International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

A System for Evaluating Big Data Analytics Performance

Ismail Mohamed, Yusmadi Yah Jusoh, Rusli Abdullah, Rozi Nor Haizan Nor, Lilly Suriani Affendey

Abstract: ***Despite of advances in systems and practices, performancemeasurement remains instrumental in helping organization discoverexistingproblems and propose viable solutions. On big data Analytics (BDA), more efforts are being focused, butperformance side is still a room for improvement. This paper discusses the development and implementation of performance measurement prototype system for big data systems. With this system, organizations can continually assess the performance gains and setbacks of their big data systems. The systems were developed based on measures and metrics retrieved from the extant literature. Then it was evaluated through review of subject-matter experts and usability survey. The development process of the prototype and the results of the evaluation are presented in this paper.***

Keywords: BDA, Performance Measures, Evaluation, Prototype

I.INTRODUCTION

Literally, the word prototype roughly means "first or primitive form" which originates from the Greek word prototypos: proto "first" and typos "impression" (Blomkvist&Holmlid, (2011). Bringing it into information systems context goes to late1970s where prototype strategy was described tobe an initial and usually highly simplified prototype version of the whole system is designed, implemented, tested and put into operation (Bally et al.,1977). Then prototype concept gradually evolved and familiarized itself into information systems and software engineering. A significant study in late 90s explores the prototyping approaches in information systems where taxonomy of prototyping concepts was presented (Beynon-Davies et al., 1999). The study reportedthe difference between prototype and prototyping. The first is said to be a working version of information system, and the second is a development process as alternative to preexisting systems development lifecycle. The prototyping approach is appropriate for scenarios in which the actual user requirements are not clear or standing to reason (Spitzer et al., 2018).

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Ismail Mohamed, Faculty of Computing, SIMAD Uinversity
Yusmadi Yah Jusoh, Faculty of Computer Science and Information Technology,Universiti Putra Malaysia
Rusli Abdullah, Faculty of Computer Science and Information Technology,Universiti Putra Malaysia
Rozi Nor Haizan Nor, Faculty of Computer Science and Information Technology,Universiti Putra Malaysia
Lilly Suriani, Faculty of Computer Science and Information Technology,Universiti Putra Malaysia

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

The term, prototype,was comprehensively definedas an early version of a software system that is designated to demonstrate concepts, try out design options,
and find out more about the problem and its possible solutions(Sommerville, 2011).

From above explanation, it comes down to the recognitionthat process of developing performance measurement system can be regarded as prototyping and the resultant work product is as a prototype.

Having mentioned performance measurement, do we know how it can be conceptualized into BDA settings?

Performance measurement is the process of quantifying the efficiency and effectiveness of action (Neely et al., 1995). Performance consists of the front-end and back-end performance properties (Liu, 2014). Big data system' performance measurement is more concentrated on back-end-performance, that is to say the performance of system's functions, but performance of the front-end performance which focuses on user experience and their satisfaction also has a profound importance. This suggests the need for considering both efficiency and effectiveness in measuring the performance of big data systems. Efficiency is related to relate to system's availability and performance over efforts, whereby effectiveness is concerned with the impact of information on assisting users on performing their work (Heo&Haan, 2000).Therefore resource utilization, time related metrics and capacity (Villalpando et al, 2014), (Brunnert et al., 2014), as well as throughput, response time, latency (Onyeabor&Ta'a, 2018)can be regarded as regarded performance measures of big data systemswhereas measures, such as satisfaction, timeliness usefulness and result representation are used to scrutinize system's success from user perspective. It means individual and organizational objectives are attained along with system's objectives. BDA being viewed as process, it encompasses both system's perspective where data is being acquired, pre-processed and integrated, and analyzed, and from user's perspective, where results are presented and interpreted in business context. The two ends are where the performance of BDA should be observed. Therefore, with existing challenges, we believe performance measurement of BDA process is rightly emphasized. In addition, performance is not always assumed as standalone entity; it is determined by the capability of the system to which it belongs. In big data, there are a number of factors that partake in enhancing big data analytical capability, among them are human capital (both technical skills and managerial skills) and technology can handle the volume and speed of big data (Mikalef, et al., 2017).

Lastly, this research strives putting all above concepts together meaningfully, does ittackle issues in big data is an important question ahead.

