Journal of Business Research 70 (2017) 263–286

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 10 August 2016

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.

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).

http://dx.doi.org/10.1016/j.jbusres.2016.08.001

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).

- types of publication (e.g. research or technical paper, literature review, viewpoint).
- types of research methods employed (e.g. case study, mixed method, analytical).

This type of profiling research is necessary to develop an understanding of the BD and BDA area and the state-of-the-art growth in the theory and application of BD and BDA within different sectors and disciplines. This paper is predominantly descriptive and inductive in nature, as the authors were interested in understanding the perspectives of BD and BDA and its distinctiveness as practiced across different sectors.

2. A normative perspective of Big Data: challenges and analytical methods

The concept of *big* is problematic to pinpoint, not least because a dataset that appears to be massive today will almost surely appear small in the near future (MIT Technology Review, 2013). Adding to the complexity of the BD itself, some practitioners argue that massive datasets are not always complex and small data sets are always simple, thus highlighting that the intricacy of a dataset is a significant factor in determining whether it is *big*. In this section, the authors provide some theoretical conceptions related to Q1 and Q2.

2.1. Big Data Challenges – related to Q1

Though the benefits of BD are factual and substantial, there remain a plethora of challenges that must be addressed to fully realise the potential of BD. Some of these challenges are a function of the characteristics of BD, some, by its existing analysis methods and models, and some, through the limitations of current data processing system (Jin, Wah, Cheng, & Wang, 2015). Extant studies surrounding BD challenges have paid attention to the difficulties of understanding the notion of BD (Hargittai, 2015), decision-making of what data are generated and collected (Crawford, 2013), issues of privacy (Lazer et al., 2009) and ethical considerations relevant to mining such data (Boyd & Crawford, 2012). Tole (2013) asserts that building a viable solution for large and multifaceted data is a challenge that businesses are constantly learning and then implementing new approaches. For example, one the biggest problems regarding BD is the infrastructure's high costs (Wang & Wiebe, 2014). Hardware equipment is very expensive even with the availability of cloud computing technologies.

Furthermore, to sort through data, so that valuable information can be constructed, human analysis is often required. While the computing technologies required to facilitate these data are keeping pace, the human expertise and talents business leaders require to leverage BD are lagging behind, this proves to be another big challenge. As reported by Akerkar (2014) and Zicari (2014), the broad challenges of BD can be grouped into three main categories, based on the data life cycle: data, process and management challenges:

- *Data challenges* relate to the characteristics of the data itself (e.g. data volume, variety, velocity, veracity, volatility, quality, discovery and dogmatism).
- *Process challenges* are related to series of how techniques: how to capture data, how to integrate data, how to transform data, how to select the right model for analysis and how to provide the results.
- *Management challenges* cover for example privacy, security, governance and ethical aspects.

Fig. 1 shows the classification of BD challenges – as adapted from Akerkar (2014) and Zicari (2014). The SLR findings for Q1 are based on three categories of BD challenges.

2.2. Big Data analytical methods – related to Q2

To facilitate evidence-based decision-making, organizations need efficient methods to process large volumes of assorted data into meaningful comprehensions (Gandomi & Haider, 2015). The potentials of using BD are endless but restricted by the availability of technologies, tools and skills available for BDA. According to Labrinidis and Jagadish (2012), BDA refers to methods used to examine and attain intellect from the large datasets. Thus, BDA can be regarded as a sub-process in the whole process of *insight extraction* from BD. It is certain that for BD to realise its objectives and progress services in business environment, it requires the correct tools and approaches to be analyzed and classified effectively and proficiently (Al Nuaimi, Al Neyadi, Mohamed, & Al-Jaroodi, 2015). The potential value of BD is solved simply when leveraged to the drive decision-making process. Extant research studies have demonstrated that substantial value and competitive advantage can be attained by businesses from taking effective decisions based on data (Davenport & Harris, 2007). But, BDA is more perplexing than merely tracing, classifying, comprehending, and quoting data. Davenport and Dyché (2013) emphasize that large organizations regularly gather BD and exploit analytics for support in decision-

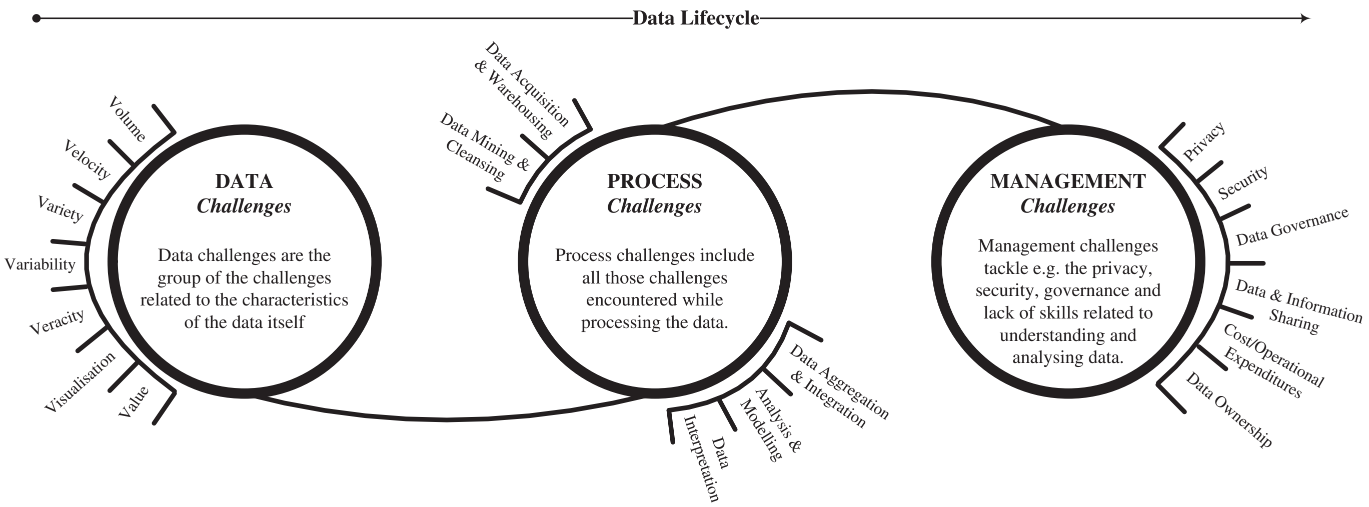
Data Lifecycle
Volume
Velocity
Variety
Variability
Veracity
Visualisation
Value
DATA
Challenges
Data challenges are the group of the challenges related to the characteristics of the data itself
Data Acquisition & Warehousing
Data Mining & Cleansing
PROCESS
Challenges
Process challenges include all those challenges encountered while processing the data.
Data Aggregation & Integration
Analysis & Modeling
Data Interpretation
MANAGEMENT
Challenges
Management challenges tackle e.g. the privacy, security, governance and lack of skills related to understanding and analysing data.
Privacy
Security
Data Governance
Data & Information Sharing
Cost/Operational Expenditures
Data Ownership

Fig. 1. Conceptual classification of BD challenges.

making as part of their usual procedures, and SMEs are the ones presently struggling to enhance top management decisions while adding more data for the analysis process. Aligning the people, technology, and organizational resources to become a data-driven company is problematic (Weill & Ross, 2009). Given BD can enhance the decision-making and increase organizational output; this is possible when a selection of analytical methods is used to extract sense from the data, such as:

- *descriptive analytics* scrutinizes data and information to define the current state of a business situation in a way that developments, patterns and exceptions become evident, in the form of producing standard reports, ad hoc reports, and alerts (Joseph & Johnson, 2013);
- *inquisitive analytics* is about probing data to certify/reject business propositions, for example, analytical drill downs into data, statistical analysis, factor analysis (Bihani & Patil, 2014);
- *predictive analytics* is concerned with forecasting and statistical modelling to determine the future possibilities (Waller & Fawcett, 2013);
- *prescriptive analytics* is about optimization and randomized testing to assess how businesses enhance their service levels while decreasing the expenses (Joseph & Johnson, 2013); and
- *pre-emptive analytics* is about having the capacity to take precautionary actions on events that may undesirably influence the organizational performance, for example, identifying the possible perils and recommending mitigating strategies far ahead in time (Szongott, Henne, & von Voigt, 2012).

Advocates assert that these types of analytical methods support in improved decision-making and organizational performance by making everything more translucent and quantifiable, while further uncovering inconsistencies as well as potential concerns and opportunities. Fig. 2 illustrates the classification of BDA methods and the SLR findings for Q2 are based on these five categories.

3. Research methodology

In an attempt to better understand and provide more detailed insights to the phenomenon of big data and bit data analytics, the authors respond to the special issue call on *Big Data and Analytics in Technology and Organizational Resource Management (specifically focusing on conducting – A comprehensive state-of-the-art review that presents Big Data Challenges and Big Data Analytics methods theorized [in extant research literature], proposed [by research scholars], and employed [by organizations])* through a SLR methodology as opposed to narrative or descriptive reviews (Tranfield, Denyer, & Smart, 2003; Kitchenham & Charters, 2007; Wang, Gunasekaran, Ngai, & Papadopoulos, 2016). In support of the former approach, Lettieri, Masella, and Radaelli (2009) report that SLR is a rational, transparent and reproducible research methodology for the analysis of extant literature. Kitchenham and Charters (2007) also highlight that SLR is a form of secondary study and it is a distinct approach to establish, explore and deduce accessible proof associated to a particular research question (e.g. Q1 and Q2) in a way that is unprejudiced and (to a certain degree) repeatable. Alternatively, meta-based-approaches can be used to conducting a literature review and include the work of Mishra, Gunasekaran, Papadopoulos, and Childe (2016), which adopt a bibliometric and network analysis approach to obtain and compare influential work in a specific domain (in this example, Big Data in Supply Chains).

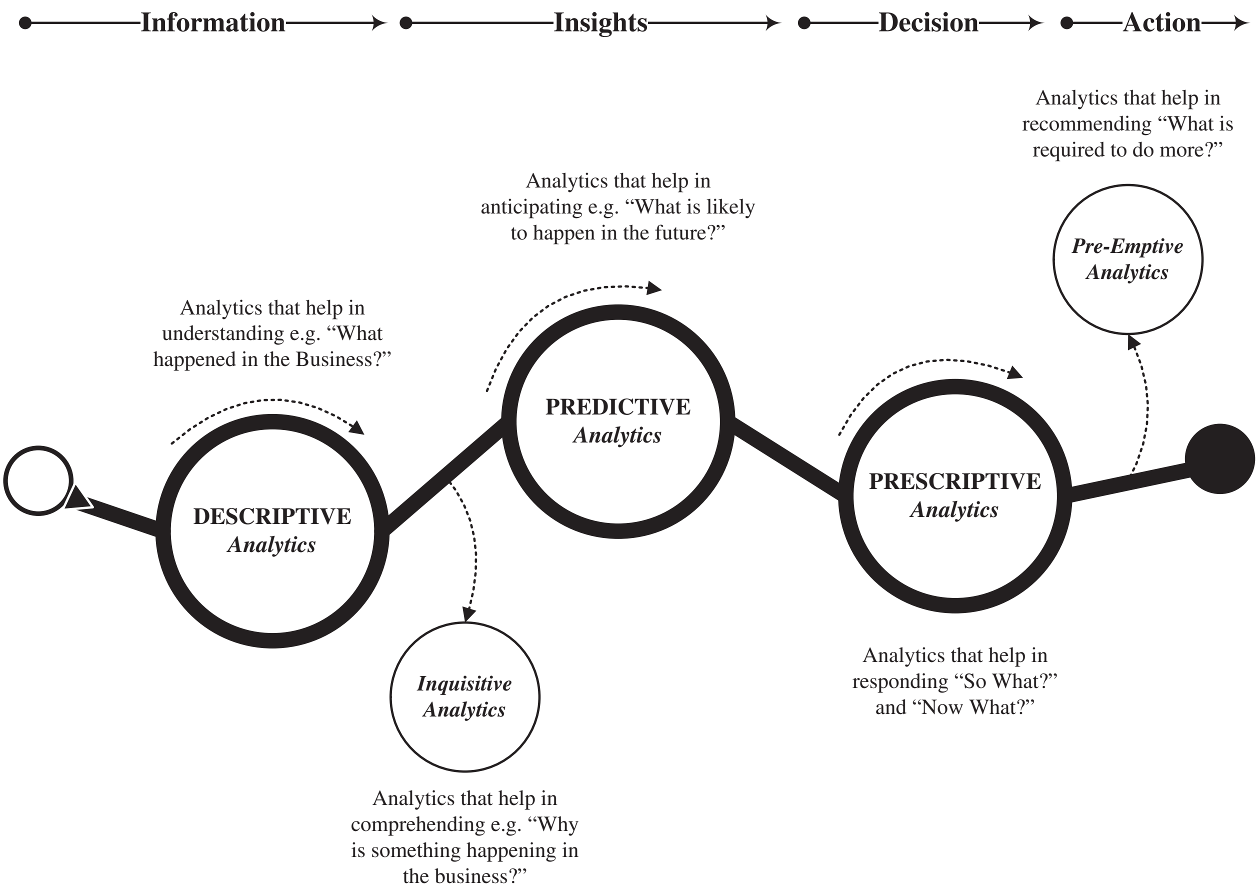
Information
Insights
Decision
Action
Analytics that help in recommending "What is required to do more?"
Analytics that help in anticipating e.g. "What is likely to happen in the future?"
Pre-Emptive Analytics
Analytics that help in understanding e.g. "What happened in the Business?"
PREDICTIVE Analytics
PRESCRIPTIVE Analytics
DESCRIPTIVE Analytics
Inquisitive Analytics
Analytics that help in responding "So What?" and "Now What?"
Analytics that help in comprehending e.g. "Why is something happening in the business?"

