

MEASURING THE PERFORMANCE OF BIG DATA ANALYTICS PROCESS

ISMAIL MOHAMED ALI ¹, YUSMADI YAH JUSOH ², RUSLI ABDULLAH ³, ROZI NOR HAIZAN NOR ⁴, AND LILLY SURIANI AFFENDEY⁵

*Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia*

¹*Ismail15@simad.edu.so*, ²*yusmadi@upm.edu.my*, ³*rusli@upm.edu.my*, ⁴*rozinor@upm.edu.my*,

⁵*lilly@upm.edu.my*

ABSTRACT

Big data analytics (BDA) in a process perspective has major benefits towards a better outcome, thereby satisfied customers and evidence-based practices. The aim of BDA is to examine and analyze raw data and to derive and extract actionable insights from it. BDA involves data and tools for processing and analyzing, and the process which data is handled and managed. BDA process is the end-to-end process which consists of phases named as data acquisition, data preparation (integration and preprocessing), data analysis, visualizations and interpretation. The performance of big data analytics is not merely dependent on having quality data input, but also on performance of the process which the data goes through from acquisition to visualization and interpretation. Measuring the process performance has the benefit of identifying problems and launching corrective actions before these problems deteriorate. The aim of this paper is to present the evaluation for BDA process performance. In view of that, the study identifies the measures, metrics, and indicators for each phase of the BDA process. A subject-matter expert review and a pilot study were conducted, and the results obtained were reported in this paper.

Keywords — *BDA Process, Process Performance, Big Data, Measures, Performance Measurement*

1. INTRODUCTION

Much has been said about the promises and potential benefits of big data. However, many challenges are still surrounding big data such as data challenges (relating to the defining characteristics such as volume, variety, and velocity etc.), management challenges (like privacy issues, security issues, governance and ethical considerations) and process challenges (concerning with how to capture, integrate, transform, and analyze data, and convey the results) [1].

As data goes through BDA process, there are several issues encountered. Lack of data provenance is one of them [1]. Having information about data upon its origins and carrying this information throughout the process is very useful. The reason is that processing errors, inconsistencies and missing information can be traced back and fixed accordingly. These problems, if not addressed, can make the subsequent analytics phases useless, and can restrict the speed to capture and store data, and the ability to extract meaningful information out of it. Heterogeneity, lack of structure, error-handling, privacy, timeliness,

provenance, and visualization also the issues that can exist through the BDA process from data acquisition to interpretation [2].

Looking at big data analytics in data quality perspective is a major focus in its still-emerging literature. However, the quality of big data analytics is not merely dependent on the data, but also on the process in which the data is collected and the way data is processed [3]. Insights revealing the existing quality issues are of important considerations. This signifies the need for measuring the performance of big data analytics, which was minimally discussed. It doesn't mean only the performance of a certain tool such as apache Hadoop and NoSQL database [4], [5], [6], but also performance of entire BDA process. Measuring process performance, as shown by other established process literature (like that in a business process), has the benefit of identifying problems and launching corrective actions before these problems deteriorate [7]. Problems can be deterioration in performance, failure to meet requirement or an opportunity for improvement.

Performance is not absolute. Instead it comes as a result of driving factors. Therefore, there is a need for measuring results and as well as the drivers of

the results. It is interesting that some performance frameworks were employing this type of performance measurement. It is the precondition for analyzing and improving processes [8].

This paper is organized as follows. Section II covers on related work of the BDA process, existing performance models and framework and performance measures. Section III includes the methodology of the pilot study conducted. In section IV, the initial model of performance measurement for the BDA process is presented. Section V shows the results and discussion of the pilot study and section VI includes the revise and propose model of this study. Finally, our work of this paper is summarized in the last section.

2. RELATED WORK

2.1. Background

Measuring the performance of the BDA process gives benefits such as better outcomes and satisfied customers. Performance measurement is said to be the process of quantifying efficiency and effectiveness of an action or a process. The importance of performance measurement is much elucidated by the old statements of “If you can’t measure it you can’t manage it” or “if you can’t measure it, you can’t improve it”. Obviously, identifying problems and performance gaps and proposing solutions accordingly have superlative importance over moving on action not guided by problem-related information and knowledge.

In context of big data, the unprecedented growth of data size echoed by the speed requirement imposes the need for performance measurement. Examples include the performance of data processing time, the performance of data transmission time (over the network) and the performance required when presenting results to users [9], as well as data acquisition time. This illuminates the point that the BDA process requires performance measurement whether it is internal performance (time, cost, resource utilization), usually referred to as efficiency, or external performance which is related to effectiveness. The explanation here agrees to the definition of process performance which involves process efficiency (i.e. productivity), and the effectiveness and compliance of the process [10].

Another point that deserves clarification is what is being captured in the performance measurement process. What to be captured is metadata, or “data about data”, about the performance of BDA process, in order to uncover performance

deficiencies and weaknesses and to discover opportunities for further improvement. A motivation is that such performance focus may lower the failure rate of big data application projects which is reportedly [11], 50% higher than IT projects failure.