II.PERFORMANCE MEASURES

Peformance measures can be specified as a metric used to quantify the efficiency and/or effectiveness of an action(Neely et al., 1995).e measures used in this prototype are presented in Table 1.The identified measures are applied to BDA which is reprensented as a process. The process of BDA extends from data acquisition and data preparation phases, to data processing and analysis, visualization, and interpretation phases. The measures used in the system are categorized under constructs, namely Efficiency, effectiveness, technology, competency, and Working conditions.

Table.1 Measure for BDA Performance

No.	**Constucts**	**Measures**
1	Efficiency	Time, Capacity, Response time, Throughput, Processing time, Accuracy, Resource utilization, Timeliness, Flexibility , Provenance
2	Effectiveness	User satisfaction, Quality of data representation, Timeliness, Perceived Usefulness , Reliability, understandability
3	Technology	Availability, Suitability, Volatility, Maturity
4	Competency	Qualification, Technical skills, Communication skills, Process knowledge, Business knowledge

III.REQUIREMENT ANALYSIS

The prototype is developedbased on the proposed model for BDA process performance measurement. The model contains the performance measurement and the factors presented in Table 1. In thisprototype, evaluators (one in charge of evaluation of performance of BDA process) will collects performance evaluation input from different users who may not be present in one geographical location. Thus, Web application was considered to be most appropriate in such environment. Therefore, ASP.net has been chosen to develop the front end of the prototype and SQL database was used to create back end of the system. Active Server Pages (ASP) was formerly created by Microsoft. ASP.net is later, more robust version of ASP. The reason to choose ASP.net is its look and feel as Microsoft Windows, and that it is faster and incorporates more features.

The database of the system has been developed in SQL Server database which is relational database which is also developed by Microsoft. For this study, the data entered by users go to SQL database which resides in the company' storage infrastructure. Also, Measures, user's particulars, and authentication information are all stored in the said database.

The Screen View of the System

Figure 5 portrays the prototypes interface. After having logged into the system, users arrive at thier homepage where they perform the performance evaluation based on metrics provided. The results will be ready as users submet thier evaluation. Manager(evaluator) also inter into the system and view the results both graphicallly or tabular manner. The cycle will be repeated as new perforamnce measureemnt for BDA is needed.

