



Contents lists available at ScienceDirect

Journal of Business Research



Critical analysis of Big Data challenges and analytical methods

Uthayasankar Sivarajah *, Muhammad Mustafa Kamal, Zahir Irani, Vishanth Weerakkody

Brunel University London, Brunel Business School, UB8 3PH, United Kingdom

ARTICLE INFO

Available online xxxx

Keywords:

Big Data
Big Data Analytics
Challenges
Methods
Systematic literature review

ABSTRACT

Big Data (BD), with their potential to ascertain valued insights for enhanced decision-making process, have recently attracted substantial interest from both academics and practitioners. Big Data Analytics (BDA) is increasingly becoming a trending practice that many organizations are adopting with the purpose of constructing valuable information from BD. The analytics process, including the deployment and use of BDA tools, is seen by organizations as a tool to improve operational efficiency though it has strategic potential, drive new revenue streams and gain competitive advantages over business rivals. However, there are different types of analytic applications to consider. Therefore, prior to hasty use and buying costly BD tools, there is a need for organizations to first understand the BDA landscape. Given the significant nature of the BD and BDA, this paper presents a state-of-the-art review that presents a holistic view of the BD challenges and BDA methods theorized/proposed/employed by organizations to help others understand this landscape with the objective of making robust investment decisions. In doing so, systematically analysing and synthesizing the extant research published on BD and BDA area. More specifically, the authors seek to answer the following two principal questions: Q1 – *What are the different types of BD challenges theorized/proposed/confronted by organizations?* and Q2 – *What are the different types of BDA methods theorized/proposed/employed to overcome BD challenges?* This systematic literature review (SLR) is carried out through observing and understanding the past trends and extant patterns/themes in the BDA research area, evaluating contributions, summarizing knowledge, thereby identifying limitations, implications and potential further research avenues to support the academic community in exploring research themes/patterns. Thus, to trace the implementation of BD strategies, a profiling method is employed to analyze articles (published in English-speaking peer-reviewed journals between 1996 and 2015) extracted from the Scopus database. The analysis presented in this paper has identified relevant BD research studies that have contributed both conceptually and empirically to the expansion and accrual of intellectual wealth to the BDA in technology and organizational resource management discipline.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The magnitude of data generated and shared by businesses, public administrations numerous industrial and not-to-profit sectors, and scientific research, has increased immeasurably (Agarwal & Dhar, 2014). These data include textual content (i.e. structured, semi-structured as well as unstructured), to multimedia content (e.g. videos, images, audio) on a multiplicity of platforms (e.g. machine-to-machine communications, social media sites, sensors networks, cyber-physical systems, and Internet of Things [IoT]). Dobre and Xhafa (2014) report that every day the world produces around 2.5 quintillion bytes of data (i.e. 1 exabyte equals 1 quintillion bytes or 1 exabyte equals 1 billion

gigabytes), with 90% of these data generated in the world being unstructured. Gantz and Reinsel (2012) assert that by 2020, over 40 Zettabytes (or 40 trillion gigabytes) of data will have been generated, imitated, and consumed. With this overwhelming amount of complex and heterogeneous data pouring from any-where, any-time, and any-device, there is undeniably an era of *Big Data* – a phenomenon also referred to as the *Data Deluge*. The potential of BD is evident as it has been included in Gartner's *Top 10 Strategic Technology Trends for 2013* (Savitz, 2012a) and *Top 10 Critical Tech Trends for the Next Five Years* (Savitz, 2012b). It is as vital as nanotechnology and quantum computing in the present era. In essence, BD is the artefact of human individual as well as collective intelligence generated and shared mainly through the technological environment, where virtually anything and everything can be documented, measured, and captured digitally, and in so doing transformed into data – a process that Mayer-Schönberger and Cukier (2013) also referred to as *datafication*.

In line with the datafication concept and ever increasing technological advancements, advocates assert that in the future a majority of data

* Corresponding author at: Brunel University London, College of Business, Arts and Social Sciences, Brunel Business School, UB8 3PH, United Kingdom.

E-mail addresses: Sankar.Sivarajah@brunel.ac.uk (U. Sivarajah), Muhammad.Kamal@brunel.ac.uk (M.M. Kamal), Zahir.Irani@brunel.ac.uk (Z. Irani), Vishanth.Weerakkody@brunel.ac.uk (V. Weerakkody).

will be generated and shared through machines, as machines communicate with each other over data networks (Van Dijck, 2014). Regardless of where BD is generated from and shared to, with the reality of BD comes the challenge of analysing it in a way that brings *Big Value*. With so much value residing inside, BD has been regarded as today's *Digital Oil* (Yi, Liu, Liu, & Jin, 2014) including the *New Raw Material* of the 21st century (Berners-Lee & Shadbolt, 2011). Appropriate data processing and management could expose new knowledge, and facilitate in responding to emerging opportunities and challenges in a timely manner (Chen et al., 2013). Nevertheless, the growth of data in volumes in the digital world seems to out-speed the advance of the many extant computing infrastructures. Established data processing technologies, for example database and data warehouse, are becoming inadequate given the amount of data the world is current generating. The massive amount of data needs to be analyzed in an iterative, as well as in a time sensitive manner (Jukić, Sharma, Nestorov, & Jukić, 2015). With the availability of advanced BD analysing technologies (e.g. NoSQL Databases, BigQuery, MapReduce, Hadoop, WibiData and Skytree), insights can be better attained to enable in improving business strategies and the decision-making process in critical sectors such as healthcare, economic productivity, energy futures, and predicting natural catastrophe, to name but a few (Yi et al., 2014).