It is common tradition that IS research to be based on existing IS theory or to be developed a new one, and researches in BDA should respect this common tradition. However, arguments are arisen on how existing IS theories fit into BDA research either being theory-driven or process-driven [12]. It was argued that one of these two approaches is not sufficient without the other. Therefore, lightweight theory has been suggested throughout the stages of BDA process. This is, at least, to addresses the epistemological challenges surrounding BDA research.

Shifting the focus to performance measurement side, an outstanding IS theory is a DeLone and McLean’s IS Success Model [61], [62]. This is because the keyword “success” roughly approximates the meaning of “performance”. Nevertheless, the applicability of any theory to be adopted is not determined by the semantical approximation of words only, but also by the relevance of constructs and contexts. Integrating DeLone and McLean model with the Structure-Process-Outcome (SPO) framework by [13] has much to say about this as process mediates performance contributing factors and resultant outcomes [14].

Generally, there are two approaches for process performance measurement. One of them is clean sheet approach where the performance measurement of process is started from the scratch. The other is to consider existing processes as starting point. Adapting existing approaches is more convenient whereas clean sheet is riskier but more rewarding [15].

2.2. Big Data Analytics Performance Measurement

Analyzing performance of big data applications and identifying the factors that influence their quality has been indicated to be one of the challenges existing in Big Data settings [16]. Existing performance evaluation studies are focused on comparing several Big Data frameworks such as Hadoop, Spark, Flink ([17], [18]). This involves conducting experiments using some datasets which are executed through selected big data frameworks, and then the corresponding performance metrics are extracted. According to

Veiga et al. [18], end-users rarely gain benefits from the research findings provided by these studies. There are two scenarios that can be into consideration. First, performance evaluation of such big data frameworks fails to consider that the field of big data is still emerging and tools in use today may not be relevant tomorrow. Second, such performance evaluation on BDA systems lacks the consideration of users' role in performance measurement. In fact, systems success is not only measured in terms of the performance of its features, but also the satisfaction of its users (e.g. users' level of satisfaction with reports, web sites, and support services [19]).

This gives rise to the necessity to discover measures that give the users the stake in performance measurement. It also illuminates the need for holistic solution that doesn't necessarily pertain to specific replaceable big data frameworks as those mentioned above, but for BDA systems in general. Therefore, using existing performance measures and metrics, this study investigates how to bridge this gap.

2.3. The BDA Process

The success of big data depends on the factors including people, technology and process. Unlike other traditional projects, big data requires different process approach which pertains to exploration and exploitation of data [20]. The need for process methodology has been a prominent discussion. Discussions center around whether to define new process which it is said to be possible, or to utilize, where applicable, the existing process methodologies such as Knowledge Discovery in Database (KDD) and Cross Industry Standard for Data Mining (CRISP-DM). The traditional ETL (extract, transform, and load) process is also another example of preexisting analytics processes. ETL process is criticized to be batch-oriented [21], a characteristic that complicates its applicability to fast analytics in the era of big data. Elsewhere, the applicability of agile methods was noted. Agile methods were formerly created for software development and suggested for data analysis (e.g big data analysis) as a better process guidance [22]. Agile method has been perceived to be suitable to business intelligence lifecycle, unlike to fast analytics lifecycle in era of big data, unless short-cycle agile approaches are employed. Short-cycle agile approach is concerned with faster and more flexible sprints [23].

Agile approach suggests that collaboration and interactions among team members over processes and tools have utmost importance. Coordinating

overall efforts of big data seemingly requires a process but in a different perspective. Therefore, some researchers stressed the need for big data team process methodology [24].

However, It is should also be noted the possible difference between big data analytics process (which data-intensive and recurrently executed), and big data projects, for example big data project lifecycle, which is temporary endeavor and that its success depends on maximum utilization of resources in a predefined timeframe. Moreover, the differentiating trait of BDA process is that it is a step-by-step process of understanding and doing BDA which may not necessarily pertain to specific process methodology. This also gives rise to the need for process standardization and integration [3], the ability to integrate processes, standardize tasks and data results and consequently achieve more benefits including minimization of costs and efforts of using big data.

Debates on BDA process can possibly be traced to the infancy of big data in general, and the future is promising as more attention is increasingly directed to big data from both the industry and academic settings.

The existence of the process is one thing and having standardized, well-defined process is another thing. The clarification of this point has pivotal importance since the researchers have vastly mentioned the notion of BDA process in the literature. For example, Ur Rehman et al. [25] presented, a big data analytics process which consists of six phases namely data collection, data preparation, modelling, evaluation, deployment and monitoring. According to them, big data analytics processes differ in terms of descriptive, prescriptive, and predictive analytic models. The said process was aimed at creation of learning models through predictive analytics. Erl et. al. [26] elaborated BDA lifecycle which consists of nine stages starting with business case evaluation and data identification stages followed by data acquisitions, and then proceeds to several stages that can be considered as data preparation. Their process, after preparation, contains other three stages, namely, data analysis, data visualization and utilization of the results. Elsewhere, the overall process of extracting values from big data is divided into five stages. The five stages are classified under two main sub-processes, data management and analytics [27]. Jagadish et al. [28] provided a process which is comprised of acquisition, information extraction and cleaning, data integration, modeling, analysis, interpretation, and

deployment. Table 1 summarizes the BDA processes from different authors.