Fig. 2. Classification of types of big data analytical methods.

There are several motivating reasons for conducting a systematic literature review (Kitchenham & Charters, 2007) such as including among other, these are to:

- précis current evidence around a technology or a treatment, such as to summarize the evidence of the benefits and drawbacks of an explicit map technique;
- determine research gaps within the extant research to propose areas for further research activities;
- recommend a frame of reference to identify current research trajectories and potential research themes.

Based on the focus of this research, the first two reasons fit the purpose of a SLR. The scope and applicability of BD and BDA phenomenon clearly indicates that this area has the potential to support organizations, for instance, at the strategic, organizational, operational as well as technological level. This SLR offers an enhanced descriptive and thematic awareness of the resulting body of knowledge, enabling the BD research area to further develop in a more cognizant and multidisciplinary approach.

Delbufalo (2012) asserts that a SLR is designed to: (a) support in generating a sense of joint effort, importance and openness between the research studies in order to impede unproductive recurrence of effort, (b) support in connecting potential research to the queries and issues that have been modelled by previous research studies (e.g. most of those paper reviewed as part of this research exercise) and, (c) develop the approaches employed to assemble and synthesize preceding pragmatic evidence. In the interest of parsimony, a meticulous though not exhaustive SLR was carried out through following a three-phase approach as described by Tranfield et al. (2003) and Kitchenham and Charters (2007) and diagrammatically illustrated in Fig. 3:

- *Phase I* – Planning the Review Process – Defining the research aim and objectives (I.1); formulating the proposal (I.2) and developing the review protocol (I.3);
- *Phase II* – Conducting the Review Process – Identifying, selecting, evaluating and synthesizing the pertinent research studies; and
- *Phase III* – Reporting and Dissemination of the Overall Research Results – Descriptive reporting of results and thematic reporting of journal articles.

Following the three-phase approach, the next subsection 3.1 summarizes the research protocol (Phase I.3) as the defining of the aim and objectives including the proposal (I.1 and I.2) have already been presented in the introduction (under subsections 1.1 and 1.2). Sub-section 3.2 describes the Scopus database searching process of the relevant articles (Phase II). Finally, the reporting and dissemination the overall results (Phase III) will be discussed in Section 4 and with Section 5 concluding the paper.

3.1. The research protocol (phase I.3)

In this paper, the authors commenced this systematic search by using an established detailed review protocol based on the guiding principles and procedures of the SLR (Tranfield et al., 2003; Kitchenham & Charters, 2007). This protocol identifies the background review, search strategy, research questions as outlined in the abstract (i.e. Q1 and Q2), data extraction, criteria for study selection and data synthesis – based on the prescriptive three phased approach. The research questions and background of this review are described above, while the following sections provide details about other elements. As this literature review focuses on *analysing*, *synthesizing* and *presenting* a comprehensive structured analysis of the normative literature on BD and BDA, it was necessary to considered the domains for this research synthesis as both conceptual and empirical (including qualitative, quantitative and mixed method) papers. The research protocol for this literature review paper provides details on the following two points, as also followed by Delbufalo (2012) and Kamal and Irani (2014):

- Point I – Conceptualizing BD and BDA research discipline, including challenges of BD and related BDA methods (as discussed in Sections 2.1, and 2.2).
- Point II – Typology of research studies to be considered in this review exercise and the appropriate measures.

Given the above, several selections in relation to the typology of research studies to be counted in and the suitability conditions (i.e. the inclusion and exclusion measures) have been made (Point II), as presented below.

- *Condition I* – The review was conducted by searching the Scopus databases. The reason for choosing the Scopus database was that it covers nearly 18,000 plus titles from over 5000 international publishers, including coverage of around 16,500 peer-reviewed journals on different areas. Therefore, it is possible to search for and locate a significant proportion of the published material in the BD and BDA area.
- *Condition II* – To focus on enhancing quality control (David & Han, 2004) only published peer-reviewed journal (including articles in press and therefore accepted post peer-review) were considered by selecting *Article* and *Articles in Press* option from the Document Type option. Other document/source types such as conferences, trade publications, books series, book or book chapter, and editorials were omitted.
- *Condition III* – Following David and Han's (2004) enhancing quality control policy, only those articles were selected that were published between 1996 and 2015.
- *Condition IV* – Articles from subject areas such as Business and Management, Computer Sciences, Decision Sciences and Social Sciences and published in the English language were only selected, excluding the articles published in Chinese, French, Korean, Spanish, German, Japanese, Portuguese and Russian. This is a recognised limitation.
- *Condition V* – It was ensured that the selected articles were not only empirical (i.e. case-study, results, analytical, etc.) but also those articles that were essentially conceptual so as to identify conceptual research developments in BD and BDA.
- *Condition VI* – Articles' applicability was confirmed by requiring that selected articles contained a number of key phrases (as listed in Section 1.2) throughout the paper, including, title, abstract, keywords and the thereafter the whole paper. In essence, the identified articles were reviewed with particular attention given to those section(s) that explicitly referred to BD and BDA. In doing so, to extract relevant perspectives on the type of BD challenges and BDA methods.
- *Condition VII* – Final substantive applicability was confirmed by reading the remaining whole article for essential research perspective and satisfactory empirical data. The latter process *forced* the alignment between the selected articles and the research review objectives.

The abovementioned conditions itemized in seven points were all prescriptively followed so as to conduct an effective and reproducible database examining process as pronounced in the following subsection.

3.2. Scopus database searching process and results – Phase II

According to Delbufalo (2012), there are four stages of database searching process. This section reports on these steps and activities of the process, demonstrating the outcomes both descriptively and synthetically by searching for relevant articles throughout the Scopus database.

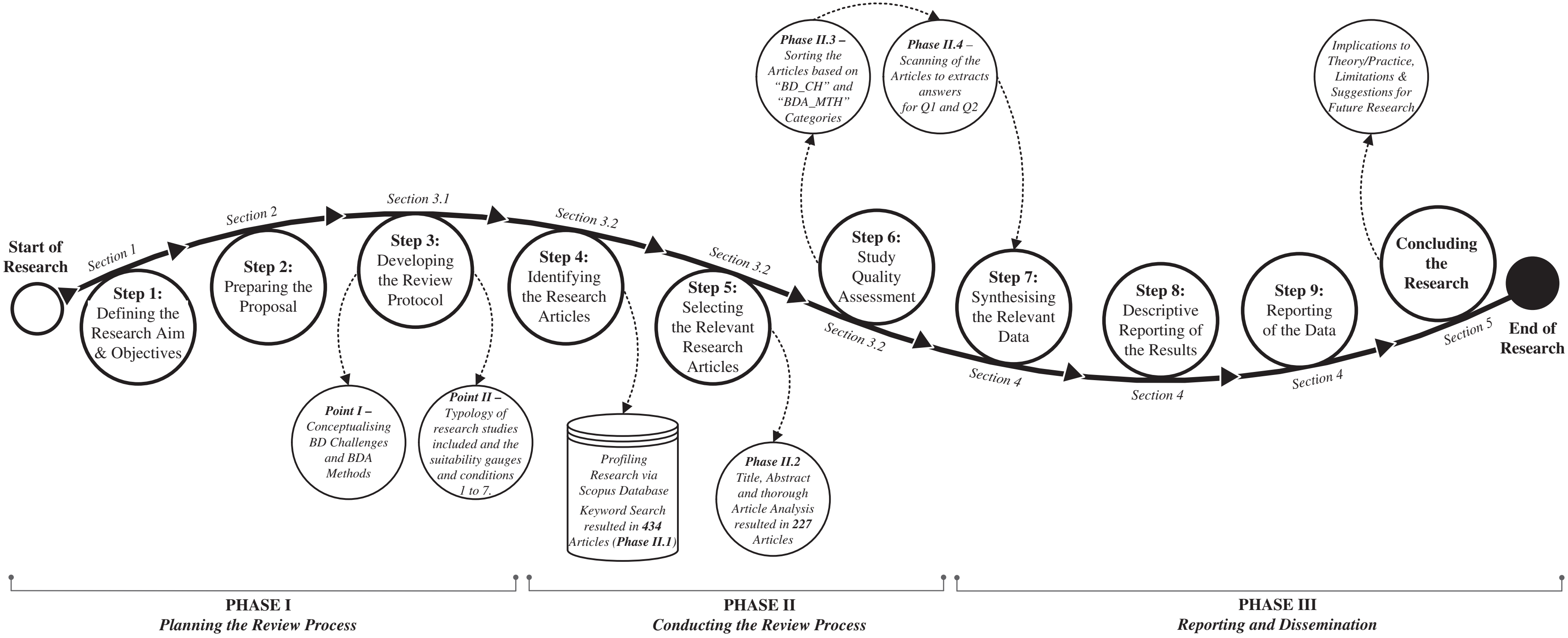
Start of Research
Section 1
Step 1: Defining the Research Aim & Objectives
Section 2
Step 2: Preparing the Proposal
Section 3.1
Step 3: Developing the Review Protocol
Point I – Conceptualising BD Challenges and BDA Methods
Point II – Typology of research studies included and the suitability gauges and conditions 1 to 7.
Section 3.2
Step 4: Identifying the Research Articles
Profiling Research via Scopus Database
Keyword Search resulted in 434 Articles (Phase II.1)
Section 3.2
Step 5: Selecting the Relevant Research Articles
Phase II.2 Title, Abstract and thorough Article Analysis resulted in 227 Articles
Section 3.2
Step 6: Study Quality Assessment
Phase II.3 – Sorting the Articles based on "BD_CH" and "BDA_MTH" Categories
Phase II.4 – Scanning of the Articles to extracts answers for Q1 and Q2
Section 4
Step 7: Synthesising the Relevant Data
Section 4
Step 8: Descriptive Reporting of the Results
Section 4
Step 9: Reporting of the Data
Concluding the Research
Implications to Theory/Practice, Limitations & Suggestions for Future Research
Section 5
End of Research
PHASE I
Planning the Review Process
PHASE II
Conducting the Review Process
PHASE III
Reporting and Dissemination

Fig. 3. Research design – systematic literature review process of BD and BDA.

- *Phase II.1* – A number of keywords were entered into the Scopus database (as stated in Section 1.2) following conditions 2, 3 and 4 in Section 3.1. This process resulted in 2360 publications, of which 433 were left as relevant after filtering according to the barring conditions.
- *Phase II. 2* – A title, abstract and thorough article analysis was thereafter conducted on the extracted articles based on conditions 5 and 6. Some further articles (i.e. 206) were discarded during this stage. At the end of this process, 227 articles were considered for further investigation.
- *Phase II.3* – For this step, the authors followed the quality criteria matrix as adopted by Pittaway et al. (2004). In this step, the selected 227 articles were further scanned, searching for both conceptual as well as empirical studies through the criteria highlighted in conditions 6 and 7. By doing so, all articles were grouped into two categories (i.e. BD_CH refers to BD challenges and BDA_MTH refers to BDA methods:
  - ○ Category BD_CH was defined to incorporate all the studies as certainly pertinent because each article either reported or discussed or evaluated the BD challenges. So for this category all the 227 papers resulted as productive.
  - ○ Category BDA_MTH was defined for those studies that were relevant for extracting information on the types of BDA methods discussed/proposed. After thoroughly analysing the 227 articles, around 115 articles discussed or proposed some form of method for BDA.

As a result of the above two categories, all 227 articles were considered applicable for responding to Q1 and Q2. The applicability assessment was considered as relative, to the degree that the authors' decrees were focused on facets defined within the scope of the review process.

- *Phase II.4* – Herein, beginning within the BD_CH category and followed by BDA_MTH category, the full-text version of 227 articles were thoroughly read by the first and second author. In order to save time, both the authors divided the articles among themselves and reviewed them for BD_CH (i.e. here the authors thoroughly reviewed the articles to identify the different types of BD challenges – either theorized/proposed/discussed/confronted by different sector organizations), and BDA_MTH (i.e. here the authors examined the articles thoroughly to identify the different types of methods discussed, proposed and or employed by organizations to overcome BD challenges), so as to confirm substantive relevance both conceptually and empirically as mentioned in conditions 6 and 7. In order to respond to each Q1 and, Q2 questions, we reviewed each paper to identify the BD challenges (Q1) and BDA methods (Q2) at the same time and noted the findings on a spread sheet.

This latter analysis was conducted descriptively, using a standard template adapted from the works of Delbufalo (2012). This descriptive investigation also produced graphs and tables designed to contain the yearly publications, geographic region of the first author and co-author(s), type of publications, and research methods employed, for all 227 articles.

4. Big Data and Big Data Analytics: findings and analysis

The findings of this study are now presented under different subsections. Each of the six subsections discuss on the findings in relation to a particular variable as set in Section 1.2.

4.1. Types of Big Data Challenges

Among the many BD challenges (as reported in Figs. 1 and 4), the large datasets (in terms of size and complexity) and the ability to process vast amount of data remains a critical challenge for outdated data processing applications and, relational database management systems (Jiang et al., 2015). According to TDWI Predictive Analytic Study (Russom, 2013), there are several BD challenges posing a peril to organizations – among these are, integrating complex and large datasets, getting started with the right BD project, developing and implementing infrastructure for managing and processing BD and a lack of skilled personnel or staff with analytics skills to make sense of BD.

