Significance of Big Data Analytics in Higher Education Institutions of Saudi Arabia

Tareq H. Aljahdali[1] & Prof. Abdullah S AL-Malaise AL-Ghamdi [2]

[1]Student, Faculty of Computing and Information Technology, King Abdulaziz University, SA
*Corresponding Author E-mail: thaljahdali@kau.edu.sa

[2]Associate Professor, Faculty of Computing and Information Technology, King Abdulaziz University
Jeddah, Saudi Arabia, E-mail: aalmalaise@kau.edu.sa

Abstract: Research is an observatory study addressing the importance of big data analytics in higher education institutions of Saudi Arabia. Quantitative data collection method was deployed to collect data from King Abdulaziz University and then it was analyzed by using PLS-SEM. The study has proven that there is a high requirement of applying big data analytics on the university networks to separate the useable data from the unusable data. This will improve network performance and make the network more stable. The study has contributed to the knowledge base theory significantly.

Keywords: Big Data, Technical Analytics, Operational Capabilities, Dynamic Capabilities, Network Performance.

I. INTRODUCTION

The research study is planned to address the key issues of big data utilization by providing best possible access time to the end-users while running the local and online applications. This modern method of arranging and distributing the data into meaning full and usable patterns is known as big data. Which is being generated at a fast scale and spread over the network and at peak times network resources might block the network traffic. Current research is planned to address big data analytics to avoid network problems and to make the network more stable and useful.

II. RESEARCH BACKGROUND

The term big data analytics is known as the most appropriate phenomena to deal with big data [1]. Big data is generally considered in some important aspects named as the volume, velocity, variety and veracity [2]. Where Volume is the amount of data to be analyzed, velocity is the speed at which data flows and variety is the variations of data received from the different sources and can be in different formats like structured and unstructured data [3]. Whereas, veracity is the degree to which data is accurate, precise and trusted. This extends the knowledge area of big data for further studying and interpreting these data patterns for better understanding and later to be used for business decision making and predictions [2].

For this purpose, research has formed the following objectives to be achieved during the study:

- To identify the important factors at KAU network servers for better utilization of network resources.
- To automate the network resource utilization with better security and performance.
- To lower the per-unit IT cost with less human intervention.
- To apply this model to higher education institutions in Saudi Arabia.

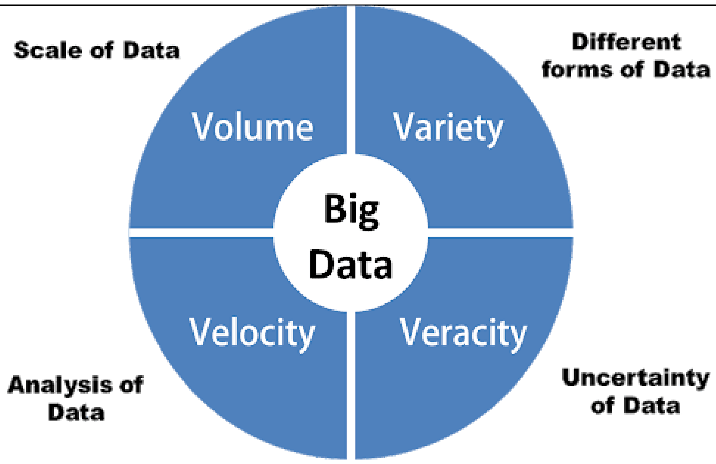

Figure 1. The four Vs of Big Data

III. PROBLEM STATEMENT

Higher Education Institutions in Saudi Arabia are facing the issues of a high volume of network traffic and affecting the business sector. Therefore, current research is initiated to focus on this core issue and will try to bridge this gap to resolve for better and more stable resource utilization in university networks. Following are some of the key problems:

- The huge costs that are paid for using internet application data.
- Waste of network infrastructure resources without any proper command and control system.
- Influencing the rest of the university's technical systems, which require an enormous amount of data due to the number of users.
- Poor response time of some internet applications that are reducing their data flow by making manual settings for resolving the issue.