TABLE1: Table 1 Analysis on Existing Processes

Author(s)	The BDA Process / Phases
[23]	Data acquisition, analyzes /visualize, model/ design, validate, deploy, among other phases for fast analytics and data science lifecycle.
[60]	Data collection and Registration, data filter/enrich/classification, data analytics/ modeling, data redelivery/ visualizations in big data lifecycle which manages data from a given data sources to consumer data analytics application.
[25]	Data collection, prepare data, model and evaluate, among other phases for big data analytics process.
[27]	Acquisition and recording, extraction/cleaning, integration/annotation, aggregation/aggregation/representation, modeling and analysis, interpretation.
[59]	Data collection, data analysis, data visualization
[58]	Discussed data acquisition, information extraction, data analysis the need for interpretation.
[57]	Mentioned some useful terms in overview of analytics workflow for big data, including are analysis, visualization, interpretation and others.
[56]	Acquisition, extraction, integration, analyzing, interpretation
[53]	Generation, acquisition, storage, analytics.

This research relies on big data analytics process with the following phases (refer to figure 1): data acquisition, data preparation, data analysis, visualization, and interpretation.

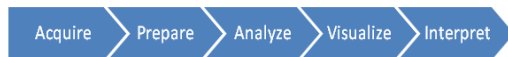


Figure 1 BDA Process (Ali et al., 2018)[29].

The phases of BDA process are described below:

Acquisition: Data Acquisition includes the selection of sources and collecting the data from diverse sources like online activities (such as tweets, retweets, web crawlers, customers' reviews, clickstreams, and sensors), log files, and data warehouses. The data is timely captured and sent to next phase for more calibration,

Preparation: Preparation phase involve activities such as data integration, pre-processing, and

cleansing. Data is checked for errors, outliers, missing, or noisy data. Then, the data are unified, dimensions are reduced, and features are extracted.

Analysis and Modeling: This phase is more on applying analytical statistical tools and methods on data to extract actionable information and business insights, thereby producing different types of analytical models descriptive, prescriptive, or predictive.

Visualization: In this phase, the result of the analysis and modeling phase are presented to users in a meaningful and understandable way; either in a tabular or graphical form or both. The visualization of results should be looked at users' perspective, as their satisfaction is a key success factor to any information systems, and analytics systems are no exception.

Interpretation: Results are interpreted and exploited by the users into their context such as operational optimization, decision making enhancement, or even creating new business models.

2.4. Existing Performance Models and Frameworks

The BDA analytics system involves two important components: data and a process. The data is handled through this process from acquisition to interpretation phase. There are two components of BDA highlighted by several studies.

Hybrid model presented by Serhani et al. [30] considers both data and process but in quality perspective. There are three process quality measures were used to evaluate the preprocessing stage and processing analysis stage: accuracy, throughput, and response time. Other data quality measures were also mentioned in their study. The **performance Analysis model** for big data applications was presented by Villalpando et al. [16]. Their study examined big data applications performance using ISO 25010 software quality concepts, namely performance efficiency and reliability. Process performance dimensions known as **the Devil's Quadrangle** is also a framework which is mainly in process redesign. The framework brings together four process performance dimensions namely time, quality, cost, and flexibility [8], [15]. The framework combines financial measures (cost) and other non-financial measures such as time, quality, and flexibility.

Going far into performance measurement literature leads to more process performance measurement frameworks and models which are mainly in business process and manufacturing perspectives. **TOPP system framework** is one of

them. TOPP system uses efficiency, effectiveness, changeability (or ability to change as some call it) as three performance measures [31]. Interestingly, TOPP System is comparable with the devil's quadrangle from the point that cost, and time can be categorized under efficiency as a measurable concept. TOPP systems' effectiveness is defined as customer satisfaction. The quality dimension in devil' quadrangle (comparatively) is aimed at the satisfaction of both customers and process participants (the staff). The similarity is also much comprehensible between changeability and flexibility. This trade-off can be made, although the two frameworks vary in their application. TOPP uses questionnaire to evaluate the performance, not only for a process but also for the entire enterprise in manufacturing areas. The summary of existing frameworks and models is provided in Table2.