Figs. 4, 5, and 6 illustrates the frequency at which the data, process and management (all three related to BD) challenges are discussed/proposed/theorized in the articles reviewed through the SLR process, as presented in Fig. 1.

4.1.1. Data challenges

Data challenges are the group of the challenges related to the characteristics of the data itself. Different researchers have distinct understandings towards the data characteristics – such as some say 3Vs [volume, velocity and variety] of data (e.g. Shah, Rabhi, & Ray, 2015), others reported 4Vs [volume, velocity, variety, and variability] of data (e.g. Liao, Yin, Huang, & Sheng, 2014) and 6Vs [volume, velocity, variety, veracity, variability, and value] of data (Gandomi & Haider, 2015). In analysing the different articles reviewed in this SLR, the authors identified 7Vs – seven characteristics of data [volume (DC_VOLM) → C = 90 (39.64% of 227 articles), variety (DC_VART) → C = 59 (25.9%), veracity (DC_VERT) → C = 44 (19.4%), value (DC_VALE) → C = 30 (13.2%), velocity (DC_VELO) → C = 18 (7.9%), visualization (DC_VISU) → C = 6 (2.6%) and variability (DC_VARB) → C = 4 (1.8%)] and these features are illustrated in Fig. 4 and discussed as follows:

- *Volume* (e.g. *large data-sets consisting of terabytes, petabytes, zettabytes of data – or even more*): Large scale and the sheer volume of data is a big challenge in its own right. The latter argument is also supported by Barnaghi et al. (2013) that state the heterogeneity, ubiquity, and dynamic nature of the different data generation resources and devices, and the enormous scale of data itself, make determining, retrieving, processing, integrating, and inferring the physical world data (e.g. environmental data, business data, medical data, surveillance data) a challenging task. This colossal increase of large-scale data (e.g. *Facebook daily generates over 500 terabytes of data, and Walmart collects more than 2.5 petabytes of data every hour from its customer transactions*) sets brings new challenges to data mining techniques and requires novel approaches to address the big-data problem (Zhao, Zhang, Cox, Duling, & Sarle, 2013).
- *Variety* (e.g. *multiple data formats with structured and unstructured text/image/multimedia content/audio/video/sensor data/noise*): Data challenges related to the variety (i.e. diverse and dissimilar forms) of data are also deemed a challenge. These articles revealed that the enormous volume of data is not consistent nor does it follow a specific template or format – it is captured in diverse forms and diverse sources e.g.: messages (text, email, tweets, blogs) – user generated content, transactional data (e.g. web logs, business transactions), scientific data (e.g. data coming from data-intensive experiments – genome and healthcare data), web data (e.g. images posted on social media; sensor data readings), and much more (Chen, Chiang & Storey, 2012; Chen et al., 2013). These different forms and quality of data clearly indicate that heterogeneity is a natural property of BD and it is a big challenge to comprehend and manage such data (Labrinidis & Jagadish, 2012). For instance, during the Fukushima Daiichi nuclear disaster, when the public started broadcasting radioactive material data, a wide variety of inconsistent data, using diverse and uncalibrated devices, for similar or neighboring locations was reported – all this add to the problem of increasing variety of data.
- *Veracity* (e.g. *increasingly complex data structure, anonymities, imprecision or inconsistency in large data-sets*): This is not merely about data quality – it is more about understanding the data, as there are integral discrepancies in almost all the data collected. IBM came up with this characteristic of data, which represents the

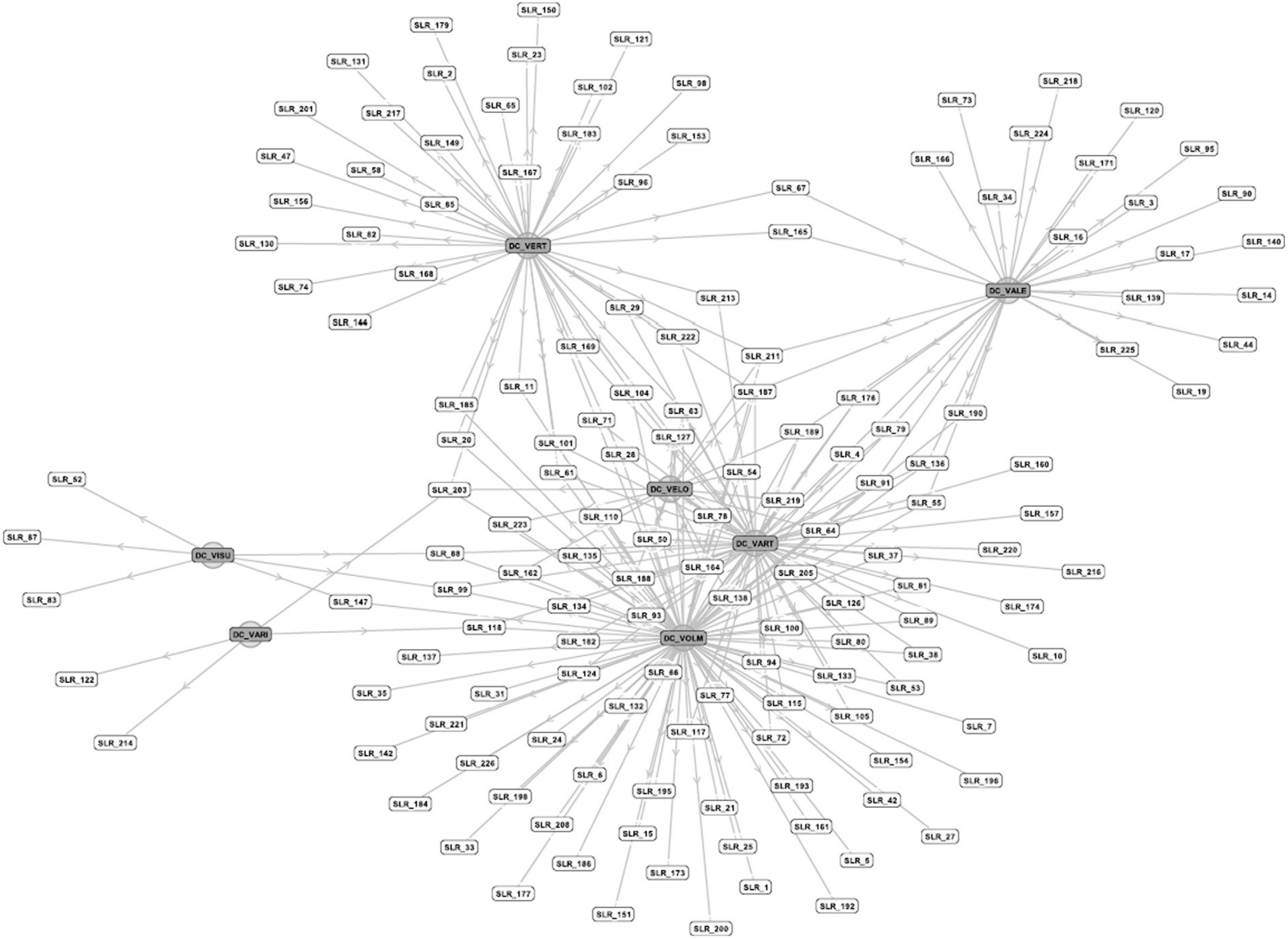

Fig. 4. Clusters of articles discussing/proposing/theorizing types of data challenges.

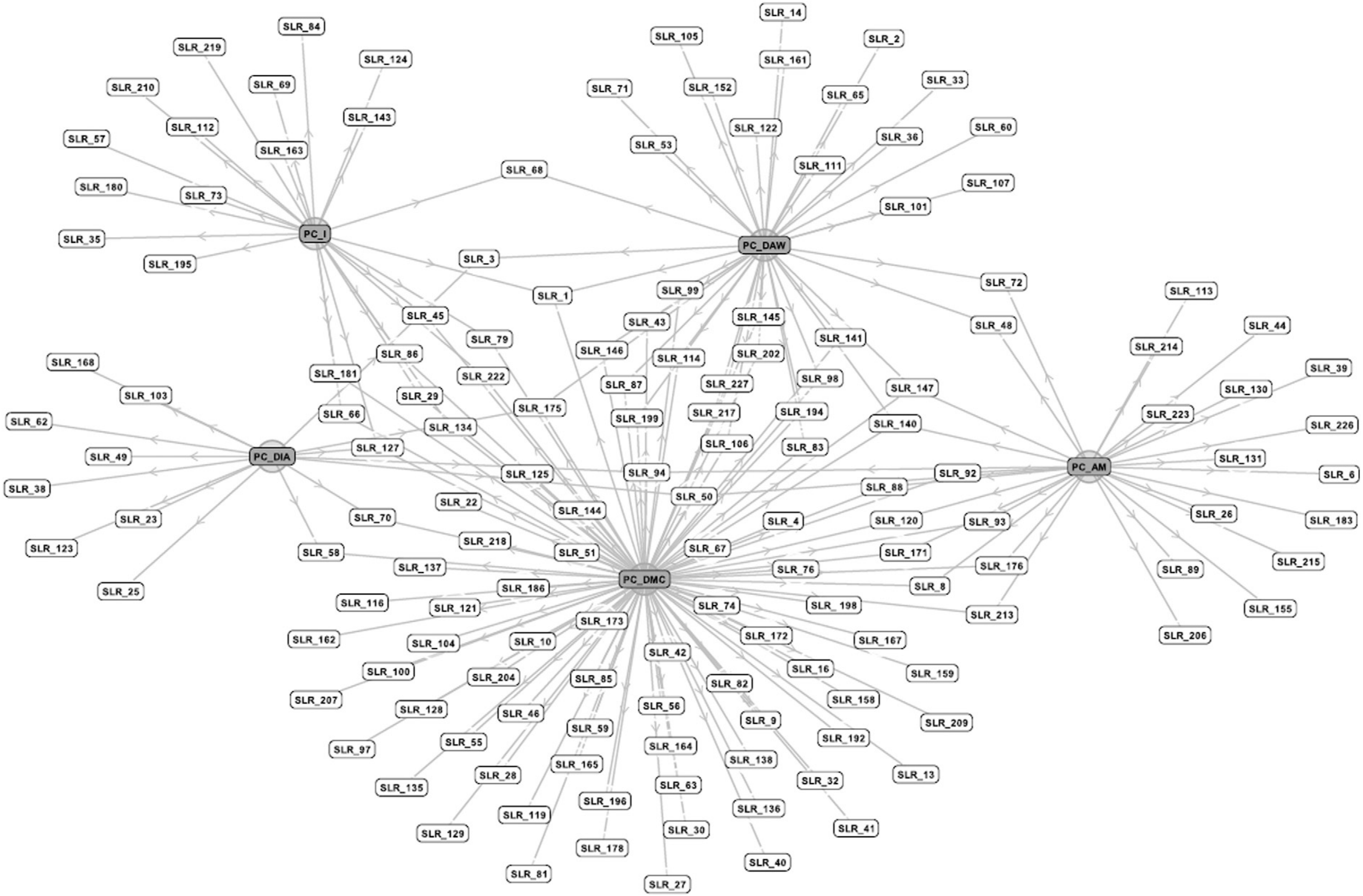

Fig. 5. Clusters of articles discussing/proposing/theorizing types of process challenges.

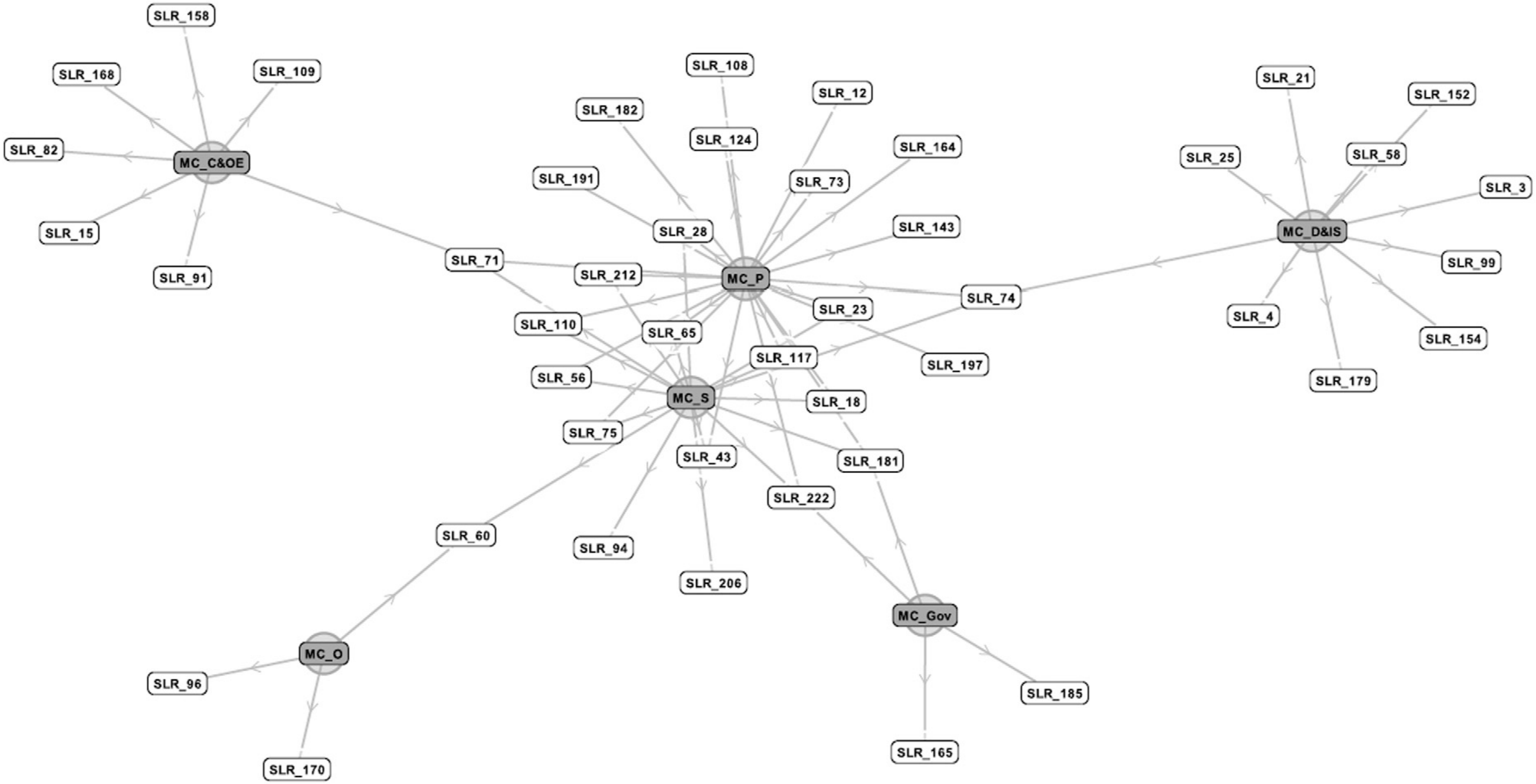

Fig. 6. Clusters of articles discussing/proposing/theorizing on types of management challenges.

