

Effectiveness of Big Data Analytics

Dr. Rekha Chouhan

Assistant Professor, Balaji Institute of Telecom & Management, Pune (India)

ARTICLE DETAILS

Article History

Published Online: 10 October 2018

Keywords

Big Data Analytics ; Information System Domain

*Corresponding Author

Email: rekha09[at]gmail.com

ABSTRACT

Extremely large data sets that may be analysed, computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. Big Data Analytics (BDA) readiness factors have been widely researched; nevertheless, few have investigated the impact and success factors of BDA implementation in the organizational context. The relevance and quality of BDA outcomes have been a significant concern to the organizational leaders in supporting them for strategic decision-making. To that end, the objective of this study is twofold. Firstly, it investigates the factors that influencing the success of BDA implementation for effective decision-making. Secondly, this study adds to the body of knowledge in the information system (IS) domain, especially with its focus on BDA implementation packages. This review has established the influencing factors that may influence the success of BDA implementation, therefore, contribute to the practice and research of BDA domain and its effectiveness towards the organizational performance enhancement. Big Data has emerged as one of the biggest business and technological buzzwords of recent years. The term Big Data applies to information that cannot be processed or analysed using traditional tools or processes.

1. Introduction

Big data, as the name suggests, refers to large datasets that are challenging to store, share, search, visualize, and analyse and here the orders of magnitude exceed conventional data processing and the largest of data warehouses. Big Data, coming from various sources, whether an airline jet collecting ten terabytes of sensor data for every half an hour of flying time, New York Stock Exchange collecting one terabyte of structured trading data each day, or a conventional structured corporate data warehouse sized in terabytes and peta-bytes, is sized in peta-, exa-, and even in zetta-bytes. It also signals that it is not just about volume, the approaches to analysis compete with data content and structure that can neither be anticipated nor predicted. There is a need to put proper analytics and the science behind them to filter low value or low-density data and divulge high value or high-density data. Also, big Data has a broad array of interesting architecture challenges and thus, new analytical techniques are required to adopt.

Usually data volume, velocity, and variety describe big data, but the unique attribute of Big Data is the manner in which the value is revealed. In conventional business intelligence, the simple summing of a known value reveals a result, such as order sales becoming year-to-date sales. But big data requires a refining modeling process to discover any value, i.e. making a hypothesis, creating statistical, visual, or semantic models, validating, and then making a new hypothesis. And for this, it either takes a person interpreting visualizations or making interactive knowledge-based queries, or by developing 'machine learning' adaptive algorithms that can discover business meaning. With the growth of big data, firms in emerging economies have also started investing in solutions that interpret consumer behavior, detect fraud, and even predict the future using big data analytics (BDA). Undoubtedly, the rise of big data in prominence has led to a titanic focus on exploring how organizations can harness

information to gain a competitive advantage. However, despite big data's well documented benefits, it would be important to investigate, how organisations across the globe are putting it to use and in which way? Recently, Big Data Analytics (BDA) has emerged as a new technology to enhance overall efficiency of management through productivity, performance, and better decision-making of the organizations in real-time.

However, recent research in information management lack focus on BDA adoption which is just one part of an adoption process, and it cannot ensure wide-scale exploitation and usage of BDA. Therefore, without wide-scale adoption, the benefits of BDA cannot be fully realized. The economic status and regulatory environment of such countries are different from developed countries where BDA technology already has established usage at huge level. Thus, it would be worthy to investigate how innovation assimilation gets influenced by contextual factors in such environments. Motivated by above theoretical gaps, present paper proposes an integrative model integrating determinants that impact BDA adoption in the context of two developing countries from Asia –China and India. The model is based on diffusion of innovation (DoI) theory, institutional theory and technology, organization and environment (TOE) framework.

Big data analytics (BDA) readiness factors have been widely researched; nevertheless, few have investigated the impact and success factors of BDA implementation in the organizational context. The relevance and quality of BDA outcomes have been a significant concern to the organizational leaders in supporting them for strategic decision-making. To that end, the objective of this study is twofold. Firstly, it investigates the factors that influencing the success of BDA implementation for effective decision-making. Secondly, this study adds to the body of knowledge in the information system (IS) domain, especially with its focus on BDA implementation packages. This review has established influencing factors that

may influence the success of BDA implementation, therefore, contribute to the practice and research of BDA domain and its

effectiveness towards the organizational performance enhancement.

