, and network analysis.

However, most of these popular data mining algorithms have been incorporated in commercial and open source data mining systems. Therefore, other advances such as neural networks for classification, prediction and clustering and genetic algorithms for optimization and machine learning have all contributed to the success of data mining in different applications.

Nevertheless, due to the success achieved collectively by the data mining and statistical analysis community, data analytics continues to be an active area of research. Statistical machine learning, often based on well-grounded mathematical models and powerful algorithms, techniques such as Bayesian networks, Hidden Markov models, support vector machine, reinforcement learning, and ensemble models, have been applied to data, text, and web analytics applications. In fact, other new data analytics techniques explore and leverage unique data characteristics, from sequential/temporal mining and spatial mining, to data mining for high-speed data streams and sensor data.

However, according to Gelfand (2011/2012) an increased privacy concerns in various e-commerce, e-government, and healthcare applications have caused privacy preserving data mining to become an emerging area of research. According to Van Der Aalst (2012), many of these methods are data driven, relying on various anonymization techniques, while others are process driven, defining how data can be accessed and used. Therefore, over the past decade, process mining has also emerged as a new research field that focuses on the analysis of processes using event data. Thus, process mining has become possible due to the availability of event logs in various industries such as supply chains and new process discovery and conformance checking techniques.

However, in addition to active academic research on data analytics, industry research and development has also generated much excitement, especially with respect to Big Data analytics for semi structured content. In fact, unlike the structured data that can be handled repeatedly through a RDBMS, semi structured data may call for ad hoc and one time extraction, parsing, processing, indexing, and analytics in a scalable and distributed MapReduce or Hadoop environment. According to Patterson (2008), MapReduce has been hailed as a revolutionary new platform for large scale, massively parallel data access. An inspired in part by MapReduce, Hadoop provides a Java based software framework for distributed processing of data, an intensive transformation and analytics. The top three com mercial database suppliers such as Oracle, IBM, and Microsoft. There are have all adopted Hadoop, some within a cloud infrastructure. According to Henschen (2011), the open source Apache Hadoop has also gained significant traction for business analytics, including Chukwa for data collection, HBase for distributed data storage, Hive for data summarization and ad hoc querying, and Mahout for data mining. However, according to Chaudhuri et al. (2011), the commercial parallel DBMS showed clear advantages in efficient query processing and high level query language and interface, whereas MapReduce excelled in ETL and analytics for “read only” semi structured data sets. New Hadoop and MapReduce based systems have become another viable option for big data analytics in addition to the commercial systems developed for RDBMS, column-based DBMS, in

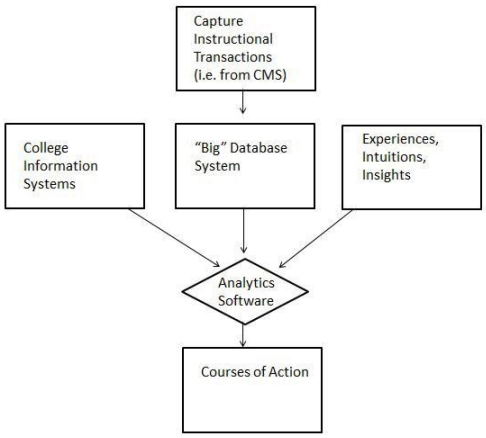
memory DBMS, and parallel DBMS.

F. Applications

The big data concepts and analytics can be applied to a variety of higher education administrative and instructional applications, including recruitment and admissions processing, financial planning, donor tracking, and student performance monitoring.

To take advantage of Big Data and learning analytics, it is almost a requirement that transaction processing be electronic rather than manual. Traditional face to face instruction can support traditional data driven decision making processes, however, to move into the more extensive and especially time sensitive learning analytics applications, it is important that instructional transactions are collected as they occur. This would be possible in the case of a course management/learning management system (CMS/LMS). Thus, most CMSs provide constant monitoring of student activity whether they are responses, postings on a discussion board, accesses to reading material, completions of a quiz, or some other assessment. However, an using the full capabilities of a basic CMS, a robust fifteen week online course could generate thousands of transactions per student. The real time recording and analysis of these transactions can be used to feed a learning analytics application. Critical to this type of application is not waiting learning analytics software application. The instructional transactions should also be integrated with other resources such as data from the university or colleage information systems (student, course, and faculty) and an analytics software program. The logic or decision trees for the latter are based on patterns as well as faculty and adviser experiences, intuition and insights that are used to develop guidelines and rules for subsequent courses of action in Figure 1. Nevertheless, one important caveat is that the data accuracy should never be compromised in favor of timeliness of the data, both for accuracy and for the end of a marking period or semester to record performance measures.