untrustworthiness inherent in many sources of structured as well as unstructured data. Akerkar (2014) and Zicari (2014) refer veracity to as coping with the biases, doubts, imprecision, fabrications, messiness and misplaced evidence in the data. Veracity feature measures the accuracy of data and its potential use for analysis (Vasarhelyi, Kogan, & Tuttle, 2015). For instance, every customer opinion on different social media networks and web is different and unclear in nature, as it involves human interaction (Sivarajah, Irani, & Weerakkody, 2015). Moreover, the web, more specifically, is a soft medium to publish and broadcast fabricated information across multiple sources and, so it is essential to isolate the wheat from the chaff when presenting quality data. Thus, the necessity to deal with inaccurate and ambiguous data is another facet of BD, which is addressed using tools and analytics developed for management and mining of unreliable data (Gandomi & Haider, 2015).

- *Velocity* (*e.g. high rate of data inflow with non-homogenous structure*): The challenge of velocity comes with the requisite to manage the high influx rate of non-homogenous data, which results in either creating new data or updating the existing data (Chen et al., 2013). This mainly applies to those datasets that are generated through large complex networks including data generated by the proliferation of digital devices, which are positioned ubiquitously resulting in driving the need for real-time analytics and evidence-based planning (Lu, Zhu, Liu, Liu, & Shao, 2014). For instance, Wal-Mart processes more than a million transactions each hour (Cukier, 2010). The data stemming from mobile devices and flowing through mobile apps or by using store cards (e.g. Sainsbury's card for collecting nectar points) generates floods of information that can be brought to use through producing real-time, *personalized* offers for customers. These data also provide sound information about customers, such as their geospatial location, buying behaviour and patterns, which can be analyzed in real-time to generate value for customers (Gandomi & Haider, 2015).
- *Variability* (*e.g. data whose meaning is constantly changing*): Among the seven pillars of BD, variability is another extremely essential feature but is often confused with variety. For instance, Google or Facebook repository stores and generates many different types of data. At the same time, if from these different types of data, one of them is brought to use for mining and making sense out of it but every time the data offers a different meaning – this is variability of data – whose meaning is constantly and rapidly changing. The volumes of machine and human-generated data constitute much greater and their rates of change and variability higher than process-mediated data. Variability is also related in performing sentiment analyzes. For example, in (almost) the same tweets a word can have a totally different meaning. In order to perform a proper sentiment analyzes, advocates assert that algorithms need to be able to understand the context and be able to decipher the exact meaning of a word in that context (Zhang, Hu et al., 2015). Nevertheless, this is yet still very challenging.
- *Visualization* (*e.g. presenting the data in a manner that is readable*): Visualising data is about representing key information and knowledge more instinctively and effectively through using different visual formats such as in a pictorial or graphical layout (Taheri, Zomaya, Siegel, & Tari, 2014). For instance, eBay has millions of users and from these many million users, even more millions of goods are sold every month – this generates a lot of data. To make all these data explicable, eBay considered the BD visualization tool – Tableau, which is capable of transforming large and complex datasets into spontaneous depictions. Based on these interactive results, eBay employees can visualize search relevance and quality, to monitor the latest customer feedback and conduct sentiment analysis. Chen and Zhang (2014) argue that for many existing BD applications that have poor performances in functionalities, scalability and response time, it is mainly problematic when conducting data visualization. This reason for this is a consequence of large sizes and high dimension of BD.
- *Value* (e.g. *extracting knowledge/value from vast amounts of structured and unstructured data without loss, for end users*): Storing BD is complex. For instance, significant values can be extracted from the stream of clicks left behind by the internet users – and this is becoming a backbone of the internet economy. Big data researchers consider value as an essential feature, as somewhere within that data, there is valuable information – extracting golden data (high-valued data), though most of the pieces of data independently may seem insignificant (Zaslavsky, Perera, & Georgakopoulos, 2012). Regardless of the number of dimensions used to describe BD, organizations are still faced with challenges of storing, managing and predominantly extracting value from the data in a cost effective manner (Abawajy, 2015).

4.1.2. Process challenges

Process challenges are the group of challenges encountered while processing and analysing the data that is from capturing the data to interpreting and presenting the end results. As large datasets are usually non-relational or unstructured, thus processing such semi-structured data sets at scale poses a significant challenge; possibly more so than managing BD (Kaisler, Armour, Espinosa, & Money, 2013). In analysing the different articles reviewed the authors identified several data processing related challenges that can be grouped into 5 steps that is data acquisition and warehousing (PC_DAW) → C = 97 (42.7%), data mining and cleansing (PC_DMC) → C = 38 (16.7%), data integration and aggregation (PC_DAI) → C = 29 (12.8%), data analysis and modelling (PC_DAM) → C = 25 (11%) and data interpretation (PC_DI) → C = 15 (6.6%). As illustrated in Fig. 5, data mining and cleansing appears to be a vital step during processing the large scale unstructured data, as 97 articles out of 227 specifically discussed and highlighted the importance of this step.

- *Step 1 – Data Acquisition and Warehousing*: This challenge is related to acquiring data from diverse sources and storing for value generation purpose. The integral complexity of BD and exponentially growing demands develop unprecedented problems in BD engineering such as data acquisition and storage (Wang & Wiebe, 2014). The latter argument is supported by Paris, Donnal, and Leeb (2014) who assert that one of the prime barriers to the analysis of BD arises from a lack of data provenance, knowledge and discrepancies of scale inherent in data collection and processing. This further restricts the speed and resolution at which data can be captured and stored. As a result, this affects the capability to excerpt actionable information from the data (Chen & Zhang, 2014). To capture related and valuable information, smart filters are required that should be robust and intelligent to capture useful information and discard useless that contains imprecisions or inconsistencies – this is a challenge in itself. For the latter, efficient analytical algorithms are required to understand the provenance of data and process the vast streaming data and to reduce data before storing (Zhang, Hu et al., 2015; Zhang, Liu et al., 2015).
- *Step 2 – Data Mining and Cleansing*: This challenge relates to extracting and cleaning data from a collected pool of large scale unstructured data. Advocates of BD and BDA perceive that in identifying a better way to mine and clean the BD can result in big impact and value (Chen, Chen et al., 2012). Due to its strident, vibrant, diverse, interrelated and unreliable features, the mining, cleansing and analysis proves to be very challenging (Chen et al., 2013). For instance, in the UK National Health Service (NHS) there are many millions of patients' records comprising of medical reports, prescriptions, x-ray data, etc. Physicians make use of such data – if for instance incorrect information is stored this may lead to physicians wrongly diagnosing conditions, resulting in inaccurate medical records. In order make use of this huge data in a meaningful way, there is a need to develop an extraction method that mines out the required information from unstructured BD and articulate it in a standard and structured form that is easy to understand. According to Labrinidis and Jagadish

(2012) developing and maintaining this extraction method is a continuous challenge.

- *Step 3 – Data Aggregation and Integration*: This process challenge relates to aggregating and integrating clean data mined from large unstructured data. BD often aggregates varied online activities such as tweets – retweets, microblogging, and *likes* on Facebook that essentially bear diverse meanings and senses (Edwards & Fenwick, 2015). This characteristically amorphous data naturally lacks any binding information. Aggregating these data evidently goes beyond the abilities of current data integration systems (Carlson et al., 2010). According to Karacapilidis, Tzagarakis, and Christodoulou (2013), the availability of data in large volumes and diverse types of representation, smart integration of these data sources to create new knowledge – towards serving collaboration and improved decision-making – remains a key challenge. Halevy, Rajaraman, and Ordille (2006) assert that the indecision and provenance of data are also a major challenge for data aggregation and integration. Another challenge relates to aggregated data in warehouses – in line with this argument, Lebdaoui, Orhanou, and Elhajji (2014) report that to enable decision systems to efficiently respond to the real world's demands, such systems must be updated with clean operational data.
- *Step 4 – Data Analysis and Modelling*: Once the data has been captured, stored, mined, cleaned and integrated, comes the data analysis and modelling for BD. Outdated data analysis and modelling centers around solving the intricacy of relationships between schema-enabled data. As BD is often noisy, unreliable, heterogeneous, dynamic in nature; in this context, these considerations do not apply to non-relational, schema-less databases (Shah et al., 2015). From the perspective of differing between BD and traditional data warehousing systems; Kune, Konugurthi, Agarwal, Chillarige, and Buyya (2016) report that although these two have similar goals; to deliver business value through the analysis of data, they differ in the analytics methods and the organization of the data. Consequently, old ways of data modelling no longer apply due to the need for unprecedented storage resources/capacity and computing power and efficiency (Barbierato et al., 2014). Thus, there is a need for new methods to manage BD for maximum impact and business value. It is not merely knowing about what is currently trendy, but also need to anticipate what may happen in the future by appropriate data analysis and modelling (Chen et al., 2013).
- *Step 5 – Data Interpretation*: This step is relatively similar to visualising data and making data understandable for users that is the data analysis and modelling results are presented to the decision makers to interpret the findings for extracting sense and knowledge (Simonet, Fedak, & Ripeanu, 2015). The astounding growth and multiplicity of unstructured data have intensely affected the way people process and interpret new knowledge from these raw data. As much of these data both instigate and reside as an online resource, one open challenge is defining how Internet computing technological solutions have evolved to allow access, aggregate, analyze, and interpret BD (Bhimani & Willcocks, 2014). Another challenge is the shortage of people with analytical skills to interpret data (Phillips-Wren & Hoskisson, 2015).

4.1.3. Management challenges

Management challenges related to BD are a group of challenges encountered, for example while accessing, managing and governing the data. Data warehouses store massive amounts of sensitive data such as financial transactions, medical procedures, insurance claims, diagnosis codes, personal data, etc. Organizations and businesses need to ensure that they have a robust security infrastructure that enables employees and staff of each division to only view relevant data for their department. Moreover, there must be some standard privacy laws that may govern the use of such personal information and strict observance to these privacy regulations must be applied in the data warehouse. In analysing the different articles reviewed in this SLR, the authors identified several data management related challenges that can be grouped into seven areas (Fig. 6) such as privacy (MC_P) → C = 23 (10.1%), security (MC_S) → C = 17 (7.5%), data and information sharing (MC_D&IS) → C = 10 (4.4%), cost/operational expenditures (MC_C&OE) → C = 7 (3.1%), data governance (MC_DG) → C = 4 (1.8%), and data ownership (MC_OG) → C = 3 (1.3%).