Table2: Existing Models and Frameworks

Author (s)	Model/ Framework	Measures/Metrics
[30]	A Hybrid model for assessing quality of big data value chain	Accuracy, throughput, and response time
[16]	Performance analysis model for Big Data Applications	Performance efficiency (time behavior) resource utilization, and, capacity), reliability (maturity, Availability, fault tolerance, and recoverability)
[8], [15]	Process performance dimensions (the Devil's quadrangle)	Time, quality, cost, and flexibility
[31]	TOPP System	efficiency, effectiveness, changeability
[47]	process performance model	overall, cost, quality, service and time as four classes of indicators with three stakeholders; namely customers, operators, and management
[46]	Process Performance Evaluation	time, quality, service,

	Methodology	efficiency, cost, and importance
[45]	A model of Process Performance	customer satisfaction, product development time and the cost for product design and manufacturing

2.5. Performance Measures

Having considered a BDA process as measurable entity, the next step is to identify performance measures that apply to it. "A performance measure is defined as a metric used to quantify the efficiency and effectiveness of an action" [32]. Time-related measures such as cycle time, response time, latency, and speed are the dominant process performance measures. Capacity, throughput and resource utilization were observed in the literature. The time behavior and resource utilization measures are classified under efficiency. Output-related measures such as user's satisfaction are classified under effectiveness. Other contributing factors also include technology, competence, compliance, and staff's working conditions. The following are the description of performance measures:

Efficiency Measures: Efficiency, as a measurable concept, is an internal process performance measure that shows how well the process transforms inputs into outputs. It includes "resource optimization (mainly cost and time) along with maximum waste reduction" [33]. Provided in Table 3 is a group of process performance measures classified as efficiency measures.

Table 3 Efficiency Measures

Measures	Definitions	Author(s)
Time	Time is a measure used to evaluate the time required for performing activities related to the process	[10]
Cost	Cost is the expense of the whole process. It is a measure related to evaluating financial resources	[44], [10]

	applied to the activities of the process (software process)	
Capacity	The maximum number of simultaneous connections and/or processes	[30],[16], [35]
Response time	Total time to complete the processing each record. The time interval between when BDA task submitted to start processing and it is Started. The time needed to complete user request	[30], [16], [42], [9], [35], [43]
Throughput	Number of records or requests completed processing over a period of time.	[30], [42], [9], [35], [43]
Latency	The time to wait for receiving the results of processed data	[30]
Processing time	Duration from started BDA process-related task until that task is completed	[16]
Accuracy	Error's rate resulted after data processing.	[30], [43]
Resource utilization	How resources such as processing power, storage, people and the money are utilized	[18],[16], [42]
Timeliness	Timeliness relates to situations where the results should be received immediately by the users. Timeliness also measured in data acquisition or data collection	[30],[43]

Effectiveness Measures: Effectiveness is an external process performance measures which shows the extent process achieves the needs of various stakeholders [33]. Process effectiveness measures the degree to which the preferred performance of the process, such as certain outcomes or results, is achieved [34]. Therefore,

effectiveness emphasizes on whether process is achieving sufficient output [8]. In information systems' perspective, it means the impact of information provided on assisting users to perform their work [35].

Flexibility Measures: Flexibility is “the ability to react to changes”. Flexibility can proudly divide into run-time and build-time flexibility [15], [36]. Also, customization and modifiability in order to meet future changes [34]. Adaptability, as a synonym for flexibility, is the degree to which analytics system can be adjusted to satisfy various needs in changing situations [37]. The flexibility in BDA process, as can be seen in Table 4, is represented in the ability to handle the increasing volumes of data, the ability to adjust to new needs and circumstances, and users' ability to view results for their preference.

Table 4: Flexibility Related Measures

Measures	Description
Input volume	The ability of BDA process to handle the large and the increasing volumes of inputs
Modifiability	Complexity of the process, ability to add new activities/sub processes or to incorporate new tools
Customizability	The users can choose their way to visualize the information (visualization phase) either in graphical or tabular form, on a computer screen or on hand-held devices.

2.6. Performance Contributing Factors

Processes, like BDA process, do not function individually. Instead, their performance depends on several factors including people with specific skills, policies and procedures that govern them, technology that enables them, and enabling work environment. The explanation of those factors provided in Table 5.

Table 5: Performance Contributing Factors

Factor	Definition	Author(s)
Technology effectiveness	Technology effectiveness is concerned with the suitability of technology to produce software which could be	[10], [52], [53]

	considered when evaluating process quality. Technological support enhances the efficiency of big data analytics process	
Competence	Big data analytics human knowledge and skills such as technical, business, relational, and business analytics. Also related to process performance management knowledge and skills: Possess knowledge, technical expertise, and communication skills	[50], [51]
Compliance	It refers to the level of which the standard process is actually implemented. This should be given a consideration as it can show the level of improvement (software process improvement). The ability to execute the process and use of the defined process	[54], [55]
Work Conditions	Work conditions related internal quality which can be discerned as process participant's satisfactions. The staffs are most productive if they have good working conditions that enable them a sustainable pace of progress with affordable workloads	[8],[48], [49]

3. THE RESEARCH METHODOLOGY

The methodology followed in this paper consists of three steps: reviewing the related work,

proposing the initial model, and conducting pilot study (expert review and survey).

Reviewing the existing and related literature was conducted to understand the background of study, highlighting its significance and its theoretical foundations. Then the focus has been put on discerning BDA process, examining process methodologies, correlate with nature of BDA, and ascertaining what constitutes BDA process in terms of process phases and descriptions of these phases. Afterwards, the literature review investigates existing performance measurement frameworks and models and presents those are related to the topic being addressed in this paper. Finally, as the result of a reviewed literature, a category of performance measures and their definitions were presented. This section is concluded with an illustration that outlines the identified performance measures, their corresponding metrics and indicators. Next, an initial model was proposed based on the related works.