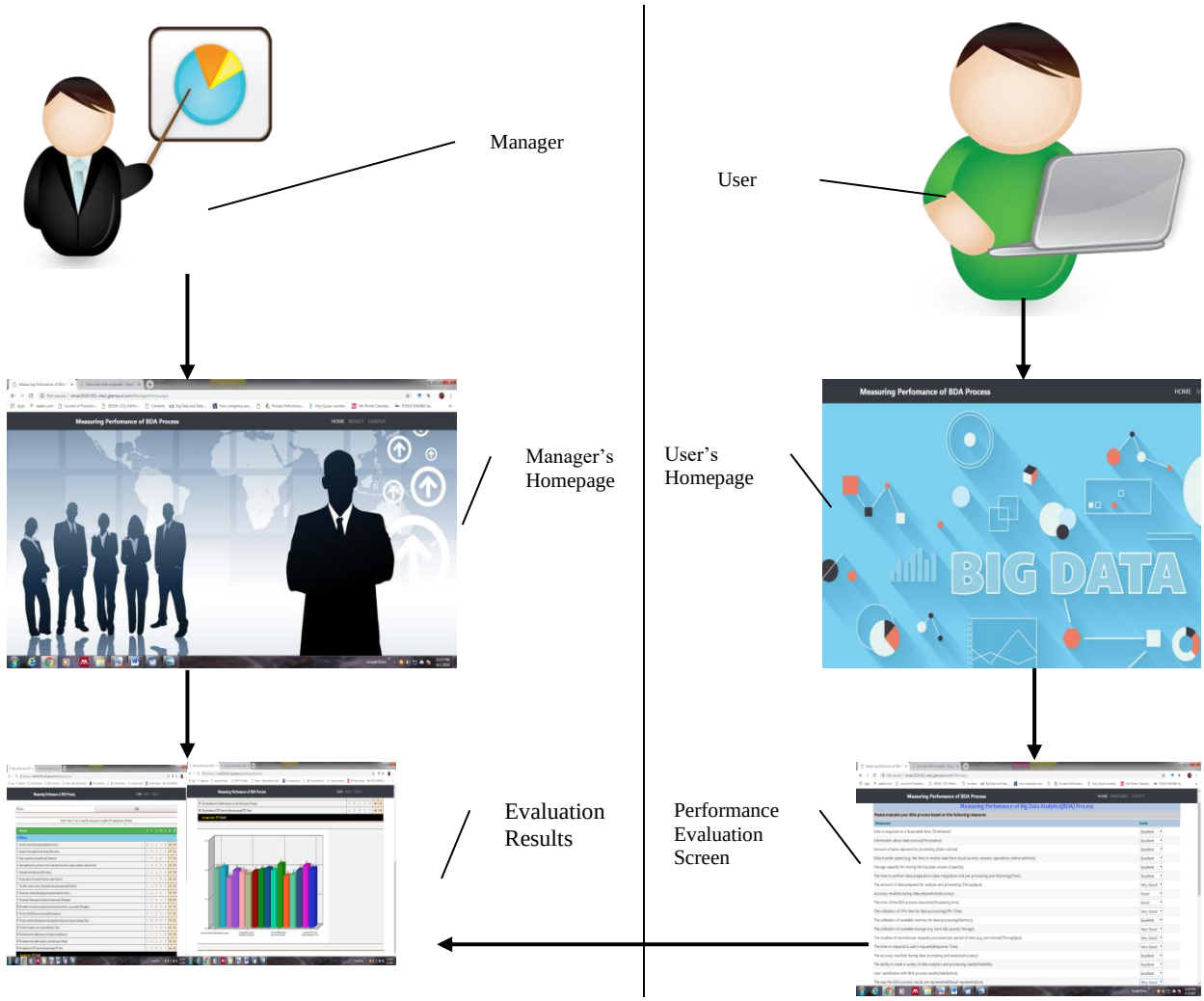

Fig. 5 The System's Screen view

IV.METHODOLOGY

This paper describes the work to develop a prototype for measuring BDA performance and present evaluation for this prototype. The prototyping approach consists of threesteps. For first step, the requirement of the prototype was identified. Performance measures and metrics were identified from the literature and showcased with experts in the field, then surveyed with practitioners of BDA. The logic of performance measurement including the formulas an rating scale has also been identified. This step also includes describing development tools, and architecture, and use cases for the system.

The second step includes the useof the refined measures and metrics to develop a performance measurement system for BDA process. The system was developed to accommodate two types of users: User and Manager: The users are normally the practitioners of big data system in any organization that introduces BDA. The user's job in this system is to evaluate their big data systems based on the measures and metrics provided, and then send the results to decision makers. The decision maker, termed here as manager, receives the results and uses them to inform their decision for improvement. As this cycle continues, more performance gaps can be discovered and systems can be improved for better. For the third step, the system was evaluated using subject-matter experts and usability survey performed by the potential users. In this regard, the system was examined by two experts who haveabackground in BDA. Similarly, the usability survey consisting of 10 questions was answered by 12 big data practitioners. The results of usability survey are reported in this paper.

V.RESULTS

Usability Survey Results

Section of the survey consisted of ten questions relating to usability of the system as per user's point of view. The respondents were asked to demonstrate the system and respond to the survey accordingly. A five-point Likert scale questions were used, where strongly Disagree=1,

Disagree=2, Slightly Agree=3, Agree=4, and Strongly Agree=5. There was commentary section where respondents provided further explanations they went through the system.

So far, in this section, the responses of respondents have been discussed by each statement in separate. The whole results are summarized in Table 2.

As shown on Table 2, respondents remarkably expressed their intention to use the system but indicated their concerns about system's complexity. The concerns about system's complexity can be traced to the lack of user manual and the overwhelming number of measures implemented in this system. This is paralleled by more than half of the respondents who asserted the system is easy to use. Also, most of users expressed they don't need the support of a technical person to be able use the system. Furthermore, majority of them responded that the system is well-integrated, which mean various parts of the system are well-suited to each other. On the flip side, a statement for system's inconsistence was no supported by the survey results. Users also remarkably endorsed the learnability of the system. This means no need for learning to many things when using the system. Perhaps it is why they perceived the system as less cumbersome and that they are confident with using it.