Types of Big Data Analytics				
	Descriptive	Diagnostic	Predictive	Prescriptive
Answers the question...	What happened?	Why did it happen?	What will happen next?	What should I do?
Level of advancement	Low	Medium	High	Very high
Incorporates AI and machine learning?	Not usually	Sometimes	Usually	Always
Level of popularity	Used by almost all organizations	Used by many organizations	Used by a smaller but growing group of organizations	Not yet widespread

Figure:1

2. Big Data Analytics Tools

By itself, stored data does not generate business value, and this is true of traditional databases, data warehouses, and the new technologies for storing big data (e.g., Hadoop). Once the data is appropriately stored, However, it can be analyzed and this can create tremendous value. A variety of analysis technologies, approaches, and products have emerged that are especially applicable to big data, such as in-memory analytics, in data base analytics, and appliances. Analytics is another buzzword in the technology industry and it “refers to our ability to collect and use data to generate insights that inform fact-based decision-making”. In the past, the data that was analysed was mainly used to predict that was analysed was mainly used to predict what might happen in the future and was fully embraced by industries such as banks and insurance companies, but not by organisations such as retailer. (Figure:1). Big Data and analytics go hand in hand in the current technological age. “Big Data analytics uses predictive and prescriptive analytics and is changing the analytics landscape”. Predictive analytics uses data from the past to predict what may happen, and the likelihood of it happening in the future. Whereas, prescriptive analytics is taking data from the past, using it to decide on what should be done next to achieve optimal results.

Big Data refers to datasets whose characteristics make typical database approaches unable to store, analyze, and manage them [5]. Big Data techniques and technologies are often seen as techniques and technologies able to work on huge amount of data with good performance. However, Big Data are much more and recently have been defined using the 5V model: Volume, Variety, Velocity, Value, Veracity [6], [7]. In short, this model highlight how Big Data technologies are oriented to the implementation of distributed analytics, that is, scalable analytics handled in architectures that adapt computational resources to the volume, variety and velocity of

data, accepting the increasing difficulty in controlling data quality and trustworthiness (veracity), and making possible to radically improve the value generated by offering results at runtime. While Big Data analytics can in principle support existing or new value propositions in a number of business domains, choosing and deploying the “right” analytics on the “right” computational infrastructure is still more an art that a science or an engineering practice. Big data and analytics are intertwined, but analytics is not new. Many analytic techniques, such as regression analysis, simulation, and machine learning, have been available for many years. Even the value in analyzing unstructured data such as email and documents has been well understood. What is new is the coming together of advances in computer technology and software, new sources of data (e.g., social media), and business opportunity[1]. This confluence has created the current interest and opportunities in big data analytics. It is even spawning a new area of practice and study called “data science” that encompasses the techniques, tools, technologies, and processes for making sense out of big data. At a high level, the requirements for organizational success with big data analytics are the same as for business intelligence in general. At a deeper level, however, there are many nuances that are important and need to be considered by organizations that are getting into big data analytics. For example, there are organizational culture, data architecture, analytical tools, and personnel issues that need to be considered. Of particular interest to information technology (IT) professionals are the new technologies, platforms, and approaches that are being used to store and analyze big data. Big data consist of three types of platforms. There are batch processing, real time processing and interactive analytics. Batch processing platforms take time to process data, there are the most deployed big data platforms, they carry out extensive computations; real time processing require faster processing of data, streaming is needed in applications that require minimal latency; interactive analytics

allow users to access dataset remotely and carry out numerous operations as needed.

The batch processing platform Apache Hadoop is the most deployed. Apache Hadoop is an open source implementation of Map Reduce. Map Reduce was designed at Google and used on applications like large scale machine learning and clustering problems; it was inspired by the map and reduce primitives of the functional language Lisp. A problem is broken down by Hadoop recursively into the smallest unit to be solved. The small chunks are then distributed on system nodes for execution through the coordination of a master node and working nodes. Hadoop is a software library with modules such as Map Reduce, HDFS and YARN. Hadoop also has abstractions that operate in its ecosystem; they are Hive, Hbase, Flume, Sqoop, Chukwa and Zookeeper [11]. These Hadoop modules and abstractions operate across the big data value chain; from acquisition, storage, computation, analysis and management. Hadoop is deployed on a huge scale and scope in companies due to scalability, cost effectiveness, flexibility and fault tolerance. There are many techniques for analyzing big data, the techniques cut across disciplines and overlap in some cases. The disciplines range from data mining, machine learning, neural networks and pattern recognition [2]. A technique from a discipline or combination of them is used to gain valuable insight from big data. These techniques in the context of big data are called analytics. Machine learning techniques are known for their predictive capability, given a previous instance of data, the technique is able to infer appropriate response to subsequent occurrence. With this ability at hand, machine learning techniques have solved traditional computing tasks. There are numerous learning techniques for diverse problem types. There is no "one size fits all" technique; this is due to unique ability, limitations of each technique and nature of problems [13]. Machine learning techniques are primarily hindered by scale of data and computation in big data. We would look at some improved learning techniques used in big data context. A Map Reduce framework for real time traffic processing was designed in a study; the challenges were distributed processing, distributed local learning and model fusion. The framework used Expectation Maximization algorithm for local learning process,