A questionnaire was prepared and distributed to the experts and potential respondents that involved in BDA. The questionnaire was sent to four experts in the area of BDA, in order to conduct content validity test. Two experts had an industry experience. The others had academic backgrounds related to BDA. One of the experts was female and the rest were male respondents. Content validity test is subject assessment of the suitability of the research instrument by subject-matter experts. It ensures that all relevant contents are included and irrelevant contents are excluded. The experts were asked to rate the relevance of survey items related to the initial model which was also presented to them. The survey consisted of 49 items excluding demographic questions. The experts' feedback and comments recorded in a table, their responses were analyzed, and the survey items were added, deleted and modified accordingly, and they were finally reduced to 41 items.

For reliability test, this research employs composite reliability (CR) because it is an estimate of a construct's internal consistency and it takes into count that indicators have different loadings [38]. The following is the formula used for CR.

$$CR = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + (\sum \delta)}$$

Where:

λ = Factor loading

δ = Measurement error

The values for CR have the following interpretations. Values that are above than 0.70 are desirable for exploratory research. Values above

0.800 or 0.900 desirable in more advanced stages of research. Values below 0.600 indicate a lack of reliability (Nunnally and Bernstein 1994) cited in [38].

A survey for pilot study was conducted with BDA practitioners. The data collection made use of online Google Form through email and printed copy of the survey physically delivered to the respondents. A total of 22 fully filled-in surveys were returned. All responses were combined in Google Form and then they were converted to Excel file. Afterwards, the necessary changes were made; for example replacing all "Strongly Agree" with "5" (This is due that Google Form stores the labels of responses not the corresponding numbers). Finally the Excel file was loaded into Smart PLS software and reliability test of the constructs, as

explained below in the results section, was performed.

4. THE PERFORMANCE MEASURES FOR BDA PROCESS

The performance measures denote the basic components that participate in the model development. Every measure contains specific elements called metrics. In addition, indicators provide more details about how the performance measurement applies to the selected domain. The details of the BDA process performance measures, metrics, and indicators are shown in Figure 2.

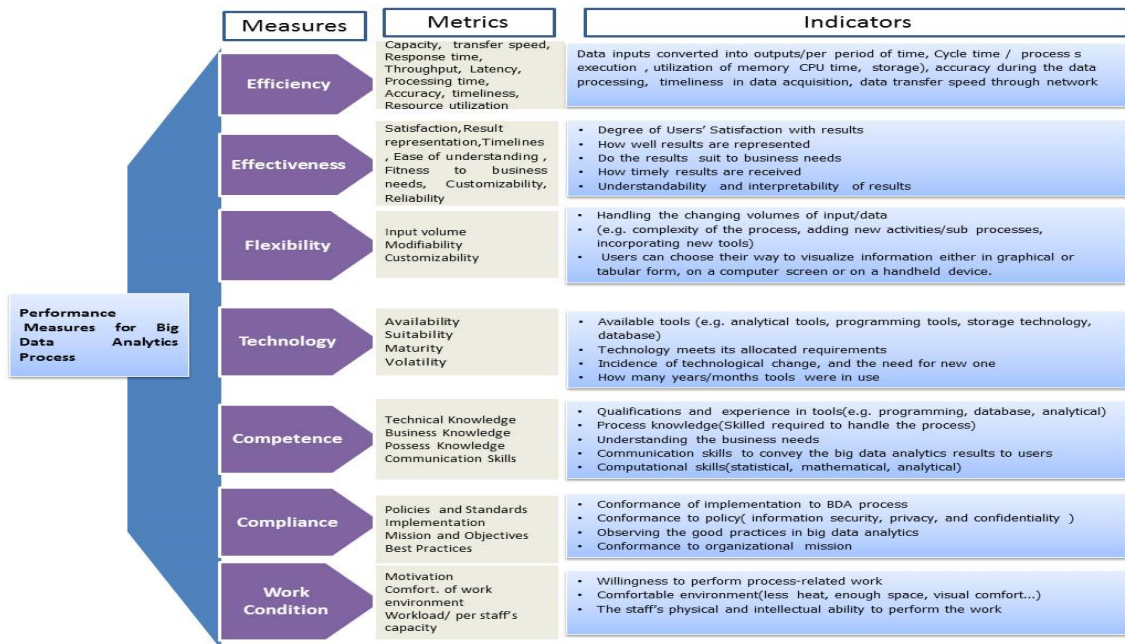


Figure 2 Measures, Metrics and Indicators for Measuring the Performance of BDA Process

5. RESULTS AND DISCUSSION

5.1. Content validity Test

As mentioned in the methodology section, expert view was conducted with four experts. Their feedback was summarized in Table 6.