- *Privacy*: BD poses big privacy concerns and how to preserve privacy in the digital age is a prime challenges. Huge investments have been made in BD projects to streamline processes; however, organizations are facing challenges in managing privacy issues, and recruiting data analysts, thus hindering organizations in moving forward in their efforts towards leveraging BD (Krishnamurthy & Desouza, 2014). In a smart city environment where sensory devices gather data on citizen activities that can be accessed, several government and security agencies pose significant privacy concerns (Barnaghi et al., 2013). Among such privacy related challenges, location-based information being collected by BD applications and transferred over networks is resulting in clear privacy concerns (Yi et al., 2014). For example, location-based service providers can identify subscriber by tracking their location information – which is possibly linked to their office or residential information. Then there is the challenge of protecting privacy – Machanavajjhala and Reiter (2012) report that failure to protect citizens' privacy is illegal and open to relevant Government oversight bodies.
- *Security*: Security is a major issue and is identified by Lu et al. (2014) who argue that if security challenges are not appropriately addressed then the phenomenon of BD will not receive much acceptance globally. Securing BD has its own distinctive challenges that are not profoundly different to traditional data. Among the several BD related security challenges are the distributed nature of large BD which is complex but equally vulnerable to attack (Yi et al., 2014), malware has been an ever growing threat to data security (Abawajy, Kelarev, & Chowdhury, 2014), lack of adequate security controls to ensure information is resilient to altering (Bertot, Gorham, Jaeger, Sarin, & Choi, 2014), analysing logs, network flows, and system events for forensics and intrusion detection has been a challenge for data security (Cárdenas, Manadhata, & Rajan, 2013), lack of sophisticated infrastructure that ensures data security such as integrity, confidentiality, availability, and accountability, and data security challenges become magnified when data sources become ubiquitous (Demchenko, Grosso, De Laat, & Membrey, 2013).
- *Data Governance*: As the demand for BD is constantly growing, organizations perceive data governance as a potential approach to warranting data quality, improving and leveraging information, maintaining its value as a key organizational asset, and support in attaining insights in business decisions and operations (Otto, 2011). According to Intel IT Centre (2012), IT managers highly support the presence of a formal BD strategy, this especially makes sense, since the issue of data governance for describing what data is warehoused, analyzed, and accessed is termed as one of the three top challenges they face (besides data growth and data centre infrastructure and the ability to provide scalability). du Mars (2012) state that a significant challenge in the process of governing BD is categorizing, modelling and mapping the data as it is captured and stored, mainly due to the unstructured and complex nature of data. Moreover, effective BD governance is essential to ensure the quality of data mined and analyzed from a pool of large datasets (Hashem et al., 2015).
- *Data and Information Sharing*: Sharing data and information needs to be balanced and controlled to maximise its effect, as this will facilitate organizations in establishing close connections and harmonisation with their business partners (Irani, Sharif, Kamal, & Love, 2014). However, where organizations store large scale datasets that have

potential analysis challenges, it also poses an overwhelming task of sharing and integrating key information across different organizations (OSTP, 2012). Al Nuaimi et al. (2015) also state that sharing data and information between distant organizations (or departments) is a challenge. For instance, each organization and their individual departments typically own a disparate warehouse (developed based on different technological platforms and vendors) of sensitive information and several departments are often reluctant to share their patented data governed by privacy conditions. According to Khan, Uddin, and Gupta (2014) the challenge here is to ensure not to cross the fine line between collecting and using BD and guaranteeing user privacy rights. The is also related to a smart city environment that entails a plethora of sectors and in such context, smart city technological systems will need to reduce the barriers to achieve seamless information sharing and exchange among different entities (Su, Li, & Fu, 2011).

- *Cost/Operational Expenditures*: The constantly increasing data in all different forms has led to a rising demand for BD processing in sophisticated data centers. These are generally dispersed across different geographical regions to embed resilience and spread risk, for example Google having 13 data centers in eight countries spread across four continents (Gu, Zeng, Li, & Guo, 2015). The significant resources have been allocated to support the data intensive operations (i.e. acquisition, warehousing, mining and cleansing, aggregation and integration, processing and interpretation) – all this lead to high storage and data processing *big costs* (Raghavendra, Ranganathan, Talwar, Wang, & Zhu, 2008). Researchers assert that cost minimization is an emergent challenge (Irani, Ghoneim, & Love, 2006; Irani, 2010), with Gu et al., 2015 explaining the challenges of processing BD across geo-distributed data centers. Advocates of BD search for cost-effective and efficient ways to handle the massive amount of complex data (Sun, Morris, Xu, Zhu, & Xie, 2014). The cost of data processing and other operational expenditures of the data center are a sensitive issue that may also impact in the way organizations adopt and implement technological solutions (Al Nuaimi et al., 2015).
- *Data Ownership*: Besides privacy, Web (2007) asserts that ownership of data is a complex issue – as big as the data itself – while sharing real time data. Kaisler et al. (2013) also claim that data ownership presents a critical and continuing challenge, specifically in the social media context such as who owns the data on Facebook, Twitter or MySpace – are the users who update their status or tweet or have any account in these social networks (Sivarajah et al., 2015; Sivarajah, Irani, & Jones, 2014). It is generally perceived that both view they (the users and the social media provider) own the data. Kaisler et al. (2013) argues that this dichotomy still needs to be settled. With ownership arise the issue or controlling and ensuring its accuracy. For instance, Web (2007) states that sensor data is too sensitive and can result in mounting errors – this may further result in capturing and revealing inconsistent data – but then who owns that data. Data ownership is a much deeper social issue. These concerns are beyond the focus on several applications, for example SensorMaps by Web (2007) requires more research since they may have deep implications.

Like other data related management challenges, data ownership is essentially vital and its issues much be addressed to realise the promise of BD.

4.2. Types of Big Data analytical methods

BD comprising of large raw data set on its own does not offer a lot of value in its unprocessed form. If its [BD] potential value is to be unlocked, businesses need efficient processes and methods to turn high volumes of structured and unstructured data to analyze these raw datasets. Analytics in this context refers to the methods used to analyze and acquire intelligence from BD. As a result, BD analytics methods can be viewed as a sub-process within the overall process of insight extraction from BD. Despite the hype about varying BDA methods, using analytics is still a labour intensive undertaking. As Assunção, Calheiros, Bianchi, Netto, and Buyya (2015) highlight the reason for this is that current solutions for analytics are often based on proprietary appliances or software systems built for general purposes. As a result, organizations need to put in significant effort to customize such BDA solutions to their individual needs, which might require integrating different data sources and setting up the software on the organization's hardware.

In analysing the different articles reviewed in this SLR, a total of 115 papers out of the 227 papers analyzed discusses and proposes some form of BDA methods and tools. The extant literature highlights a number of analytical processes and methods – such as text analytics, audio analytics, video analytics, social media analytics, predictive analysis of data (Gandomi & Haider, 2015) and others reported of descriptive analytics, inquisitive analytics, prescriptive analytics and pre-emptive data analytics (Assunção et al., 2015; Rehman, Chang, Batool, & Teh, 2016). Within these various BD analytics methods, the SLR highlights that there are a number of off the shelf software tools [e.g. Hadoop, MapRecuce, Dyrad] (Chen, Chen et al., 2012; Chen, Chiang et al., 2012; Jiang et al., 2015), that have been built using and extending off-the-shelf existing software [e.g. Hadoop based e-book conversion system, MapReduce-based Big Data Processing on Multi-GPU systems)] (Jiang et al., 2015) and finally novel solutions to tackle BD analysis [e.g. DEMass – A New Density Estimator for Big Data] (Ting, Washio, Wells, Liu, & Aryal, 2013). In studying the analyzed papers, the authors identified and classified analytics methods into 3 groups – such as descriptive analytics, predictive analytics and prescriptive analytics; however, nothing was specifically noted for inquisitive and pre-emptive analytics (Fig. 7).

4.2.1. Descriptive analytics

Descriptive analytics are the simplest form of BDA method, and involves the summarization and description of knowledge patterns using simple statistical methods, such as mean, median, mode, standard deviation, variance, and frequency measurement of specific events in BD streams (Rehman et al., 2016). Often, large volumes of historical data is used in descriptive analytics to identify patterns and create management reports that is concerned with modelling past behaviour (Assunção et al., 2015). Watson (2014) asserts that descriptive analytics, such as reporting, dashboards, scorecards, and data visualization, have been widely used for some time, and are the core applications of traditional business intelligence. Descriptive analytics are considered backward looking and reveal what has already occurred. However, a trend that is being adopted in descriptive analytics now is to make use of the findings from predictive analytics, such as forecasts of future revenues, on dashboards/scorecards. Spiess, T'Joens, Dragnea, Spencer, and Philippart (2014) highlights root cause analysis and diagnostics are also form of descriptive analysis which involve both the passive reading and interpretation of data, as well as initiating particular actions on the system under test, and reading out the results. The authors discuss that root cause analysis is an elaborate process of continuous digging into data, and correlating various insights such as to determine the one or multiple fundamental causes of an event (Spiess et al., 2014).

Another form of descriptive analysis, pointed out by Banerjee, Bandyopadhyay, and Acharya (2013) is the use of dashboard sort of application when a business routinely generates different metrics including data to monitor a process or multiple processes across times. For example, this sort of application could be useful to understand in terms of the financial strength of a business at a given point of time or to compare it with others or its own across different point of time. In descriptive analytics, there is a need for analysts to nurture the skill of reading facts from figures, connecting them with the relevant decision-making process and finally taking a data-driven decision from a business perspective. Most of the BDA is commonly descriptive (exploratory) in nature and the use of descriptive statistical methods

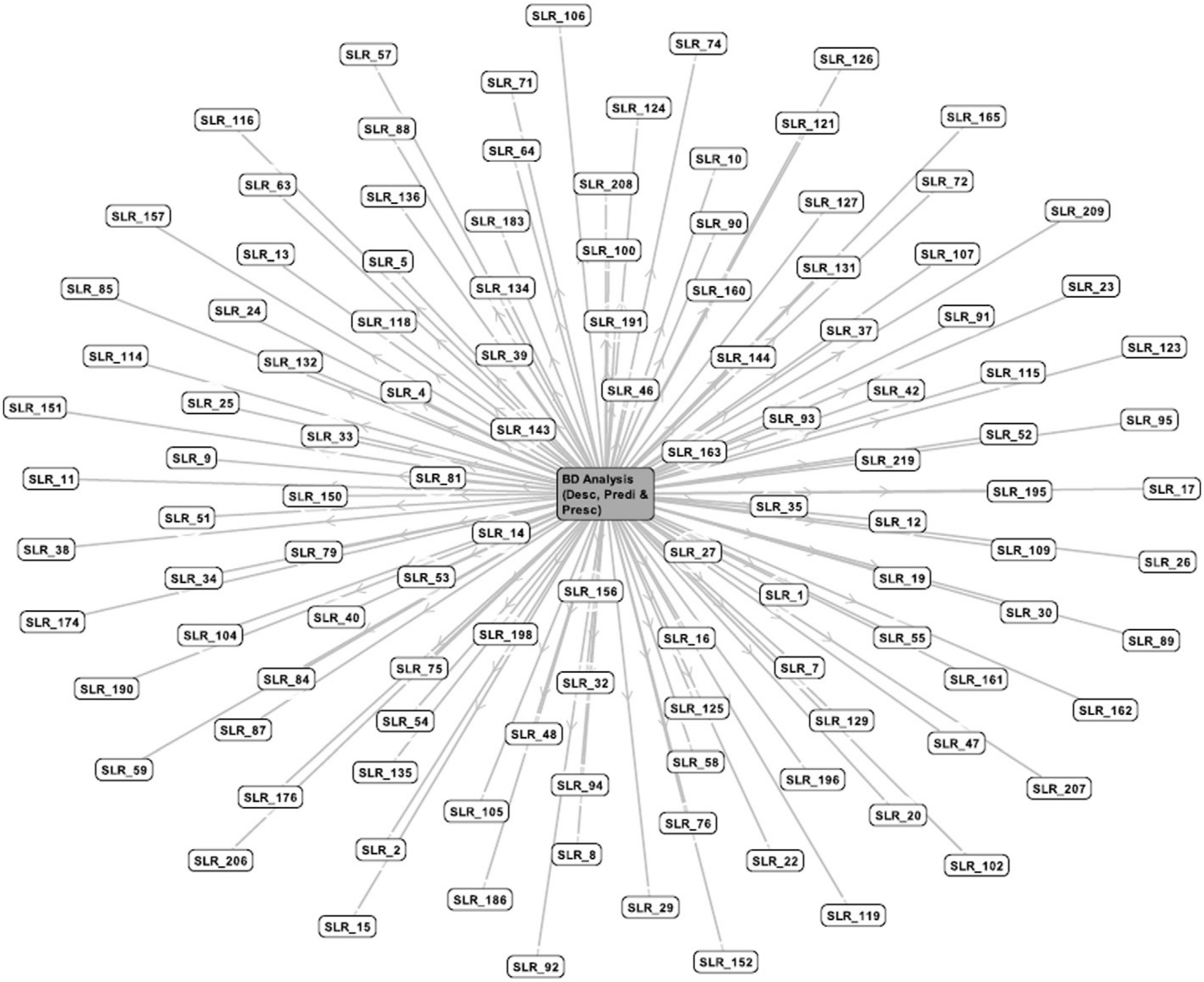

Fig. 7. Clusters of articles discussing/proposing/theorizing types of BD analysis methods.

(data mining tools) allows businesses to discover useful patterns or unidentified correlations that could be used for making business decisions.

4.2.2. Predictive analytics

This analytics is concerned with forecasting and statistical modelling to determine the future possibilities based on supervised, unsupervised, and semi-supervised learning models (Joseph & Johnson, 2013; Rehman et al., 2016; Waller & Fawcett, 2013). Gandomi and Haider (2015) asserts the need to develop new solutions for predictive analytics for structured BD. Predictive analytics are principally based on statistical methods and seeks to uncover patterns and capture relationships in data. Gandomi and Haider (2015) categorised predictive analysis into two groups – regression techniques (e.g., multinomial logit models) and machine learning techniques (e.g., neural networks). The authors highlight that some approaches, such as moving averages, attempt to identify historical patterns in the outcome variable(s) and extrapolate them to the future. Others, such as linear regression, seek to capture the interdependencies between outcome variable(s) and explanatory variables, and use them to make predictions. Hasan, Shamsuddin, and Lopes (2014) proposed a machine learning BD framework that envisaged the broad picture of machine learning in dealing with BD problems. The framework included the presentation of multi-structure input varieties from different sources, followed by the pipeline preprocessing phase prior to machine learning knowledge discovery. The authors implement the parallelism on machine learning approaches of BD predictive knowledge discovery based on Neural Network (NN) algorithms; Multiple Backpropagation (MBP) and Self-Organizing Map (SOM) using GPUMLib. In sum, predictive analytics aims to predict the future by analysing current and historical data. For example, determination of customers' propensity to churn, by correlating behaviour over a period of time with network event data such as usage records and fault indicators (Spiess et al., 2014).

