Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

A REVIEW ON THE PERFORMANCE MEASUREMENT OF BIG DATA ANALYTICS PROCESS

I. M. Ali*, Y. Y. Jusoh, R. Abdullah, and R. N. H. Nor

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), Selangor, Malaysia

Published online: 01 February 2018

ABSTRACT

Abstract— one vital aspect of big data is the process which handles big data, mainly referred to as big data analytics process. Studying this process and measuring it for better performance can have enormous benefits for those who are willing to harvest the value of big data. The focus on processes is more common in other disciplines such as business and software development but in big data analytic, it is relatively new. The purpose of this paper is to investigate big data analytics (BDA) process and identify suitable performance measures for it. Existing BDA processes and performance measurement methods are studied and the results are presented in this paper

Keywords: big data analytics process, measures, performance measurement, process performance dimensions

1. INTRODUCTION

Prior to big data, data analytics was not as challenging as it is in the era of big data. A significant number of software tools, data warehousing, and straightforward Extract, Transform, and Load (ETL) process are the ones that characterized the traditional data analytics. Big data along with its huge difference in terms of volume, velocity, and variety [25], has introduced a major shift in analytics settings.

Author Correspondence, e-mail: Ismaaciil15@gmail.com

doi: http://dx.doi.org/10.4314/jfas.v10i2s.9

This trend, some called it fast analytics [19], imposes new requirements mainly in terms of data processing technology, data collection mechanisms, and most importantly, the process that data go through, along with other requirements such as human capital and organizational culture.

Big data should be in a matter of process; the process of harnessing big data, extracting insights, and creating value out of it, thereby measuring performance and making informed decisions, for organizations such as higher learning institutions. The BDA process which extends from data acquisition and selecting the sources through data preparation, analysis, and visualization, to interpretation phase, has enormous importance, an issue that has been less focused by previous researches and practices. This review paper is a part of on-going research which is aimed at measuring the performance of BDA process and examining how this can contribute to higher learning institutions.

This research utilizes existing performance measurement techniques whether they are process-level such as process performance dimensions [3] or organizational level such as Balanced Scorecard (BSC) [6]. The above two considerations are important to clarify how an improved BDA process (process-level) can contribute to higher learning institutions (organizational level).

2. MOTIVATION AND CHALLENGES

Looking at big data analytics in a process perspective has major benefits since improving process drives a better outcome; satisfied customers, evidence-based practices, and better investment strategy. Big data, as new phenomenon, definitely requires such an enlightening study so efforts become more effective and maximum benefits could be gained. Is there well-defined BDA process? How do existing performance measures apply to BDA process? These questions draw the challenges ahead.

3. RELATED WORK

3.1. Big Data Analytics (BDA) Process

Process is defined by Cambridge Dictionary as a series of actions that you take in order to achieve a result. Process is also referred to as an approach of converting inputs into outputs [27]. Elsewhere, a process was defined as "a series of related activities and conversions that

are designed to gather input, and convert it into a desired result"[1]. A similar definition provided in [4], is a sequence of activities that produce a specific output based on a given input. Zairi[27] stated that any process should have the following defining features: 1) A linear, logical sequence or flow, 2) A set of clearly definable tasks or activities, 3)A predictable and desired outcome or result. These features draw that process is series of tasks which have clear inputs and intended results.

Big data analytics process which uses data to generate learning models that enable organizations to anticipate changes and to reveal the constantly shifting knowledge patterns [16]. Lifecycle, as a term that is comparable with process, is "the progression of something from the conception to end of lifecycle or when something no longer provides value" [19].

Although this study chose the naming of “Big data analytics process”, Very few authors, refer to [16], stated “Big data analytics process”. Other descriptions include “a process for extracting insights from big data Gandomi and Haider[5], process of analyzing unstructured data Wang et al[21], "big data lifecycle" as indicated in [22], [23] and [2] and "Analytics Workflow for Big Data" Assunção et al [24].

Other related terms include “phases of big data analytics” [15], “the stages of extracting values from big data” [20], and “steps of big data processing” Mello et al[10].