Table 6 Experts' Feedback

Construct	Relevancy (Rated by the Experts)
Efficiency	There are 3 items for this construct. Two times: Resource utilization and Time were rated to relevant. The third one was productivity, and based on the experts' feedback, it was removed. More elaboration was made on time behavior and resource utilization, and a result, specific metrics of each were identified
Effectiveness	This construct contained 5 items, 4 times, relating to users' satisfaction,

	timeliness, fitness to the needs, understandability and interpretability of output, were rated relevant, were rated relevant. One item, accuracy was seen irrelevant. Some amendments were made based on the experts' comments and literature findings
Flexibility	3 items, namely the ability to handle the changing volumes of data, modifiability, customizability deemed relevant based on the experts' ratings. One item which is the ability to handle the structural variations (structured data vs. unstructured data, and different data types, was removed.
Technology	Availability, suitability, volatility, and maturity were rated relevant. One item was excluded from the 5, original items.
Competence	There were 6 items for competence and all of them were seen necessary and relevant (refer to Figure 3).
Work Conditions	5 items were presented to the experts, 3 were endorsed by them and 2 were excluded.
Compliance	The construct had 5 items. Based on the experts' judgment, all of them remained relevant.

5.2. The reliability test of the model constructs

Reliability can be defined that the measurements are free from error and, therefore yield consistent results. After running the pilot study data of this research in Smart PLS 3.0, it was found the CR test yields acceptable results. The generated results range from 0.681 to 0.927. Table 5 shows the CR test of pilot study for this research.

Table 7 Reliability Test Results

Constructs	Composite Reliability
Effectiveness	0.801
Flexibility	0.681
Technology	0.84
Competence	0.927
Working Conditions	0.863
Compliance	0.848
Efficiency	0.746

The Reliability Test yielded comfortable scores based on the suggested threshold. One exception is

that the value of 0.681 for flexibility which is less than 0.70. However, there are arguments that range is still acceptable [39]. The value for competence (0.927) is seemingly high but also in an acceptable range. Very high scores that are above 0.950 are said to be suspicious than those in the middle alpha ranges as they may involve common method bias [40].

6. PERFORMANCE MEASUREMENT MODEL FOR BDA PROCESS

The initial model represented in Figure 3 consists of five variables selected among the seven factors mentioned earlier in this paper. The selection is based on the expert's comments, the survey results, and the rationale of how different factors can join together and meaningfully interact. The selected factors are technology, competency, work conditions, efficiency and effectiveness. The essence of these factors lies in their relationship. The relationship is represented in the following sequence and order. First, put the required competence to perform BDA tasks and activities in a place. Then equip the staff, given their competence, with the required tools to perform the job. Make sure that the staff is happy and motivated. Then come down to assess performance using efficiency measures which focuses on how BDA process is internally functioning. Finally evaluate user's satisfaction based on effectiveness measures.

Another perspective of the model components is that they are divided into two parts: Global and local measures. Global measures are technology, competence and work conditions. The global measures, or factors in statistical terms, have holistic effects on all activities in BDA process. Think about the need for technological tools, for example, in data acquisition, data preparation, data analysis, visualization and interpretation. Also having competence and good working conditions are indispensable for every tiny work being performed amid the chain of big data analytics. Local measures are efficiency and effectiveness, these measures are not a holistic as global measures but on other hand, they are measures for individual BDA process phases.

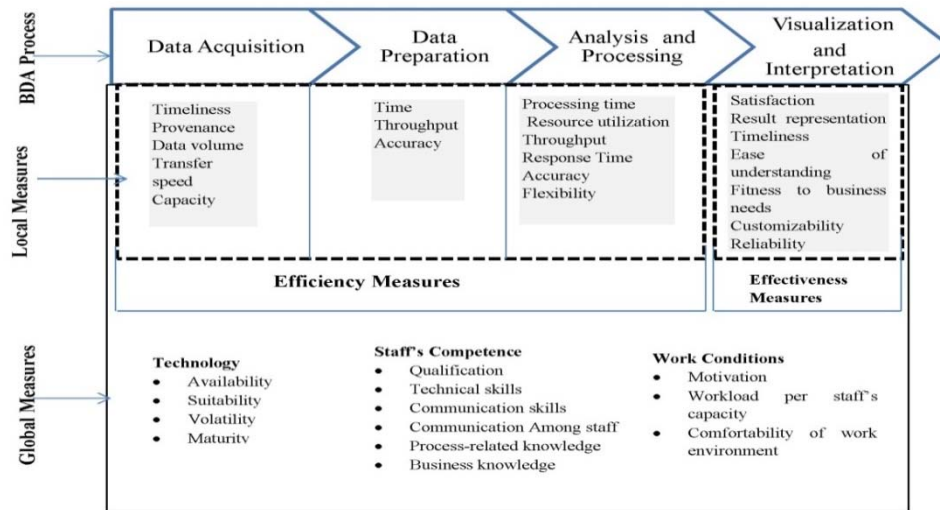


Figure 3 Performance Measurement Model for BDA Process

7. CONCLUSION AND FUTURE WORK

Researches on big data have many directions including big data analytics (BDA), big data infrastructure, and transformation and impact [41]. BDA, in turn, involves the data, the tools and techniques for data processing and analytics, and most importantly, the process which connects all things together.