IV. BIG DATA ANALYTICS CAPABILITIES

Big data analytics is the capability of exploring business data using data management tools and transform into more meaningful information for the business decision-makers [5]. Literature has proved that most of the researchers in the last few years have focused on a better understanding of these kinds of capabilities. This shows that big data analytics in university networks needs to be looked into more critically to make the network more stable and avoid vulnerability. It also provides a better insight of the network by developing a key concept of hierarchical model components upon which big data infrastructure can be applied in the higher education sector of King Abdulaziz university networks [6].

V. RESEARCH METHODOLOGY

The research study is purely exploratory and based on quantitative data, a cross-sectional survey questionnaire was used for data collection purpose from the users of King Abdulaziz university including students, faculty and technical professionals. The questionnaire was distributed to university students, faculty and technical professionals of different levels and academic major for collecting the data. Furthermore, an online survey software tool was also used for collecting data from the participants to whom researcher himself could not reach. The survey questionnaire was formed initially 7 demographic questions and 42 survey questions and then it was reviewed by an academic professional, who has the experience and he has made some changes under the research objectives and questions. Hence the questionnaire was updated accordingly and the final survey questionnaire was based on 34 questions.

VI. CONSTRUCTION OF STUDY VARIABLES

Study variables are an important aspect in terms of reviewing the literature and adopting or constructing the study constructs or variables. In this regards, detail literature has been reviewed and study variables have been constructed from the available literature regarding big data analytics. Below is the detail illustrated for the study variables, empirical evidence of the literature in the previous few years and the variables status.

Table 1. Conceptual Model for the Study

Study Variables	*Empirical Evidence from Literature*	*Variable Status*
Network Performance (NC)	(Maroufkhani et al., 2019), (Kadhim, Yusof, & Mahdi, 2018), (Brien, 2017), (Alhamed, 2017), (Klooster, 2016)	Constructed
Technical Analytical Capabilities (TAC)	(Banihashem et al., 2018), (Danielsen & Framnes, 2017), (Ramasamy, 2014)	Constructed
Dynamic Capabilities (DC)	(Maroufkhani et al., 2019), (Attaran et al., 2018), (Haan, 2018), (Ghasemaghaei et al., 2017)	Constructed
Operational Capabilities (OC)	(Raguseo & Vitari, 2018), (Mukthar & Sultan, 2017), (Ghasemaghaei et al., 2017), (Danielsen & Framnes, 2017)	Constructed

VII. DATA COLLECTION AND SAMPLING METHOD

The general investigation outline, the particular research questions, techniques, rationality and consistency of findings influence the nature of study outcome [7]. Data collected through the quantitative method by conducting surveys, polls and reviews or from some previous data by putting some mathematical procedures. Quantitative research is known as making a more accurate analysis as it is dealing with numbers, so the error margin is lesser in terms of data collection or human error. [8]. In this regards, non-probability based sampling technique is being opted for data sampling where it is not identified that which person from the population will be chosen as a sample. So for this purpose, King Abdulaziz university network was used as an organization. Around 125 samples were collected manually conducted by the researcher himself. Email method was also adopted to reach the technical professionals for getting their viewpoint and total of 26 emails were also sent with a request to participate in the online survey and this leads to a total of 151 responses.

VIII. PROPOSED STUDY MODEL

Based on the detailed literature review and discussion above, a research model has been proposed for the current study to evaluate and contribute to the knowledge and theory based on the study outcome. This model is to assess, test and verify the structural model fit and standard part coefficients to look for the supported and non-supported hypothesized relationships in the proposed study model:

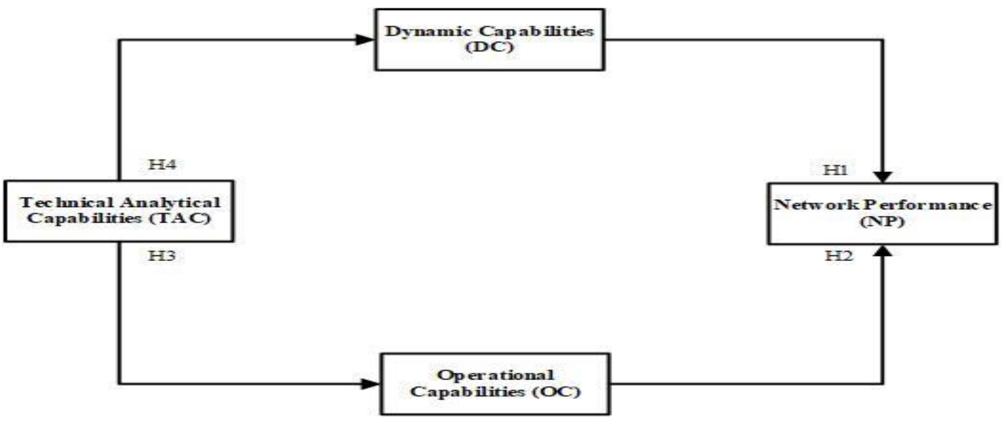

Figure 2. Proposed Model for the Research Study

IX. DATA ANALYSIS AND INTERPRETATION

Furthermore, this collected data was viewed carefully and after going through the basic analysis for the data normalization. Then the data was used for performing the analysis using PLS-SEM. Survey responses are a key phase of the study, it is hard to collect data and particularly for this research study. To collect the relevant data to perform more suitable and accurate analysis as per the propose research questionnaire and research study model.

X. FORMATIVE MEASURES

Smart PLS is used to calculate the path coefficients/weights. To further establish the validity and reliability of the outer model to calculate the t-values of all the formative indicators as a two-sided test, as recommended [9]. P-values over 0.05 are included in the table as additional info to show the relevance of these different weights. Then the Variance Inflation Factor (VIF) is assessed. These values should be below 3.3 as recommended by Petter et al., 2007. Adequacy coefficient (R^2) were calculated and these values should be above 0.50 as recommended by [10]. This detail is given below:

Table 2. Formative Indicators Values

Constructs	*Measures*	*Weight*	*t-Value*	*P*	*VIF*	*R2*
Network Performance (NP)	NP1	0.54	2.33	p<0.5	1.648	0.61
	NP2	0.54	2.33	p<0.6	1.648	
	NP3	0.54	2.33	p<0.7	1.648	
Operational Capabilities (OC)	OC1	0.54	2.33	p<0.18	1.648	0.77
	OC2	0.54	2.33	p<0.19	1.648	
	OC3	0.54	2.33	p<0.17	1.648	
Dynamic Capabilities (DC)	DC1	0.54	2.33	p<0.15	1.648	0.7
	DC2	0.54	2.33	p<0.16	1.648	
	DC3	0.54	2.33	p<0.17	1.648	
Technical Analytical Capabilities (TAC)	TAC1	0.54	2.33	p<0.18	1.648	0.71
	TAC2	0.54	2.33	p<0.19	1.648	
	TAC3	0.54	2.33	p<0.19	1.648	

XI. CONSTRUCT AND CONVERGENT VALIDITY

Content validity is defined as the level to which the proposed items suitably measure the concept of the construct that they are designed to measure [9]. Hence, research study items are considered through a detailed literature review [11]. According to Hair *et al.* (2010), convergent validity refers to the degree to which a group of variables converge in measuring a particular concept. Factor Loadings, Composite Reliability (CR) and Average Variance Extracted (AVE). As such, the loadings of the entire items were examined and confirmed where AVE must be > 0.50 CR must be > 0.70, Cronbach's must be > 0.70 and Loadings must be > 0.60. However, if all other criteria of measurement model fulfil the validity and reliability requirements, the researcher can retain some item's loading even as low as 0.40 as an acceptable level [11] and refer by [12]. The factor loadings were all significant with 0.01 level of significance, shown in the table above.