A consideration can also be given to how pre-existing process models can be aligned with fast analytics imposed by emergence of big data. Larson and Chang [19] reviewed agile methodology, business intelligence and fast analytics and proposed agile framework for BI delivery and fast analytics and data science. It was indicated that fast analytics comes with big data trend. The interesting thing here is that authors provided BI lifecycle and fast analytics lifecycle those were presented in parallel with agile methodology lifecycle. Larson and Chang[19] also expressed that the lifecycle consists of phases that represent the sequence of something from start to the end and , highlighted the similarities between of (Business Intelligence)BI lifecycle and agile methodology with respect to their lifecycle phases. However, fast analytics which is trigged by big data is said to have different approach because of the speed of the technology and the acquisition of data. The phases of fast analytics mentioned by Larson and Chang [19] include scope, data acquisition and discovery, analysis and visualization, modeling, validation, deployment, support and feedback. The phases will be discussed later in the section of phases of big data analytics.

According to LaValle et al cited in [16], BDA processes differ depending on the analytics model which can either be descriptive, prescriptive, or predictive. Rehman et al[16] presented big data analytics process which is broken into a number of stages including data collection, data preparation, modeling(learning model creation), evaluation, deployment and monitoring.

3.2. Performance Measurement

Glavan[26] defines performance as an achievement of a particular task measured against predetermined standards of completeness, accuracy , cost, and speed. Performance Measurement was described as the task of quantifying the efficiency and effectiveness of action [28]. Another definition of performance measurement is "the process of measuring efficiency, effectiveness and capability, of an action or a process or a system, against given norm or target"[29]. The two above definitions agree on two components of Performance Measurement those are effectiveness and efficiency. Effectiveness is said to be the measure of doing the right thing, while the efficiency refers to doing things in a right way. The second definition adds one more component which is capability. Nightingale [29] describes capability as the ability required for achieving for both effectiveness and efficiency.

According to literature analysis presented in [26] , process performance measurement systems is typically a new topic in area of performance measurement, and it becomes necessary to include the descriptions of other popular concepts such as business process management, business performance measurement, and performance measurement systems. Process performance can be related to process efficiency (such as its productivity), the compliance and effectiveness of the process (that could be sorted as process quality) [12].

Performance measurement can also be linked to traditional Business Intelligence (BI), the predecessor of big data analytics. There are two aspects of measuring business intelligence (BI), namely determining how BI efforts are worth undertaking and managing BI process. Existing measures examine the value of BI rather than focusing on BI process. Existing BI process models vary depending on the number of process phases included, the structure in which the process cycle is presented and sources and storage of information. For measuring the performance of BI process, Prism performance approach was indicated Lönnqvist and Pirttimäki [9].

Other performance measurement approaches include process performance dimensions, TOPP Performance Measurement System, Balanced Score Card (BSC), Statistical Process Control, and the Performance Pyramid. The process performance dimensions, some call them devil's quadrangle, are discussed later in findings; this approach is more related to the purpose of this study.

4. METHODOLOGY

This paper is based on literature review findings of BDA process and process performance measurement. The paper followed three-stage method [11], to conduct and report the literature review findings. The first stage is to investigate and collect the related articles. The next stage is to decide on how to capture and analyze data. The final stage is to write up the literature-based findings.

5. FINDINGS

5.1. The Phases of Big Data Analytics Process

Most of the studies that contained big data analytics process didn't address a problem related to a process, although the processes were mentioned in some context. Fewer studies [20], [2], [15] and [31] are in the context of education. One study [19] addressed the comparison between agile methodology, BI lifecycle, and fast analytics lifecycle.

5.1.1. Mapping Existing Big Data Analytics Processes

Determining a measurable big data analytics process is one major goal of this study. To achieve this goal, this study examined the existing processes and highlighted their similarities and differences. This concept is shown in Table1.