4.2.3. Prescriptive analytics

This type of analytics is performed to determine the cause-effect relationship among analytic results and business process optimization policies. Thus, for prescriptive analytics, organizations optimize their business process models based on the feedback provided by predictive analytic models (Bihani & Patil, 2014). Although difficult to deploy, prescriptive analytics contribute to handling the information shift and the continuous evolution of business process models (Rehman et al., 2016). There are very limited examples of good prescriptive analytics in the real world. One of the reasons for this shortage is that most databases are constrained on the number of dimensions that they capture (Banerjee et al., 2013). Therefore the analysis from such data provides, at best, partial insights into a complex business problem. Few initial studies have applied the simulation optimization methods to the BDA. For instance, Xu, Zhang, Huang, Chen, and Celik (2014) proposed a framework called multi- fidelity optimization with ordinal transformation and optimal sampling (MO2TOS). The framework provides a foundation for descriptive and prescriptive analytics under the BD environment. In the MO2TOS framework, two set of high- and low-resolution models were developed. The authors highlighted that the high resolution model development can be very slow due to the large amount of data. On the other hand, the low-resolution models were much faster and can be developed using only a sample of data. The proposed MO2TOS framework is able to efficiently integrate the both the resolution models to optimize targeted systems under the BD environment.

In general, prescriptive solutions assist business analysts in decision-making by determining actions and assessing their impact regarding business objectives, requirements, and constraints. For example, *what if* simulators have helped provide insights regarding the plausible options that a business could choose to implement in order to maintain or strengthen its current position in the market.

4.3. Yearly publications

Using the keywords as stated in Section 1.2, initial search resulted in 2360 articles from 1996 until 2015 based on the number of subject areas including material sciences, energy, neuroscience, chemistry, etc. However, this research focused on only four subject areas such as business and management, computer science, decision science and social science (that directly relate to the special issue theme (i.e. *Big Data and Analytics in Technology and Organizational Resource Management*) – and following the systematic literature review steps (explained and illustrated in Section 3 and Fig. 3, respectively) – this research resulted in 227 articles. As presented in Fig. 8, the largest number of publications were recorded for year 2015 (with C = 114, 50.2%), followed by year 2014 (with C = 63, 27.7%) and year 2013 (with C = 43, 18.9%). With fewer publications (i.e. below the 5 mark) were recorded from 2009 until 2012 and zero articles recorded from 1996 to 2008.

Fig. 8 illustrates an abrupt increase in number of journal articles in the BD and BDA research area from 2013 onwards until 2015. Even through the initial search for articles (resulting in 2360 articles), there are more articles published from 2012 (e.g. 99 articles noted) until 2015 (e.g. 1156 articles noted). Regardless, the rapid increase in the articles highlights the awareness and importance of this area among the academic community, practitioners, and even governments worldwide (see e.g. Chen, Chen et al., 2012; Chen, Chiang et al., 2012; Joseph & Johnson, 2013). Despite the increase in the number of articles on BD and BDA, this research domain is still emerging (e.g. as noted from Scopus Database that from January 2016 to-date so far 295 articles have been published). With the significance of BD and BDA from a strategic perspective and the increasing number of articles, it appears that this research domain requires further in-depth conceptual as well as empirical, especially case study and survey based research studies.

4.4. Number of regions (geo-spatial coverage)

Fig. 9 highlights that the number of articles published on BD and BDA area represent 42 different geographical regions across the globe between 1996 and 2015. The total number of regions of the 227 articles is 790 as it takes into account of the geographical regions of the co-authors as well. It was considered appropriate to include the regions of the co-authors in order to avoid misrepresenting that each paper was single authored. From the total number of articles (i.e. 227) analyzed, the largest number of scholarly contributions came from the Chinese region (C = 241 scholars, representing 30.51%) – the 241 figure is the total number of authors and co-authors from China across all the 227 publications. This is followed by USA (C = 145, 18.35%), and then there is Australia (C = 51, 6.45%), UK (C = 49, 6.20%), and Korea (C = 37, 4.68%). The results in Fig. 5 evidently specify that China and USA have a lead on BD and BDA research area that is the upward trends in the first three to four regions noticeably indicate that there are clear signals of the growing interest in the BD and BDA area in those regions.

Whereas, from Belgium (with C = 1, 0.12%) to Italy (with C = 17, 2.15%) there is slow increase in the number of papers on BD and BDA. Nevertheless, the huge difference between two extremes clearly raises a vital research agenda for BD and BDA researchers and practitioners to explore: whether this position is a result of a global sector based BD and BDA divide or whether it is due to a lack of essential knowledge and proficiency to undertake BD and BDA research within such countries (i.e. more specifically those regions with five or less publications). In either case, the problem of a potential global BD and BDA area needs to be further studied (and or creating awareness) among the academics from countries such as the Belgium, Czech Republic, Denmark, Hong Kong, Norway and Russia. Researchers and scholars from China, USA, Australia, UK, Korea and Spain for instance should contemplate collaborating with researchers from under-represented regions so as to undertake more productive research and contributes towards the extant BD and BDA research area.

4.5. Types of publications

This section categorizes the list of 227 papers based on the publication type. The authors employed a analogous list of publication types as employed by Dwivedi and Mustafee (2010). This list is also similar to those identified by the publisher – Emerald. The data presented in Fig. 10 demonstrates that the vast majority of the publications are research papers (C = 159, 70.04%), followed by general review (with C = 27, 11.89%) and technical and conceptual papers (with C = 15, 6.60% and C = 9, 3.96%, respectively). A large number of research papers clearly indicate the significance of the BD and BDA area in different

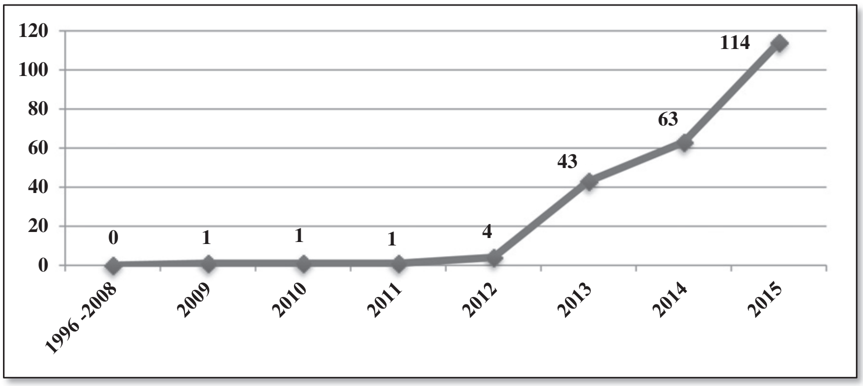

Fig. 8. Total number of papers published (from 1996 to 2015).

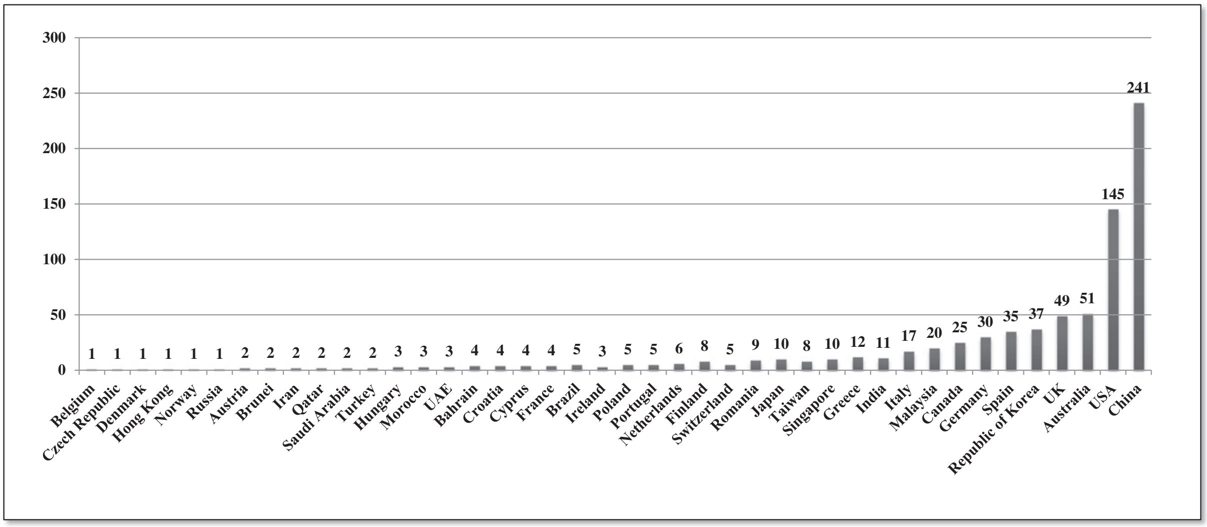

Fig. 9. Frequency of researchers from different geographical locations (from 1996 to 2015).

sectors (e.g. healthcare, government, and telecommunication). However, most of these research papers are analytical in nature (as explained in the following section) mainly focusing on experiments, performing simulations and proposing algorithms. The authors perceive that there is a need for more research considering using in-depth case studies in different sector organizations. Researchers and practitioners need to focus on developing and proposing sound solutions to BD challenges (Chen & Zhang, 2014).

4.6. Types of research methods employed

The research methods employed by the BD researchers in the selected 243 papers and were coded under different categories as suggested by Dwivedi, Kiang, Lal, and Williams (2008) and Dwivedi and Mustafee (2010). The findings suggest that although a total of 11 different types of research methods were recorded from our data analysis, the majority of studies were analytical in nature (C = 103, 45.37%). This was then followed by articles that are either conceptual/descriptive or theoretical in nature (C = 64, 28.19%), and design research (C = 12, 5.28%) methods. With regard to the analytical methods (with C = 103, 45.37%) – it was denoted as a combination of five different methods such as statistics, computer programming, simulation, algorithm and mathematical modelling, as also followed by Dwivedi and Mustafee (2010) and Kamal and Irani (2014). A big proportion of analytical articles clearly indicate that conducting experiments and simulations and or proposing algorithms have emerged as an alternative powerful meta-learning tool to accurately analyze the massive volume of data generated by modern applications (Chen, Chen et al., 2012; Chen, Chiang et al., 2012).

A small number of the selected articles employed interview and survey research approach to conduct their study – perhaps this is due to the nature of the BD and BDA research discipline that requires technical and methodical analysis of the huge type and format of data involved. Most of the studies reported using survey as tool to study the literature (i.e. secondary research) as opposed to seeking responses from the respondents (e.g. Chen, Mao, & Liu, 2014). The other categories with their associated counts and percentages are presented in Fig. 11.

5. Conclusions

The authors of this paper have presented a holistic view of BD practices and application of BDA methods as presented in a normative slice of literature. Based on the findings from existing research studies, the presented research has sought to analyze, synthesize and present a comprehensive structured analysis on BD and BDA to support the signposting of future research directions. The SLR methodology adopted demonstrated to be a convenient tool for conducting a descriptive literature reviews, with contributions including the synthesis of core conclusions of the literature, the literature voids, and the formation of a foundation for future research. The findings of this structured

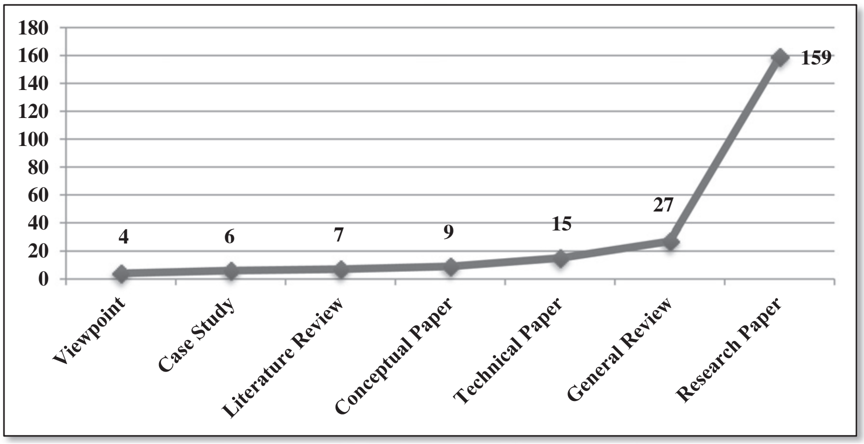

Fig. 10. Classification of publication types (from 1996 to 2015).

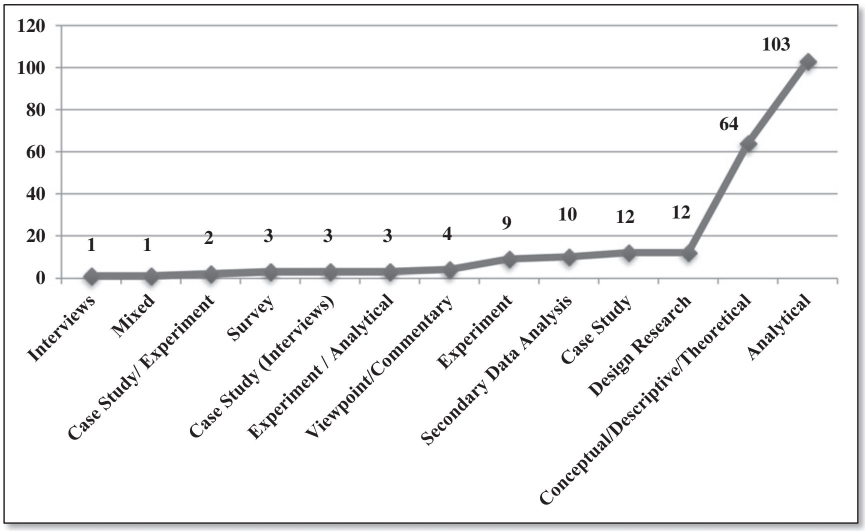

Fig. 11. Classification of research methods (from 1996 to 2015).

literature review will assist both BD and BDA academics and practitioners to develop new solutions based on the challenges identified in this paper. BD is still an emerging phenomenon but in the recent past years its significance in different industries and countries (as evident from Fig. 9) makes it a pertinent research area for academic and management studies. It is evident from the review conducted that it has significantly changed the data management landscape with scope for further profound changes.