the global model is compared to the distributed models in order to merge into a final model for traffic forecasting. Support Vector Machine was also improved with a probabilistic distribution for a large scale data application; the problem required an active learning mechanism due to sample complexity. Cross domain learning incorporating Support Vector Machine was used to develop a new and efficient method for learning classification of images across domains, the method works despite a small training data and different distributions. The method was evaluated against other cross domain ones on large video datasets. Other scaled machine learning algorithms are fast learning for ranking and sparse approximations through boosting. Machine learning are used by another set of techniques called Data Mining in big data problems. Data mining techniques are used for deducing information from data. There are used for clustering, regression and association rule mapping. Data mining techniques equally face computational and data complexity inherent in big data.

3. Conclusion

Big data platforms and tools are the cornerstone of the data revolution taking place. Volume, velocity and variety are the main features that make big data outstanding from traditional computing. Big data has been premised to make profound insights in all domains. Innovative models, products, services and huge cost savings are some of the opportunities big data present. Big data due to its huge potential and user base has multiple definitions reflecting wide perspective of its stakeholders. Platforms for big data have diverse purposes with regard to speed at which data needs to be produced and processed. The platforms are batch processing, stream processing and interactive analytics. Big data analytics particularly focus on the prediction of future outcomes and embedding predictive models in business operations. While businesses have always been trying to predict the future, big data opens up new possibilities. New sources of data mean that it is possible to make predictions in new areas and machine learning techniques increase the accuracy of prediction models. As a result, greater reliance can be placed on models.

References

1. Agneeswaran, V., 2012. Big-data - Theoretical, engineering and analytics perspective. In: Srinivasa, S., Bhatnagar, V. (Eds.), *Big Data Analytics*. Vol. 7678 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 8-15.
2. Anastasia, February 2015. Big data and new product development. *Entrepreneurial Insights* <http://www.entrepreneurial-insights.com/big-data-new-product-development/>, Accessed on June 15, 2015. Boyd, D., Crawford, K., 2012. Critical questions for big data. *Information, Communication & Society* 15:5, 662-679.
3. D. Che, M. Safran, and Z. Peng, "From big data to big data mining: challenges, issues, and opportunities," in *Database Systems for Advanced Applications*, B. Hong, X. Meng, L. Chen, W. Winawater, and W. Song, Eds., vol. 7827 of *Lecture Notes in Computer Science*, pp. 1-15, Springer, Berlin, Germany, 2013.
4. P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, 2011.
5. S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4-6, 2012.
6. P. Russom, *Big Data Analytics*, TDWI best practices report, TDWI Research, 2014, http://www.iso.org/iso/home/news/index/news_archive/news.htm?refid=Ref1821.
7. P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, 2011.
8. V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon
9. Dolan/Houghton Mifflin Harcourt, 2013.
10. Chen M, Mao S, Liu Y. Big data: A survey. *Mob Networks Appl*. 2014; 19(2): 171-209.

11. Chardonnes T, Cudre-Mauroux P, Grund M, Perroud B. Big data analytics on high Velocity streams:
12. A case study.2013 IEEE IntConf Big Data. 2013: 784–787.
13. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. and Kruschwitz, N. (2011), Big Data, Analytics and the Path from Insights to Value, MIT Sloan Management Review, Vol 52(2), pp. 21–31.
14. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh C. and Beyers, A. (2011), Big Data: The Next Frontier for Innovation, Competition and Productivity, McKinsey Global Institute.
15. Kwon, O., Lee, N., & Shin, B (2014): "Data quality management, data usage experience and acquisition intention of big dataanalytics", International Journal of Information Management, 34, 387-394.
16. Codd, E.F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13(6), 377-387.