TABLE 1. The Mapping the Existing Big Data Analytics Process

Authors	Process Phases
Larson and Chang(2016)	Data acquisition, analyzes /visualize, model/ design, validate, deploy, among other phases for fast analytics and data science lifecycle
(Demchenko et al, 2015).	Data collection and Registration, data filter/enrich/classification, data analytics/ modeling, data redelivery/ visualizations for big data lifecycle
(Rehman et al, 2016)	Data collection, prepare data, model and evaluate, among other phases for big data analytics process
(Gandomi & Haider, 2015)	Acquisition, extraction/cleaning, integration/ aggregation, modeling and analysis, interpretation. NB all keywords appeared are not mentioned here.
(Daniel, 2014).	Data collection, data analysis, data visualization
Labrinidis & Jagadish(2012)	Discussed data acquisition, information extraction, data analysis the need for interpretation
(Assunção et al, 2014)	Mentioned a number of useful terms in overview of analytics workflow for big data, including are analysis, visualization, interpretation... and more
(Tulasi, 2013).	Acquisition, extraction, integration, analyzing, interpretation
(Gama et al, 2016)	Acquisition, access, analytics application
(Hu & Wen, 2014).	Generation, acquisition, storage, analytics

5.1.2. Selecting the Phases

BDA process consists of several phases. It starts with acquisition where sources of data are selected and data is collected whether manually or automated. Then data is sent to preparation phase which includes activities such as cleaning, filtering, classification, etc. Analysis consists of modeling. Visualization includes graphical or tabular presentation of

data. In interpretation phase, the meaning and the understanding of data are conveyed to users. The BDA process is shown in Fig. 1.

Fig.1. Big Data Analytics Process

Acquisition Data acquisition is carried out using both on board sensory and non-sensory data sources [16]. The advance of the technology has enabled data acquisition even when the meaning of the data is not fully understood, this is different from previous practices of business intelligence where the meaning of data was to be understood prior data storage[19] Acquisition includes data captured and obtained from numerous and diverse sources [10]. Data acquisition could be divided into three parts data collection, data transmission, data pre-processing [30]. In Higher education, data can be obtained from different sources including "social networking sites (like Facebook, Twitter, Blogs) , Course Management systems (CMS), Learning Management System (LMS) and physical world data like library usage" [15].

Preparation Big data include unstructured data which requires some work to be done before it becomes ready for analysis. Big data can also be thought of its distinct characteristics such the large volume and the variety. All these necessitate additional work, unlike structured, "ready-to- analyze" data in traditional databases. A number of researchers have addressed the data preparation irrespective of how differently they approached it. For example, Demchenko et al [2] used filtering, enrichment or classification. Assunção et al(24) stated data management phase which encompasses preprocessing, filtering, aggregation and transformation. Alternatively, Labrinidis and Jagadish[7] called it information extraction. Other researchers explained the purpose of this stage. For example, see [16], preparation stage includes data preprocessing and integration operations which are meant to enhance data quality. Integration deals with the differences brought by big data; the difference in data structure, and in semantics, and more importantly organizes data in a way this is understandable by computers [15].

Analysis The data is transformed into usable information. The analysis contains complex process; it demands linking separate data sets for extracting useful information [20]. Data

analysis uses "analytical methods or tools to inspect, transform, and model data to extract value"[30].

Visualization Gartner [33] defines visualization as "the illustration of information objects and their relationships on a display". The visualization stage is where information is presented in a form that is interpretable by users in order to leverage the insights into existing processes and to make informed decisions [20]. Its purpose is to" communicate information clearly and effectively in graphical means "[30], that is graspable even by non-technical staff [13]. In higher education context, It is indicated [8] that visualization facilitates understanding of analytics results for senior management or what is called C-level executives. Visualization of data in tabular and graphical manner can also help educators, students and administrators to easily access and interpret information [15].

Interpretation After visualization, it is necessary to be able to extract insights and understandings from the presented information in order to create values for the respective users. Analyzing big data has not much value if the results cannot be interpreted by users [15]. This prompts the availability of new skills, among them is inductive reasoning [13].