This SLR paper has revealed the past and current state of BD and BDA research published, thereby focusing on the past trends and current patterns in BD and BDA practices. Following Tranfield et al. (2003) and Kitchenham and Charters (2007) Systematic Review Approach, this paper extracted and reviewed 227 journal articles from 1996 to 2015 from the Scopus database – as a result fulfilling the aim of this literature review paper (as indicated in Section 1.1). Figs. 4 to 11 clearly indicate the past trends and current patterns in the number of articles published on BD and BDA. Moreover, the continuing interest (as indicated from Fig. 8 – increasing number of BD and BDA articles over the years) specifies that in future research studies; academics, researchers and practitioners may focus on the BD challenges to further propose robust solutions to the challenges of acquiring and storing, mining and cleansing, aggregating and integrating, analysis and modelling and interpreting data. The intention in conducting this detailed investigation was to provide a useful and usable resource of information for future researchers.

5.1. Research implications to research and practice

- Implications to Research: This SLR offered a number of useful insights into the extant status of research into BD and BDA, how it is defined and conceptualised, and the key types of research methodologies employed to date. The prime emphasis of these SLR based articles has been on using analytical and or conceptual/descriptive/theoretical research methods; however, due to the emerging nature of this area, there is a need to develop and understand BD and BDA in an intensive way using case studies and survey based research where appropriate. The authors assert that more practical insights into BD and BDA can be attained by utilising the findings of this SLR to enlighten and direct research towards a more holistic view of the BD as a research discipline. In this paper the authors have not restricted their focus on identifying specific lines of enquiry on BD and BDA, but rather focused on synthesizing and presenting a comprehensive analysis of the normative literature on BD challenges and the types of BDA methods discussed, proposed and or employed by organizations. This paper extends the research stream on BD and BDA by demonstrating and analysing the key trends related to the challenges of BD and BDA methods.
- Implications to Practice: The authors of this paper have presented the practice community with an insight to the plethora of BD and BDA methods available and, insight to their application. While there is no one advocated robust approach, the descriptive insight presented will offer an opportunity for practitioners and applied researchers to align their approached to the application pursued by others.

5.2. Limitations

The authors recognise that our study has limitations, and readers and future academics and researchers should be aware of these and indeed interpret the material presented in this paper within the context of the limitations. By explanation, a meta-analysis rest on the existing as well as accessible research studies (both conceptual and empirical). While the authors conducted a thorough literature search through the Scopus database to identify all possible relevant articles, it is possible that some research articles could have been missed in this review from some *other* leading databases (i.e. Web of Science and EBSCO). So to avoid duplication, every effort was exhausted to acquire and analyze all relevant information essential, regarding the two questions (i.e. Q1 and Q2) from the articles reviewed from the Scopus database. Additionally, the analysis and synthesis are based on the research team interpretation of the selected articles. The authors attempted to avoid these issues by cross-checking papers independently and thus deal with embedded bias but errors might have occurred but this research is considered robust as every effort to mitigate error was taken.

5.3. Suggestions for future research

Building upon the rich underpinning of the research findings described and overall understanding acquired in this paper, the authors presents the concerns that merit further research and anticipate that these issues may hold the potential in contributing towards the future research studies. The analysis of the selected articles reveals that the opportunity clearly exists to strengthen empirical research based on in-depth case study based qualitative and survey based quantitative approach, as most of the articles analyzed followed an analytical approach. Furthermore, there is need for stronger infusion of generic theory into the BD and BDA debate. BD is a cross-cutting theme, and many connections exist with established topics across computing, engineering, mathematics, business and management, social sciences, etc. It would be valuable to expand the scope of the subject area and to repeat this exercise to identify and draw links with established theoretical contributions in other different associated areas. A publication based on such analysis would provide an extremely valuable platform for the BD and BDA research and practitioners' community.

Appendix A

Paper code	Citation
SLR_1	Jacobs. A., (2009). The pathologies of bis data. Communication: of the ACM. 52(8), 30–44.
SLR_2	Li. X. Lillibridge, M. & Uvsal. M. (2011). Reliability analvsis of deduplicated and erasure-coded $torse.e. ACMS1GMETRICS Performance Evaluation Review. 38(3). i-9.
SLR_3	Reddi, V.J. Lee. B.C. Chflimbt. T. Sc Vaid. K. (2011). Mobile processors for energy-efficient web search. ACM Transactions on Computer Svstems (TOCS). 29(3). 9.
SLR_4	Chen, H. Chians. R. H Sc Storey. V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. Quanerh. 35(4). 1105-11SS.
SLR_5	Chen. G. Clien. K. Jians. D. Ooi. B. C. Shi. L. Vo. H. I. 8c Wu. S. (2012). E3: an elastic execution ensme for scalable data processins. Journo! of Information Processing. 20(1). 05–70.
SLR_6	Longley, P. A. (2012). Geodemosraphics and die practices of seosraphic information science. International Journal of Geographical Information Science. 25(12), 2227–2237.
SLR_7	Doersch.C.: Sinsh. S. Gupta. A., Sivic. J. Sc Efros. A. A. (2012). What makes Pari3 look like Paris^0. Communications of die ACM. 58(12), pp. 103–110.
SLR_8	Riedel. M.AVittenburg. P. Reetz, J. yan de Sanden. M.,Rybkki, J. von St Vieth. B. Fiameni, G. Martani, G., Michelini. A., Cacctari. C. & Hbers. W. (2013). A data infrastructure reference model with applications: towards realization of a ScienceTube vision wrth a data replication service. Journo! of Internet Serf ices and Applications. 4(1). 1–17.
SIR_9	Mershad. K. Altai. H. Sashir. M. Hajj. H. & Awad. M. (2013). A mathematical model to analyze the utilization of a doud datacenter middleware. Journal ofXetworkand Computer Applications. 39.399–415.
SLR_10	Huans. G. . Sons. S. Gupta. J. N. & Wu. C. (2013). A second order cone prosrammins approach for semi-supervised leamins. Pattern Recognition. 45(12). 354S—355S.
SLR_11	Lin. J. Yin. J. Cai. Z. Liu. Q. Li. K. & Leune. V. (2013). A s ecure and practical mechanism of out30urdns extreme leamins machine in doud computms. IEEE Intelligent Systems. 28(6). 35–3 S.
SLR_12	Kim. K. J. Hons. S. P. Sc Kim. J. Y. (2013). A study of privacy protection from risk of hijackins data. Internationa! Journal of Multimedia and Ubiquitous Engineering. 5(1), 235–244.
SLR_13	Chen. Z. Xiao. N. & Liu. F. 2013. An SSD-based accelerator for directory parsins in storage systems containing massive files. Peer-to-Peer Networking and Applications. 5(4). pp.39″ —40S.
SLR_14	Fischer. F. Fuchs. J. Mansmann. F. & Keim. D. A. (2013). BANKSAFE: Visual analytics for big data in large-scale computer uetworks. Information Visualization. 74(1). 51–61.
SLR_15	Ln. J. Sc Li. D. (2013). Bias correction in a small sample from big data. IEEE Transactions on Knowledge and Data Engineering: 25(11). 265S-2663.
SLR_16	Joseph. R. C. Sc Johnson. N. A. (2013). Big data and transformational government. IT Professional. 15(6). 43-4S.
SLR_17	Sukumar. S. R. Sc F errell. R. K. (2013). 'Big Data' collaboration: Exploring, recording and sharing enterprise knowledge. Information Servkes and Use. 33(3–4). 257–270.
SLR_I8	Weber. S. (2013). Big data privacy and security challenges. In Proceedings of the ACM Works hop on Building Anahs is Datasets and Gathering Experience Returns for Securing pp. 1–2.
SLR_19	Hone. T. H. Yun. C. H. Park. J. W., Lee. H. G. Jung. H. S. &Lee. Y. V. (2013). Big data processing with MapReduce for *E*-book. Internationa! Journal of Multimedia and Ubiquitous Engineering. 5(1). 151–162.
SLR_20	Choo. J. Sc Paik. H. (2013). Customizing computational methods for visual analytics with big data. Computer Graphks andApplkations. IEEE. 33(4). 22–28.
SLR_21	Crowe. J. Sc Candhsh. J. R. (2013). Data analytics: the next bis thing m information. Proceeding: of the 14* International Conference on Grev L her azure. Rome. Italy, pp. 139–142.
SLR_22	Grolinget. K. Higashino. W. A. Tiw'ari. A. .Sc Capretz. M. A. (2013).Data management in doud environments: NoSQL andNewSQL data stores. Journal of Cloud Computing:. idxances. Systems and.ipp!kaiions. 2(1). 1–24.
SLR_23	Miller. H. (2013). Bia-Data in Cloud Computing: A Taxonomv of Risks. Information Research. 18(1).
SLR_24	Tme. K. M. Washio. T. Wells, J. R. Liu. F. T. Sc. Arval. S. (2013). DEMass: a new density estimator for bis data. Knowledge and Information Svstems. 33(3). 493–524.
SLR_25	Shen. Y. Sc VarveL, V. E. (2013). Developing data management services at the Johns Hopkins University. The Journal Of Academic Librarianship. 39(6). 5 52–5 57.
SLR_26	Mansell. R. (2013). Employing digital crowdsourced information resources: Managing the emerging information commons. International Journal of the Commons. 7(2), 255–277.
SLR_27	Kraska. T. (2013). Finding the needle in the big data svstems haystack. IEEE Internet Computing. 1. S4-S6.
SLR_28	Bamaghi. P. Sheth A. & Henson. C. (2013). From Data to Actionable Knowledge: Big Data Challenges in die Web of Things [Guest Editors' Introduction]. IEEE Intelligent Svstems. 28(6). 6–11.
SLR_29	Kumar. A., Niu, T. & Re. C. (2013). Hazy: Making it easier to build and maintain brg-data analytics. Communkaiions of the ACM, 55(3). 40–49.
SLR_30	Rabkm. A. Sc Katz. R. H. (2013). HowHadoop dusters break. IEEE Software. 30(4), SS-94.
SLR_31	Baumgarten. M. Mulvama, M. Rooney, N. & Reid. J. (2013). Kevword-Based Sentiment Mining using Twitter. Imernarional Journal ofAmbient Computing and Intelligence. 5(2). 56–69.
SLR_32	Yang. Y. Long, X. Sc Jiang. B. (2013). K-Meams method for grouping in hybrid MapReduce duster. Journal of Computers. 5(10). 2648–2655.
SLR_33	Xia. S. Xie. J., Dai. D. Zhang. H. Nie. Q. Kawata. S. & Zhang. W. (2013). Rum combined with Hadoop application based-on cpse-bio. Journal of Next Generation Information Technology. 4(3). 160.
SLR_34	Lee. C. H. Sc Chien. T. F. (2013). Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. Journal of Information Science. 39(4). 523–543
SLR_35	Zhao. Z., Zhang. R. Cox. J. Dulmg. D. Sc Sarle. W. (2013). Masstvdy parallel feature selection: an approach based on variance preservation. Machine Learning. 92(1). 195–220.
SLR_36	Yfldtrim, E Kim. J. Sc Kosar. T. (2013). Modelling throughput sampling size for a doud-hosted data scheduling and optimization service. Future Generation Computer Svstems. 29(7). 1795-1S07.
SLR_37	Ouzounis. G. K, Syrris. V. & Pesaresi. M. (2013). Multiscale quality assessment of Global Human Settlement Laver scenes against reference data using statistical learning. Pattern Recognition Letters. 34(14). 1636–1647.
SLR_38	Karacapilidis. X. Tzagarakis. M. Sc Christodonlou. S. (2013). On a meaningful exploitation of machine and human reasoning to tackle data-mtensive decision miking. Intelligent Decision Technologies. 7(3), 225–236.
SLR_39	Wen, D. Guo-min. G., Tian-jun, W., & XLn-ju, Y. (2013). Organization and Management of Meteorological Sensor Network Collected BigData. Information Technolog}' Journal, 12(22), 6636–6640.
SLR_40	Deng, S. G, Huang, L. T., Wu, B. & Xiong. L. R. {2013). Parallel optimization for data-intensive service composition. Journal of Internet Technology, 14(5), 817–824.
SLR_41	Procter. R., Crump. J. Karstedt. S. Voss, A, & Cantijoch,M. (2013). Reading the riots: What were the Police doing on Twitter?. Policing and Society, 23(4). 413–436.
SLR_42	Procter. R. Vis, F. & Voss, A (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. International Journal of Social Research Methodology', 16(3), 197–214.
SLR_43	Qin. H. F. & Li. Z. H. (2013). Research on the Method of Big Data Analysis. Information Technology? Journal, 72(10), 1–7.
SLR_44	Small. S. G. &. Medsker, L. (2013). Review of information extraction technologies and applications. Neural Computing and Applications. 25(3–4), 533–548.
SLR_45	Tan, W., Blake, M.B., Saleh, I., & Dustdar. S. (2013). Social-network-sourced big data analytics. IEEE Internet Computing, 5,62–69.
SLR_46	Yang, H. (2013). Solving problems of imperfect data streams by incremental decision trees. Journal of Emerging Technologies in Web Intelligence, 5(3), 322–331.
SLR_47	Wang. W. Lu, D. Zhou.X., Zhang, B.,& Mu, J. (2013). Statistical wavelet-basedanomalv detection in big data with compressive sensing. EURASIP Journal on Wireless Communications and Networking, 2013(1), pp. 1–6.