The reason this is important is that monitoring student transactions on a real time basis allows for real time alerts. Instructors may take actions or intervene in time to assist students. A CMS or something similar therefore becomes critical for collecting and feeding this data into a “big” database for processing by timeliness are important and need to be present in the learning analytics application Crush M. (2011).



Source: The evolution of big data and learning analyttics in American Higher Education by Professor Anthony G. Picciano, 2011

Fig. 1. Learning Analytics Flow Model

G. Data Security

According to Chaowei Yang et al. (2017), the increasing dependence on computers and Internet over the past decades makes businesses and individuals vulnerable to data breach and abuse. Thus, Big Data poses new security challenges for traditional data encryption standards, methodologies and algorithms. However, the previous studies of data encryption focused on small to medium size data, which does not work well for Big Data due to issues of the performance and scalability (Chen et al. 2014b). In addition, according to Villars et al. (2011), data security policies and schemes to work with the structured data stored in conventional DBMS are not effective in handling highly unstructured, heterogeneous data. Therefore, effective policies for data access control and safety management need to be investigated in Big Data and these need to incorporate new data management systems and storage structures (Cavoukian and Jonas 2012; Chen et al. 2014a). Nevertheless, in the cloud era, since data owners have limited control on virtualized storage, ensuring data confidentiality, integrity and availability becomes a fundamental concern (Kaufman 2009; Wang et al. 2009; Feng et al. 2011; Chen and Zhao 2012).

IV. STRUCTURED AND UNSTRUCTURED DATA

Data sources in Big Data technology can be structured and unstructured data. There are;

1. Structured data: has defined data types, formats, and structures. Can be in the form of transactional data, OLAP data, traditional RDBMS, CSV files, simple spread-sheets (De Cnudde et al., 2015).
2. Unstructured data: textual data with an uncertain format or lack of inherent structure, so to make structured data requires more effort, tools, and time. This data is generated by internet applications, such as URL log data, social media, e-mail, blogs, videos, audio and semantic data (De Cnudde et al., 2015).

V. CONCLUSION

Big data is an important to explore for Vs (Volume, Velocity, Valence, Variability, Variety and Value) for the effectiveness and efficiency data to store with the structured or not structured in digital contents of online learning. This is to encourage students for expose the environment of digital contents for everywhere and anywhere there are needs without bottleneck of the data during online session of learning. Therefore, the Vs can support the platform to enhance the device to use for more structured. However, for the mobile application can help

student to access the contents precisely without buffering. The big data is evolution of the concept of big data for data, technology, process, subscriber data management and big data analysis. Therefore, for the data source in big data technology for support the mobile application in online learning is included of structured data and unstructured data to make the technology going smoothly in the contents during the accessing.

The mobile application for the online learning is part of the solution to this problem but student attrition in universities and colleges is at unacceptable rates and needs to be addressed as well. Thus, data driven decision making is already being used to help universities and colleges identify and evaluate strategies that can improve retention. Therefore, as data driven decision making enters the big data and learning analytics era, these new approaches, while not silver bullets, may be part of the solution. However, an academicians will do well by evaluating whether they can be used in their universities or colleges and determining the role they can move to the mobile application for the transforming and enhancement online learning.