5.2. Process Performance Measures

Dumas et al[3] proposed what they called four process performance dimensions namely time, quality cost and flexibility. These terms were earlier elaborated in [28]. Process information that contains the various dimensions like those above can help address the limitations of conventional performance measures [18]. Other authors argue that processes represent the important drivers behind three success factors namely cost, quality and time [32]. Each of the four process performance dimensions stated above can be further elaborated into a number of performance measures or key performance indicators (KPIs)[3].

It has been observed that the four dimensions have been widely cited in the literature. It was also indicated that, like organizational performance, a process should also have clear performance indicators [17]. One popular widely-cited organizational measurement is Balanced Scorecard (BSC) proposed by Kaplan and Norton [6]. Studies of BSC applied in higher education have been spotted in the literature; however, elaborating this is beyond the scope of this paper.

Below is the explanation of the four process performance dimensions; time, quality, cost and flexibility:

Time Time is defined as the "The indefinite continued progress of existence and events in the past, present, and future regarded as a whole"(oxforddictionaries.com). Monteiro and de Oliveira [12] linked the time to software development process, describing it as a measure used to evaluate the period of time required for performing activities related to software development process. One popular performance measure, related to time, is cycle time. Cycle is the time spent to handle a situation or a "case" from the start to the end [3]. It is also referred to as the time it takes to transform inputs into outputs [14].The performance benefit here is to reduce cycle time. Dumas et al [3] expressed some items related to cycle time. Processing time is one of them. Processing time, also known as service time, is the time it takes the resources (e.g. staff and tools) to complete the process. Another performance item is waiting time. Waiting time is the process's idle time including the time spent for querying.

Quality Quality is shortly described as "how good or bad something is"(Cambridge Dictionary). Quality is also expressed as the "standard of something as measured against other things of a similar kind; the degree of excellence of something" (oxforddictionaries.com). Quality is referred to the measures to evaluating the ability of software process and products to answer the stakeholders' needs upon predefined acceptance criteria [12]. According to Dumas et al [3], quality can be viewed in two perspectives; internal quality and external quality. The internal quality of process is from process participants' side. It can mean the ability of the staff involved in managing the work of the process and the level of their process-related experience and how-well challenges of process -related work are perceived and worked on. External quality, on the other hand, is related to process clients, that is to say the users, and the degree of their satisfaction.

Cost The cost is the expense of whole process [14]; one example is operational cost [3]. It measures the financial resources used for software development process activities [12]. Generally, Process should be cost-effective and it should add value to organizations.

Flexibility Flexibility is the "ability to change to suit new conditions or situations" (oxfordlearnersdictionaries.co). It is commonly referred to the capability to respond to changes [3], which should, in business process perspective, accommodate the ability to respond to the changing circumstances in the organization; the ability to perform varied tasks and adopt to different situations and workloads. Flexibility, according to Dumas et al

[3], can also be viewed in the difference between runtime flexibility and build time flexibility.

Time, quality, cost and flexibility are process performance dimensions mainly discussed in the literature. Elaborating how they apply to BDA process is not in the scope of this paper

6. CONCLUSION AND FUTURE WORK

Several conclusions can be drawn from this paper. First, big data analytics is performed through a process which starts from data acquisition and selection of sources to interpretation. This process which handles big data is worth efforts, measuring its performance is one of them. Time, quality, cost, and flexibility are four process performance measures mainly discussed in the literature. These terms have been previously used to measure processes in other areas such as business and it is believed that they are applicable to BDA process. It is also possible that other measures do exist. The future work of this research will involve in further investigation of how existing process performance measures apply to BDA process.

7. REFERENCES

[1] Cates, J. E., Gill, S. S., & Zeituny, N. (2005). The Ladder of Business Intelligence (LOBI): *a framework for enterprise IT planning and architecture. International Journal of Business Information Systems*, 1(1-2), 220-238.

[2] Demchenko, Y., Gruengard, E., & Klous, S. (2014, December). *Instructional model for building effective Big Data curricula for online and campus education. In Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on* (pp. 935-941). IEEE.