XII. THE DISCRIMINANT VALIDITY ANALYSIS

Discriminant Validity refers to the level to which items can differentiate among different constructs to show that the items of different constructs are not overlapping. In this study, discriminant validity of measures was established through Fornell and Larcker's (1981) method, where the square root of AVE for all constructs was replaced at diagonal elements of relation matrix [13].

Table 3. Composite Reliability, Cronbach's alpha and Indicator Reliability

1st Order	*Items*	*Loadings*	*AVE*	*Composite Reliability CR*	*Cronbach's α*
Network Performance (NP)	NP1	0.7583	0.5046	0.7532	0.8138
	NP2	0.7662			
	NP3	0.7424			
	NP4	0.7101			
	NP5	0.7201			
	NP6	0.8012			
	NP7	0.7301			
Dynamic Capabilities (DC)	NTS1	0.7223	0.5743	0.8542	0.8432
	NTS2	0.7341			
	NTS3	0.8101			
	NTS4	0.7701			
	NTS5	0.6995			
	NTS6	0.4053			
	NTF1	0.8196	0.5321	0.7351	0.7121
	NTF2	0.7858			
	NTF3	0.7735			
	NTF4	0.7229			
	NTF5	0.7399			
	NTT1	0.6896	0.5546	0.7553	0.7138
	NTT2	0.7679			
	NTT3	0.7583			
	NTT4	0.7662			
	NTT5	0.7424			
Operational Capabilities (OC)	OC1	0.6995	0.5139	0.7353	0.7081
	OC2	0.6605			
	OC3	0.8196			
	OC4	0.7858			
	OC5	0.7735			
Technical Analytical Capabilities (TAC)	TAA1	0.6896	0.5845	0.8921	0.7966
	TAA2	0.7229			
	TAA3	0.7399			
	TAA4	0.7679			
	TAA5	0.6442			

XIII. SIGNIFICANCE OF THE STRUCTURAL MODEL

Convergent validity is assessed by looking at AVE (Average Variance Extracted) values. These numbers should be above 0.50 [9] and are included in table 4.13. This is established discriminant validity by creating a cross-loading overview and checked that the indicators are measured accurately during the study phase.

Finally, path coefficients statistical significance can be determined via bootstrapping methods in Smart PLS 3. In this regard, the t-values of each path coefficient were produced through such method and p-values were eventually obtained during the analysis phase of the study. Further detail of coefficients produced is presented in the following table 4.15.

Table 4. Significance of the Structural Model

Relations	*Hypothesis*	*Statement*	*Beta*	*t-Statistics*	*Decision*
DC->NP	H1	DC has direct and positive impact to NP.	0.2431	3.771	Strong
OC->NP	H2	TAC has direct and positive impact to DC and indirect positive impact on NP.	0.2923	2.928	Strong
TAC->OC->NP	H3	TAC has direct and positive impact to NP through OC.	0.2041	2.874	Moderate
TAC->DC->NP	H4	OC has direct positive impact to OC.	0.3721	4.810	Moderate

Table 5. Effect Size (R2) and (Q2) on Network Performance

Variable	*Variable Type*	*R2*	*Q2*	*Cross-Validated Communality*	*Cross Validated Redundancy*
Network Performance (NP)	Endogenous	0.736	0.251	0.664	0.355
Operational Capabilities (OC)	Exogenous	0.648	0.166	0.625	0.504
Dynamic Capabilities (DC)	Exogenous	0.543	0.114	0.441	0.453
Technical Analytical Capabilities (TAC)	Exogenous	0.543	0.114	0.441	0.453

XIV. THEORETICAL CONTRIBUTION

The study has extended moderate to a strong level of significance in terms of evaluating and assessing network performance by using quantitative analysis. The study results have shown consistency with the previous studies result. Therefore, it is assumed that current research supports big data analytics capabilities to improve network performance. It is important to mention here that the network performance has a significant impact through both Operational and Dynamic Capabilities, as shown by the study results. Quantitative results have shown some implications regarding its planning for data collection and analysis.