As evident, much has been written on the BD phenomenon. The majority of academic research articles reviewed are analytical in nature (also evident from the findings – see Figs. 10 and 11) that is either focusing on using experiments, simulations, algorithms and or mathematical modelling techniques in tackling BD. Regardless of their research approach, these articles present BD as a source that when appropriately managed, processed and analyzed, have the potential to generate new knowledge thus proposing innovative and actionable insights for businesses (Jukić et al., 2015). There is an ever-growing discourse about BD offering both *Big Opportunities* and *Big Challenges* through the plethora of sources from different domains; extending from enterprises to sciences. For instance, the opportunities include value creation (Brown, Chui, & Manyika, 2011), rich business intelligence for better-informed business decisions (Chen & Zhang, 2014), and support in enhancing the visibility and flexibility of supply chain and resource allocation (Kumar, Niu, & Ré, 2013). On the other hand, the challenges are significant such as data integration complexities (Gandomi & Haider, 2015), lack of skilled personal and sufficient resources (Kim, Trimi, & Chung, 2014), data security and privacy issues (Barnaghi, Sheth, & Henson, 2013), inadequate infrastructure and insignificant data warehouse architecture (Barbierato, Gribaudo, & Iacono, 2014), and synchronising large data (Jiang, Chen, Qiao, Weng, & Li, 2015). Advocates such as Sandhu and Sood (2014) perceive that the potential value of BD cannot be unearthed by simple statistical analysis. Zhang, Liu et al. (2015) support this perspective and state that to tackle the BD challenges, advanced BDA requires extremely efficient, scalable and flexible technologies to efficiently manage substantial amounts of data – regardless of the type of data format (e.g. textual and multimedia content).

1.1. Research scope

BD and BDA as a research discipline are still evolving and not yet established, thus, a comprehensible understanding of the phenomenon, its definition and classification is yet to be fully established. The extant progress made in BD and BDA not only revealed a lack of management research in the field but a distinct lack of theoretical constructs and academic rigor – perhaps a function of an underlying methodological rather than academic challenge. At large, there has also been a lack of research studies that comprehensively addresses the key challenges of BD, or which investigates opportunities for new theories or emerging practices (e.g. George, Haas, & Pentland, 2014). Thus, there exists the need to culminate the BD challenges and associated BDA methods to allow signposting to take place. Following the earlier limited normative

research studies conducted by Polato, Ré, Goldman, and Kon (2014) – mainly focusing on Apache Hadoop; Frehe, Kleinschmidt, and Teuteberg (2014) – BD logistics; Eembi, Ishak, Sidi, Affendey, and Mamat (2015) – on data veracity research for profiling digital news portal, and Abdellatif, Capretz, and Ho (2015) – on software analytics (a distinct branch of BDA), this paper attempts to *broaden the scope of their reviews by further investigating and assessing the different types of BD challenges and the analytical methods employed to overcome the challenges*. Although these research studies provide worthy understanding on some aspects of BD and BDA area, there seems to be a lack of comprehensive and methodical approaches to understand the phenomenon of BD – more precisely the types of BDA methods thus an aide memoir will act as a suitable frame of reference. Moreover, explicitly in respect of the conclusions offered by these existing review articles, this research specifically aims to:

analyze, synthesize and present a state-of-the-art structured analysis of the normative literature on big data and big data analytics to support the signposting of future research directions.

1.2. Academic challenge

This SLR research aims to evaluate the existing research published on BD and BDA by employing an established profiling approach and to investigate and analyze different BD challenges and BDA technologies, techniques, methods and or approaches. To identify the relevant articles through the Scopus database, the following keywords search criteria was used:

- *Big Data OR Big Data Analytics OR Big Data Analysis AND Challenge OR Challenges OR Barrier OR Barriers OR Obstacle OR Obstacles OR Problem OR Problems OR Impediment OR Impediments AND Technology OR Technologies OR Technique OR Method OR Methods OR Approach OR Approaches.*

Through using the abovementioned list of keywords and focusing on four subject areas that is *business and management, computer science, decision science, and social science*; initially 433 journal articles were identified from the Scopus database and relating to articles published during the period from 1996 to 2015. However, from period 1996 until 2002, there were no papers recorded on BD and BDA in these four subject areas. After assessing the 434 articles (from refereed journals), 206 papers were discarded, and finally 227 papers were selected and taken forward for further interrogation. As reflected in Fig. 9, contributors from across the world have made contributions to the BD and BDA area. Nevertheless, given the limitations in the existing BD and BDA literature review studies (as reported earlier in Section 1.1), the rationale for undertaking this research is to provide a systematic state-of-the-art literature analysis of the BD and BDA area. In doing so to better understand the different types of BD challenges and associated BDA methods. Thus, the two underlying academic challenges orientate around identify the:

- different types of BD challenges theorized/proposed/discussed/confronted by organizations.
- different types of BDA methods theorized/proposed/discussed/employed to overcome BD challenges.

To supplement this research and the above objectives, the authors also identified the:

- yearly publications from 1996 until 2015.
- geographic location of each publication (this includes the geographical location of each author as well as the co-author(s) in each paper reviewed).

Download English Version:

<https://daneshyari.com/en/article/5109820>

Download Persian Version:

<https://daneshyari.com/article/5109820>

[Daneshyari.com](https://daneshyari.com)