There are different process methodologies for different artifacts of data science, big data as an advance in data science, has its own processes and ways of organizing things. Nevertheless, the step-by-step process of performing big data analytics and process of managing and coordinating big data project into two different concepts. The first is a cycle of durable and repeatedly executed BDA process for extracting knowledge and insights. The second is a temporary endeavor for coordinating teams, time and resources to address an issue of a concern.

BDA process, as mentioned above, produces the knowledge and the insights that businesses need, the efforts to improve and optimize this process have arguably a sound justification. Performance measurement is an important topic to embark on. Looking at existing performance models and frameworks and other performance literature provided a number of performance measures and performance contributing factors. The identified performance measures as well as the contributing factors give rise to the understanding that BDA process requires a specific set of skills to perform the BDA process related activities, technology that enhances BDA process execution, supportive work environment, and performance measurement ways

for spotting performance deficiencies and exploiting opportunities.

A reliability test conducted yields that the performance measures and factors have sound reliability scores. The future work of this research will be focusing on evaluation of the proposed model based on a larger sample of BDA community.

REFERENCES

- [1] 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.
- [2] Alguliyev, R. M., Gasimova, R. T., & Abbasli, R. N. (2017). The obstacles in Big data process. *International Journal of Modern Education and Computer Science*, 9(3), 28.
- [3] Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338-345.
- [4] Joshi, S. B. (2012, April). Apache hadoop performance-tuning methodologies and best practices. In *Proceedings of the 3rd acm/spec international conference on performance engineering* (pp. 241-242). ACM.
- [5] Dede, E., Govindaraju, M., Gunter, D., Canon, R. S., & Ramakrishnan, L. (2013, June). Performance evaluation of a mongodb and hadoop platform for scientific data analysis. In *Proceedings of the 4th ACM workshop on Scientific cloud computing* (pp. 13-20). ACM.
- [6] Gandini, A., Gribaudo, M., Knottenbelt, W. J., Osman, R., & Piazzolla, P. (2014, September).

- Performance evaluation of NoSQL databases. In *European Workshop on Performance Engineering* (pp. 16-29). Springer, Cham.
- [7] Kueng, P. (2000). Process performance measurement system: a tool to support process-based organizations. *Total quality management, 11*(1), 67-85.
- [8] Leyer, M., Heckl, D., & Moormann, J. (2015). Process performance measurement. In *Handbook on Business Process Management 2* (pp. 227-241). Springer, Berlin, Heidelberg.
- [9] Liu, Z. (2014, July). Research of performance test technology for big data applications. In *Information and Automation (ICIA), 2014 IEEE International Conference on* (pp. 53-58). IEEE.
- [10] Monteiro, L. F. S., & de Oliveira, K. M. (2011). Defining a catalog of indicators to support process performance analysis. *Journal of Software Maintenance and Evolution: Research and Practice, 23*(6), 395-422.
- [11] Lai, S. T., & Leu, F. Y. (2017, November). A Critical Quality Measurement Model for Managing and Controlling Big Data Project Risks. In *International Conference on Broadband and Wireless Computing, Communication and Applications* (pp. 777-787). Springer, Cham.
- [12] Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data, 4*(1), 19.
- [13] Cornford, T., Doukidis, G. I., & Forster, D. (1994). Experience with a structure, process and outcome framework for evaluating an information system. *Omega, 22*(5), 491-504.
- [14] Iannacci, F., & Cornford, T. (2018). Unravelling causal and temporal influences underpinning monitoring systems success: A typological approach. *Information Systems Journal, 28*(2), 384-407.
- [15] Dumas, M., Van der Aalst, W. M., & Ter Hofstede, A. H. (2005). *Process-aware information systems: bridging people and software through process technology*. John Wiley & Sons.
- [16] Villalpando, L. E. B., April, A., & Abran, A. (2014). Performance analysis model for big data applications in cloud computing. *Journal of Cloud Computing, 3*(1), 19.
- [17] Veiga, J., Expósito, R. R., Pardo, X. C., Taboada, G. L., & Tourifio, J. (2016, December). Performance evaluation of big data frameworks for large-scale data analytics. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 424-431). IEEE.
- [18] Veiga, J., Expósito, R. R., & Touriño, J. (2018). Performance Evaluation of Big Data Analysis. *Encyclopedia of Big Data Technologies*, 1-6.
- [19] DeLone, W. H., & McLean, E. R. (2016). Information systems success measurement. *Foundations and Trends® in Information Systems, 2*(1), 1-116.
- [20] Koronios, A., Gao, J., & Selle, S. (2014). Big Data Project Success-A Meta Analysis. In *PACIS* (p. 376).
- [21] Chen, H. M., Kazman, R., & Haziyevev, S. (2016, January). Agile big data analytics development: An architecture-centric approach. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on* (pp. 5378-5387). IEEE.
- [22] Ponsard, C., Touzani, M., & Majchrowski, A. (2017). Combining Process Guidance and Industrial Feedback for Successfully Deploying Big Data Projects. *Open Journal of Big Data (OJBD), 3*(1), 26-41.
- [23] 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.
- [24] Saltz, J. S. (2015, October). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In *Big Data (Big Data), 2015 IEEE International Conference on* (pp. 2066-2071). IEEE.
- [25] 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.
- [26] Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals: Concepts, Drivers & Techniques*. Prentice Hall Press.
- [27] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.
- [28] Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM, 57*(7), 86-94.
- [29] Ali, I. M., Jusoh, Y. Y., Abdullah, R., & Nor, R. N. H. (2018). A review on the performance