REFERENCES

1. Brown, M. S. 2014. Big Data, Mining, and Analytics. Components of Strategic Decision Making. In S. Kudyba (Ed.), Big Data, Mining, and Analytics. Components of Strategic Decision Making (pp. 211-230). Boca Raton: CRC Press Taylor & Francis Group.
2. Butler, D. 2008. Web Data Predict Flu. Nature, 456 (7220), 287-288. doi:10.1038/456287a
3. Caragliu, A., Del Bo, C., & Nijkamp, P. 2011. Smart Cities in Europe. Journal of Urban Technology, 18(2), 65-82. doi:10.1080/10630732.2011.601117
4. Cavoukian, A., and J. Jonas. 2012. Privacy by Design in the Age of Big Data. Information and Privacy Commissioner of Ontario, Canada. Accessed December 3 2015. https://privacybydesign.ca/content/uploads/2012/06/pbd-big_data. pdf.
5. Chan, J. O. 2016. Big Data Customer Knowledge Management. Communications of the IIMA, 14 (3), 5.
6. Chaudhuri, S., Dayal, U., and Narasayya, V. 2011. "An Overview of Business Intelligence Technology," Communications of the ACM (54:8), pp. 88-98.
7. Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu. 2017. Big Data and cloud computing: innovation opportunities and challenges, International Journal of Digital Earth, 10:1, 13-53.
8. Chen, J., Tao, Y., Wang, H., & Chen, T. 2015. Big Data Based Fraud Risk Management at Alibaba. The Journal of Finance and Data Science, 1(1), 1-10.Chen, M., S. Mao, Y. Zhang, and V. C. Leung. 2014a. Chapter 1, Big Data: Related Technologies, Challenges and Future Prospects .Heidelberg: Springer.
9. Chen, Z. K., S. Q. Yang, S. Tan, H. Zhao, L. He, G. Zhang, and H. Y. Yang. 2014b. The Data Allocation Strategy Based on Load in NoSQL Database. Applied Mechanics and Materials. 513–517: 1464–1469.
10. Coppersmith, D. 1994. The Data Encryption Standard (DES) and Its Strength Against Attacks. IBM Journal of Research and Development 38 (3): 243–250.
11. Crush, M. Monitoring the PACE of student Learning: Analytics at Rio Salado Community College.
12. *Campus Technology* (December 14, 2011).
13. De Cnudde, S., & Martens, D. 2015. Loyal to Your City? A Data Mining Analysis of a Public Service Loyalty Program. Decision Support Systems, 73,
14. 74-84. doi:10.1016/j.dss.2015.03.004.
15. Fan, S., Lau, R. Y. K., & Zhao, J. L. 2015. Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. Special Issue on Computation, Business, and Health Science,
16. 2(1), 28-32. doi:10.1016/j.bdr.2015.02.006
17. Fuchs, M., Hopken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. Journal of Destination Marketing & Management, 3(4), 198-209.
18. Gelfand, A. 2011/2012. "Privacy and Biomedical Research: Building a Trust Infrastructure—An Exploration of Data-Driven and Process-Driven Approaches to Data Privacy," Biomedical Computation Review, Winter, pp. 23-28 (available at http://biomedicalcomputationreview.org/content/privacy-and-biomedical-research-building-trust-infrastructure, accessed August 2, 2012).
19. Henschen, D. 2011. "Why All the Hadoopla?" Information Week, November 14, pp. 19-26.
20. Helbing, D., & Balietti, S. (2011). From social data mining to forecasting socio-economic crises. The European Physical Journal Special Topics, 195(1), 3-68. doi:10.1140/epjst/e2011-01401-8
21. Kaufman, L. M. 2009. Data Security in the World of Cloud Computing. IEEE Security & Privacy Magazine 7 (4): 61–64.
22. Nadeem, A., and M. Y. Javed. 2005. A Performance Comparison of Data Encryption Algorithms. In ICICT 2005. First International Conference on Information and Communication Technologies, 2005, 84–89.
23. Patterson, D. A. 2008. "Technical Perspective: The Data Center Is the Computer," Communications of the ACM (51:1), p. 105.
24. Redlich, R. M., and M. A. Nemzow. 2006. U.S. Patent No. 7,103,915. Washington, DC: U.S. Patent and Trademark Office.
25. Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data' Collaboration: Exploring, Recording and Sharing Enterprise Knowledge. Information Services & Use, 33(4), 257-270. doi:10.3233/ISU-130712
26. Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What Can Big Data and Text Analytics Tell Us about Hotel Guest Experience and Satisfaction? International Journal of Hospitality Management, 44, 120-130.
27. Van Der Aalst, W. 2012. "Process Mining: Overview and Opportunities," ACM Transactions on Management Information Systems (3:2), pp. 7:1-7:17.
28. Villars, R. L., C. W. Olofson, and M. Eastwood. 2011. Big Data: What It Is and Why You Should Care. White Paper, IDC.
29. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D. J., and Steinberg, D. 2007. "Top 10 Algorithms in Data Mining," Knowledge and Information Systems (14:1), pp. 1-37.

AUTHORS PROFILE

I am Zairina Ibrahim, I am associated with Information Technology and Innovation Centre, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia and and my area of interest is mobile technology and big data.

My name is Md Gapar Md Johar, I am associated Information Technology and Innovation Centre, Management and Science University, University Drive Seksyen 13,
Shah Alam 40100, Malaysia and my area of interest is mobile technology and big data.

My name is Che Ku Nurul Hasmaria Che Ku Yahaya, I am from Information Technology and Innovation Centre, Management and Science University, University
Drive Seksyen 13, Shah Alam 40100, Malaysia and my area of interest is mobile technology and big data.

My name is Normy Rafida Abdul Rahman, I am from Faculty Business Management and Professional Studies, Management and Science University, University Drive Seksyen 13, Shah Alam 40100, Malaysia.