Table. 2 Summary of Usability Survey Results

Item	Strongly Disagree	Disagree	Slightly Agree	Agree	Strongly Agree
I think I would like to use this system frequently	0	3(25%)	2(16.7%)	5(41.7%)	2(16.7%)
I found the system unnecessarily complex.	1(8.3%)	3(25%)	5(41.7%)	2(16.7%)	1(8.3%)
I thought the system was easy to use.		1(8.3%)	4(33.3%)	3(25%)	4(33.3%)
I think that I would need the support of a technical person to be able to use this system.	4(33.3%)	5(41.7%)	1(8.3%)	1(8.3%)	1(8.3%)
I found the various functions in this system were well integrated.	0	0	4(33.3%)	5(41.7%)	2(16.7%)
I thought there was too much inconsistency in this system.	3(25%)	6(50%)	3(25%)	0	0
I would imagine that most people would learn to use this system very quickly.			3(25%)	3(25.0%)	6(50%)
I found the tool very cumbersome to use.	2(16.7%)	5(41.7%)	4(33.3%)	0	1(8.3%)
I felt very confident using the system.	0	2(16.7%)	2(16.7%)	4(33.3%)	4(33.3%)
I needed to learn a lot of things before I could get going with this system.	3(25%)	5(41.7%)	4(33.3%)	0	0

Further Comments and Discussion

The section discusses some comments that users provided. The respondents mainly expressed the comments on user interface (UI) design and user experience (UX) design, and need for more clarification for system's features. One comment reads as follows: "The UI & UX must be enhanced to capture the attention of the user; more explanation should be attached with the questions being asked". Another user writes similar observation as "The home page is empty, it was better if the description of the system is explain at the home page and some details can give about the measurement scales so whoever assess your system can understand what this system for". Another user recommends that control boxes should be replaced with checkboxes, appropriate charting tools should be used. How the type of preferred charts should be was not specified. Two users suggested that presentation of the measures should be simplified by splitting them into pages. One of them states" evaluation /measures should be divided into categories or displayed in pages". The other says "the measures, the amount of data have been shown to users are too much and can be intimidating. I suggest breaking all of them to several pages"

Some users gave overall recommendation for the system. A comment of one reads as "very nice system to use in big data analytics measurement". Another remarked as "good system and useful"

A user questions the commercial value of the systems as saying "what's the commercial value of the system?" This can be attributed to that users come from the industry and system's contribution to their businesses in quantifiable manner is at their priority. The research is focused on performance of big data analytics its self, rather than big data contribution to business performance. The general idea is that strong and robust systems will lead to business and organizational performance. Therefore, the research is expected to contribute to the success of big data analytics, thereby realize the promise of big data in improving decision making, optimizing business processes, and creating new business models. Achieving this will undoubtedly level up business successes and contribute to the digital economy at large.

VI.CONCLUSION

Big data with promises and potential benefits, is taking points in both academic researches and industry practices. Both it could be more encouraging if theses promises are paralleled with success stories of big data initiatives in which striking number of failure is reported. Is it the computational algorithms that can solve the buzzle? Or the scarcity of skills is the reason, or advanced tools have the answer or epestimological challenges is to be blamed. More researches of this type are increasingly directed to big data. But performance side which intrumental to any success is minimally disccused. This paper discussed a prototype which tries to implement existing performance measures into big data context and facilitate performance evaluation for big data systems. The prototype was evaluated by experts and practiotioners in big data field and the results were presented in this paper.

REFERENCES

1. Bally, L., Brittan, J., & Wagner, K. H. (1977). A prototype approach to information system design and development. *Information & Management*, *1*(1), 21-26.
2. Spitzer, M., Nanic, I., &Ebner, M. (2018). Distance Learning and Assistance Using Smart Glasses. *Education Sciences*, *8*(1), 21.
3. Blomkvist, J., &Holmlid, S. (2011). Existing prototyping perspectives: considerations for service design. *Nordes*, (4).
4. Beynon-Davies, P., Tudhope, D., & Mackay, H. (1999). Information systems prototyping in practice. Journal of Information Technology, 14(1), 107-120.
5. Sommerville, I. (2011). Software engineering 9th Edition. *ISBN-10*, *137035152*.
6. 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
7. Heo, J., &Haan, I. (2000). Measuring Performance of Information Systems in Evolving Computing Environments: An Empirical Investigation. *PACIS 2000 Proceedings*, 25
8. Brunnert, A., Vögele, C., Danciu, A., Pfaff, M., Mayer, M., &Krcmar, H. (2014). Performance management
9. 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.
10. 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.
11. 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.
12. Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. (2017, July). Big Data Analytics Capability: Antecedents and Business Value. In *PACIS* (p. 136).
13. Heckl, D., &Moormann, J. (2010). Process performance management. In *Handbook on business process management 2* (pp. 115-135). Springer, Berlin, Heidelberg.
14. Mohamed Ali, I., Yah Jusoh, Y., Abdullah, R., Nor, H., & Nor, R. (2017). A Conceptual Framework For Measuring The Performance of Big Data Analytics Process. ActaInformatica Malaysia (AIM), 1(2), 13-14.