XV. LIMITATIONS AND FUTURE PROSPECT

The current study has been initiated to address the very important issue of application of technical data analytics in the higher education institutions of Saudi Arabia. Big data analytics is a vast field and it needs lots of technical measures for assessing evaluating both technical analytical applications in university network performance through operational and dynamic capabilities as mediator. Hence, Technical Analytical Capabilities has moderate to a strong level of impact on the network performance. Study results have shown positive a significant impact under the precious literature. Unfortunately, moderate level of significance of big data analytics at H1 and H4 mediating through both Operational and Dynamic Capabilities has shown a bit worry and need to be addressed in future to improve and enhance the network performance by adding more factors in the research model.

XVI. CONCLUSION

A detailed study has been done to review the relevant literature addressing big data analytics in the computer network. For this purpose, research questions have been constructed and respondent's feedback through the survey questionnaire in both cases have played an essential role in conducting the study for betterment of the university networks. Results have been evaluated by using the smart PLS and supported by quantitative analysis. The study results have shown moderate to strong level impact during the analysis, which shows that study can be extended further to use more relevant factors for better outcome. Future studies could further extend by adding the more suitable factors and technical solutions for better study outcome.

REFERENCES

[1] J. N. Undavia, S. Patel, and A. Patel, "Future trends and scopes of Big Data Analytics in the field of Education," Int. J. Eng. Technol., vol. 9, no. 3S, pp. 9–14, 2017.

[2] V. Heilala, "Framework for Pedagogical Learning Analytics," Univ. Jyväskylä Fac. Inf. Technol., pp. 1–84, 2018.

[3] G. Johnes, J. Johnes, T. Agasisti, L. López-Torres, T. Agasisti, and A. J. Bowers, "Data analytics and decision making in education: towards the educational data scientist as a key actor in schools and higher education institutions," Handb. Contemp. Educ. Econ., pp. 184–210, 2018.

[4] M. Ghasemaghaei, K. Hassanein, and O. Turel, "Increasing firm agility through the use of data analytics: The role of fit," Decis. Support Syst., vol. 101, pp. 95–105, 2017.

[5] V. Vatsala, R. Jadhav, and S. R, "A Review of Big Data Analytics in Sector of Higher Education," Int. J. Eng. Res. Appl., vol. 07, no. 06, pp. 25–32, 2017.

[6] O. F. . Klooster, "Increasing agility in Big Data analytics through implementation of the BASE / X Increasing Agility in Big Data analytics through implementation of the BASE / X framework : A In partial fulfillment of the requirements for the degree of Master of Science In," 2016.

[7] A. Bryman, Research Methods and Organization Studies, vol. 20. 2005.

[8] A. Bryman, "Integrating quantitative and qualitative research: How is it done?," Qual. Res., vol. 6, no. 1, pp. 97–113, 2006.

[9] J. F. Hair, M. Sarstedt, L. Hopkins, and V. G. Kuppelwieser, "Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research," Eur. Bus. Rev., vol. 26, no. 2, pp. 106–121, 2014.

[10] C. Baudoin, E. Dekel, and M. Edwards, "Interoperability and Portability for Cloud Computing: A Guide," Cloud Stand. Cust. Counc., pp. 1–20, 2014.

[11] J. F. Hair, B. J. Babin, and N. Krey, "Covariance-Based Structural Equation Modeling in the Journal of Advertising: Review and Recommendations," J. Advert., vol. 46, no. 1, pp. 163–177, 2017.

[12] H. Hassan, M. Herry, M. Nasir, N. Khairudin, and I. Adon, "Factors Influencing Cloud Computing Adoption in Small and Medium Enterprises," J. ICT, vol. 16, no. 1, pp. 21–41, 2017.

[13] I. Ghadi, N. H. Alwi, K. Abu Bakar, and O. Talib, "Construct Validity Examination of Critical Thinking Dispositions for Undergraduate Students in University Putra Malaysia," High. Educ. Stud., vol. 2, no. 2, pp. 138–145, 2012.