- measurement of big data analytics process. *Journal of Fundamental and Applied Sciences*, 10(2S), 103-115.
- [30] Serhani, M. A., El Kassabi, H. T., Taleb, I., & Nujum, A. (2016, June). An hybrid approach to quality evaluation across big data value chain. In *2016 IEEE International Congress on Big Data (BigData Congress)* (pp. 418-425). IEEE.
- [31] Browne, J., Devlin, J., Rolstadas, A., & Andersen, B. (1997). Performance measurement: the ENAPS approach. *International Journal of Business Transformation*, 1, 73-84.
- [32] 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.
- [33] "Process effectiveness and efficiency" form <http://isoconsultantpune.com/iatf-169492016-process-effectiveness-and-efficiency>
- [34] Kahrović, E. (2013). *Design and Implementation of Business Process Performance Measurement System*. Retrieved from: <http://facta.junis.ni.ac.rs/eao/eao201303/eao201303-06.pdf>
- [35] Heo, J., & Haan, I. (2000). Measuring Performance of Information Systems in Evolving Computing Environments: An Empirical Investigation. *PACIS 2000 Proceedings*, 25.
- [36] Jansen-Vullers, M. H., Kleingeld, P. A. M., & Netjes, M. (2008). Quantifying the performance of workflows. *Information Systems Management*, 25(4), 332-343.
- [37] Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, 55(17), 5011-5026.
- [38] Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information technology theory and application*, 11(2), 5-40.
- [39] Wong, K. K. K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin*, 24(1), 1-32.
- [40] Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information systems*, 13(1), 24.
- [41] Goes, P. B. (2014). Editor's comments: big data and IS research. *Mis Quarterly*, 38(3), iii-viii.
- [42] Brunnert, A., Vögele, C., Danciu, A., Pfaff, M., Mayer, M., & Krcmar, H. (2014). Performance management work. *Wirtschaftsinformatik*, 56(3), 197-199.
- [43] Onyeabor, G. A., & Ta'a, A. (2018, June). A Model for Addressing Quality Issues in Big Data. In *International Conference of Reliable Information and Communication Technology* (pp. 65-73). Springer, Cham.
- [44] Ohlsson, J., Han, S., Johannesson, P., & Rusu, L. (2014, June). Developing a Method for Prioritizing Business Process Improvement Initiatives. In *PACIS* (p. 342).
- [45] Syamil, A., Doll, W. J., & Apigian, C. H. (2004). Process performance in product development: measures and impacts. *European Journal of Innovation Management*, 7(3), 205-217.
- [46] Tan, W., Shen, W., Xu, L., Zhou, B., & Li, L. (2008). A business process intelligence system for enterprise process performance management. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(6), 745-756.
- [47] Longo, A., & Motta, G. (2005, September). Design processes for sustainable performances: a model and a method. In *International Conference on Business Process Management* (pp. 399-407). Springer, Berlin, Heidelberg.
- [48] Górný, A. (2017). The use of working environment factors as criteria in assessing the capacity to carry out processes. In *MATEC Web of Conferences* (Vol. 94, p. 06011). EDP Sciences.
- [49] Mahnic, V., & Vrana, I. (2007). Using stakeholder driven process performance measurement for monitoring the performance of a Scrum based software development process. *Electrotechnical Review*, 74(5), 241-247.
- [50] Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and E-Business Management*. <https://doi.org/10.1007/s10257-017-0362-y>
- [51] Blasini, J., & Leist, S. (2013). Success factors in process performance management. *Business Process Management Journal*, 19(3), 477-495.
- [52] Sheng, J., Amankwah-Amoah, J., & Wang, X. (2018). Technology in the 21st century: New

- challenges and opportunities. *Technological Forecasting and Social Change*.
- [53] 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.
- [54] Unterkalmsteiner, M., Gorschek, T., Islam, A. K. M., Cheng, C. K., Permadi, R. B., & Feldt, R. (2014). A conceptual framework for SPI evaluation. *Journal of Software: Evolution and Process*, 26(2), 251-279.
- [55] Statz, J. (2005). Measurement for process improvement. *Practical Software and Systems Measurement*, 1-17.
- [56] Tulasi, B. (2013). Significance of Big Data and analytics in higher education. *International Journal of Computer Applications*, 68(14).
- [57] 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.
- [58] Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- [59] Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, 46(5), 904-920.
- [60] 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.
- [61] DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, 3(1), 60-95.
- [62] DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.