[3] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management* (Vol. 1, p. 2). Heidelberg: Springer.

[4] Ergin, K. T. (2011). *Evaluation of automated business process optimization* (Master's thesis).

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

[6] Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: translating strategy into action*. Harvard Business Press.

[7] Labrinidis, A., & Jagadish, H. V. (2012). *Challenges and opportunities with big data. Proceedings of the VLDB Endowment*, 5(12), 2032-2033.

[8] Liebowitz, J. (2017). Thoughts on Recent Trends and Future Research Perspectives in Big Data and Analytics in Higher Education. *In Big Data and Learning Analytics in Higher Education* (pp. 7-17). Springer International Publishing.

[9] Lönnqvist, A., & Pirttimäki, V. (2006). *The measurement of business intelligence. Information systems management*, 23(1), 32.

[10] Mello, R., Leite, L. R., & Martins, R. A. (2014, January). Is Big Data the Next Big Thing in Performance Measurement Systems?. In *IIE Annual Conference. Proceedings* (p. 1837). Institute of Industrial Engineers-Publisher..

[11] Miskon, S., Bandara, W., Gable, G., & Fielt, E. (2011, November). Success and failure factors of shared services: An IS literature analysis. *In Research and Innovation in Information Systems (ICRIIS), 2011 International Conference on* (pp. 1-6). IEEE.

[12] Monteiro, L. F. S., & de Oliveira, K. M. (2010). Defining a catalog of indicators to support process performance analysis. *Journal of Software Maintenance and Evolution: Research and Practice.*

[13] Morabito, V. (2015). Big data and analytics. *Strategic and organisational impacts.*

[14] Ohlsson, J., Han, S., Johannesson, P., & Rusu, L. (2014, June). Developing a Method for Prioritizing Business Process Improvement Initiatives. In PACIS (p. 342).

[15] Tulasi, B. (2013). Significance of Big Data and analytics in higher education. *International Journal of Computer Applications*, *68*(14).

[16] ur Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*(6), 917-928.

[17] Van Looy, A., & Shafagatova, A. (2016). Business process performance measurement: a structured literature review of indicators, measures and metrics. *SpringerPlus*, *5*(1), 1797.

[18] Vukšić, V. B., Glavan, L. M., & Suša, D. (2015). The Role of Process Performance Measurement in BPM Adoption Outcomes in Croatia. *Economic and Business Review*, *17*(1), 117-143.

[19] Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), 700-710.

[20] Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, *46*(5), 904-920.

[21] Wang, Y., Kung, L., & Byrd, T. A. (2016). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*.

[22] Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, *50*, 408-413.

[23] Chandio, A. A., Tziritas, N., & Xu, C. Z. (2015). Big-data processing techniques and their challenges in transport domain. *ZTE Communications*, *1*(010).

[24] 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.

[25] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*, 70.

[26] Glavan, L. (2011). Understanding process performance measurement systems. *Business Systems Research*, *2*(2), 25-38.

[27] Zairi, M. (1997). Business process management: a boundaryless approach to modern competitiveness. Business Process Management Journal, 3(1), 64-80.

[28] Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: a literature review and research agenda. *International journal of operations & production management*, *15*(4), 80-116.

[29] Nightingale, D. (2005). Metrics and Performance Measurement System for the Lean Enterprise.

[30] Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, *2*, 652-687.

[31] Hussain, M., Al-Mourad, M. B., & Mathew, S. S. (2016, March). Collect, Scope, and Verify Big Data--A Framework for Institution Accreditation. In*2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)* (pp. 187-192). IEEE.

[32] Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. *IT Professional*, *15*(6), 29-35.

[33] "Visualization." *Gartner IT Glossary (*Accessed 01 Mar. 2017). From: http://www.gartner.com/it-glossary/visualization/

How to cite this article:
Ali I M, Jusoh Y Y, Abdullah R, Nor R N H. A review on the performance measurement of big data analytics process. J. Fundam. Appl. Sci., 2018, *10(2S), 103-115.*