Appendix A (*continued*)

Paper code	Citation
SLR_48	Hu, B., Carvalho, N.,& Matsutsuka. T. (2013). Towards Big Linked Data: A Large-Scale, Distributed Semantic Data Storage. International Journal of Data Warehousing and Mining (TJDWM), 9(4), 19–43.
SLR_49	Jimei, L., Yuzhou. H., & Meijie, D. (2013). XBRLinthe Chinese Financial Ecosystem. IT Professional, 15(6), 36–42.
SLR_50	Chen. J., Chen, Y.Du. X. Li, *C. Lu*. J. Zhao. S. & Zhou. X. (2013). Big data challenge: a data management perspective. Frontiers of Computer Science, 7(2), 157–164.
SLR_51	Lee, B., & Jeong,E.{2014). A design of a patient-customize dhealthcare svstenbased on the Hadoopwithtextmining{PHSHT) for an efficient disease management and pre diction. Imernaiional Journal ofSoftware Engineering & Applications, #(8), 131–150.
SLR_52	Faria, F. A. Dos Santos, J. A. Rocha. A. & Torres. R. D. S. (2014). A framework for selection and fusion of pattern classifiers in multimedia recognition. Pattern Recognition Letters, 39, 52–64.
SLR_53	Chen. Z. Lu. Y., Xiao.N. and Liu. F. 2014. Ahvbrid memory built bv SSD and DRAM to support in-memorv Big Data analvtics. Knowledge and Information Systems, 41(2). 335–354.
SLR_54	Lin. C. Y., & Liao. J. K (2014). A fob-oriented load-distribution scheme for cost-effective NameNode service in HDFS. International Journal of Web and Grid Services, 10(4). 319–337.
SLR_55	Antonie, A,Marjanovic,M.,Pripuzic, K., & Zarko, I. P. (2014). Amobile crowd sensing ecosystem enabledby CUPUS: doud-basedpublish' subscribemiddleware forthe internet of things Future Generation Co mp uier Systems, 56,607–622
SLR_56	Ulltveit-Moe,N. (2014). A roadmap towards improving managed security services from a privacy perspective. Ethics and Information Technology, 16(3), 227–240.
SLR_57	Fahad. A., Ashatn, N. Tari, Z.: Aamri, A., Khalil, I., Zomava, A.Y.,Foufou, S., Bouras, A (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE Transactions on Emerging Topics in Computing, 2(3), 267–279.
SLR_58	Liu. S. Cui. W. Wu, Y. & Liu, M. (2014). A survey on information visualization: recent advances and challenge s. The Visual Computer. 39(12). 1373–1393.
SLR_59	Zhang. F. Cao. J. Khan, S. U. Li. K. & Hwang. K. (2015). A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications. Future Generation Computer Systems. 43. 149–160.
SLR_60	Hofrnan, W., & Rajagopal, M. (2014). A technical framework for data sharing. Journal of Theoretical and Applied Electronic Commerce Research, 9(3), 45–58.
SLR_61	Kuang. L.,Hao,F. Yang. L. T. Lin. M.Luo. C. &Min, G. (2014).Atensor-basedapproachforbig data representation and dimensionality reduction. IEEE Transactions on Emerging Topics in Computing, 2(3), 280–291.
SLR_62	Lebdaoui. I. Qrhanou. G. & Elhajfi, S. (2014). An Integration Adaptation for Real-Time Datawarehousing. Internationa! Journal of Software Engineering and its Applications, 2(11). 115–128.
SLR_63	AgrawaL D. (2014). Analvtics based decision making. Journal of Indian Business Research, 6(4), 332–340.
SLR_64	Song. M. Kim, M. C. Jeong. Y. K (2014). Analysing the political landscape of 2012 Korean Pre sidential Ele ction in Twitter. IEEE Intelligent Systems. 29(2), 18–26.
SLR_65	Liu, C., Chen, J., Yang.L. T, Zhang, X, Yang, C,& Rao,K (2014). Authorized public auditing of dynanicbig datastorage on cloud with efficientverifiable fine-grained updates. IEEE Transactions on Parallel and Distributed Systems, 25(9), 2234–2244.
SLR_66	Gandomi. A. & Haider. M. (2015). Bevondthe hvpe: Big data concepts, methods, and analytics. International Journal of Information Management. 35(2), 137–144.
SLR_67	Tinati. R. Halford. S. Can. L. & Pope. C. (2014). Big data: methodological challenges and approaches for sociological analysis. Sociology.48(4), 663–681.
SLR_68	Yin, H. Jiang. Y. Lin. C. Luo, Y. & Liu. Y. (2014). Big data: transforming the design philosophy of future internet. IEEE Network, 28(4), 14–19.
SLR_69	Krishnamurthv.R., & Desouza, K C.{2014). Big data analytics: The case of the social security administration. Information Polity, 19(3,4), 165–178.
SLR_70	Diamantoulakis. P. D., Kapinas. V. M.,& Karagiannidis. G. K_ (2015). Big data analytics for dynamic energy* management in smart grids. Big Data Research, 2(3), 94–101.
SLR_71	Wang, Y., & Wiebe, V. J. (2014). Big Data Analvtics on the Characteristic Equilibrium of Collective Opinions in Social Networks. International Journal of Cognitive Informatics and Natural Intelligence, 8(3), 29–44.
SLR_72	Fernandez, A, del Rio, S., Lopez, V′., Bawakid, A, del Jesus, M. J., Benitez, J. M.,& Herrera, F. (2014). BigData with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks. Wilev Interdisc iplinarv Reviews: Data Mining and Knowledge Discovery, 4(5), 3 80–409.
SLR_73	Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C, & Choi, H. (2014). Big data, open government and e-Govemment: Issues, policies and recommendations. Information Polity, 29(1,2), 5–16.
SLR_74	Kim. G. H., Trimi, S., Chung, J. H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), 78–85.
SLR_75	Yi, X., Liu. F. Liu, J., & Jin, H. (2014) Building a network highway for big data: architecture and challenges. IEEE Network, 28(4), 5–13.
SLR_76	Zhang, Y., Chen, M., Mao. S. Hu. L. & Leung. V. (2014). Cap: Community activity prediction based on big data analysis. IEEE Netw ork. 28(4). 52–57.
SLR_77	Leeflang. P. S. Verhoef. P. C. Dahlstrom. P. & Freundt T. (2014). Challenges and solutions for marketing: in a diaital era. European Management Journal. 32(1). 1–12.
SLR_78	Imran. A. SrZoha. A. (2014). Challenges in 5G: howto empower SON with big data for enabling 5G. IEEE Network. 26(6). 27–33.
SLR_79	Hu. R. Dou. W„ & Liu, J. (2014). ClubCF: A Clusterms-Based Collaborative Filterina Approach for Big Data Application. IEEE Transactions on Emerging Topics in Computing. 2(3), 302–313.
SLR_80	Hurlburt. G. Bojanova. 1. & Berezdivm. R. (2014). Computational Networks: Challenging Traditional Proaram Management IT Professional. 16(6). 66–69.
SLR_81	Li. Z. Sharaf. M.A. Sitbon. L. Du. X. & Zhou. X. (2014). Core: a context-aware relation extraction mediod for relation completion. IEEE Transactions on Knowledge and Data Engineering. 26(4). S36-S49.
SLR_82	Zen?. D. Gu. L. & Guo. S. (2015). Cost minimization for bis data proces sins in seo-distributed data centers. In Cloud 'Networking for Big Data. Sprmser International Publi3hins. pp. 59–7 S.
SLR_83	Chen. C. P. & Zhsns. C. Y. (2014). Dana-intensive applications, challenses. Techniques and technologies: A survey on Bis Data. Information Sciences. 275.314–347.
SLR_84	Kirn. Y. Shim. K_. Kim. M. S. 5c Lee. J. S. (2014). DBCURE-MR: an efficient densitv-based dusterins alsorithm for larse data using MapReduce. Information Svs terns. 42.15–35.
SLR_85	Jung. B. & Lim. S. (2014). Designing a Smart Consumption Tracking Model. International Journal of Software Engineering and its Applications, 6(10). 167–178.
SLR_86	Bkimsni. A. & WBlcocks. L. (2014). Digitisation. "Bis D ata and die trans formation of accounting information. Accounting and Business Research. 44(4). 469–490.
SLR_87	Ryiavy, S. J. Bromlev. D. & Daggett. V. (2014). DIVE: A eraph-based visual-anal',lies framework for big data. Computer Graphics and,-Applications. IEEE. 34(2). 26–37.
SLR_88	Sn, Y. Agrawal. G. Woodrmg. J. Myers, K. Wendelberger. J. & Ahrens. J. (2014). Effective and efficient data sampling using bitmap indices. Cluster Computing. 17(4), 10S1–1100.
SLR_89	Guo. T. Papaioannon. T. G. 5c Aberer. K. (2014). Efficient Indexing and Query Processing of Model-View Sens or D ata in die Cloud. Big Data Research. 1. 52–65.
SLR_90	Ellis. J. Fokoue. A. Hassanzadeh. O. Kementsietsidis. A. Srinivas. K. & Ward. M. J (2015). Exploring Big Data with Helix: Finding Needles in a Bia Haystack ACMSIGMOD Record. 43(4). 43–54.
SLR_91	Sim, N. Moms. J.G. Xu. J. Zhu. X. & Xie. M. (2014). iC ARE: A framework for big data-based bankmg customer anal\tics. IBM Journal of Research and Development. 56(5.6). 1–9.
SLR_92	Dobre. C. Sc Xhafa. F. (2014). Intelligent sendees for big data science. Future Generation Computer Systems. 37. 267–281.
SLR_93	Meng. S. Dou. W._ Zhang. X. & Chen. J. (2014). Kasr: A keyword -aware service recommendation mediod on MapReduce for big data applications. IEEE Transactions on Parallel and D is iribuied Systems, 25(12), 3221–3231.
SLR_94	Abawgjv. J. H. Kelarev. A. Chowdhury. M. (2014). Large iterative multitier ensemble classifiers for security of big data. IEEE Transactions on Emerging Topics in Computing. 2(3). 352–363
SLR_95	Hasan. S. Shamsuddin. S. M. Sc Lopes. N. (2014). Machine learning big data framework and anal',tics for big data problems. International Journal of Advances in Soft Computing and its Applications, 6(2). 1–14.
SLR_96	Goldberg.D. Olivares. M. Li. Z. and Klein A.G. 2014. Maps &. GIS data libraries m the era of big data and doud computing. Journal of Map & Geography Libraries. 10(1). pp. 100–122.
SLR_97	Bravo-Marquez. F. Mendoza. M. Sc Poblete. B. (2014). Meta-level sentiment models for big social data analysis. Knowledge-Based Systems, 69. 86–99.

(*continued on next page*)

Appendix A (*continued*)

Paper code	Citation
SLR_98	Otji, U. A. Remscrim, Z., Schantz. C., Donna]. J., Paris. J. Gillman. M., Surakitbovom, K, Leeb. S. B.; & Kirtley. J. L. (2015). Non-intm3ive induction motor speed detection. Electric Power Applications. 1ET Electric Power Applications. 9(5), 3SS-39i6.
SLR_99	Taheri. J. Zomava. A. Y., Siegel. H. J. & Tari. Z. (2014). Pareto frontier for job execution and data transfer tune in hybrid clouds. Future Generation Computer Systems. 37. 321–334.
SLR_100	Barbierato. E. Gribaudo. M. Sc laoono. M. (2014). Performance evaluation of NoSQL big-data applications using multi-formalism models. Future Generation Computer Systems. 37. 345–353.
SLR_101	Ma. S. Meng. X. Sc Wang. F. (2014). Report on the Sixth International Workshop on Cloud Data Management. A CM SIGMOD Record. 43(2). 53–56.
SLR_102	Ji. C. Li. Z. Qu. W.Xu. Y., Sc Li, Y. (2014). Scalable nearest neighbour query processing based on Inverted Grid Index. Journal of Network and Computer. ipplications. 44.172–182.
SLR_103	Chelmis. C. Wu. H. S crathia. V. Sc Prasanna. V. K (2015). Semantic sodal network analysis for die enterprise. Computing andInformancs. 33(3). 479–502.
SLR_104	Kourtesis. D. Alvarez-Rodriguez. J. M. 8c Paraskakis. I. (2014). Semantic-based QoS management in doud systems: current 3tatu3 and future challenges. Future Generation Computer Systems. 32.307–323.
SLR_105	Wang. C'. Li. X. Zhou. X. Wang. A. Sc Xedjah. N. (2015). Soft computing in big data intelligent transportation systems,. ippliedSoft Computing. 38,1099-110S.
SLR_106	Jardak. C. Mahdnen. P. Sc Riihijarvi. J. (2014). Spatial bie data and wireles s networks: experiences. Applications, and res earch challenges. IEEE Network. 28(4). 26–31.
SLR_107	He. B. Li. Y. Husns. H. Sc Tang. H. (2014). Spatial-temporal compression and recovery in a wireles s sensor network in an underground tunnel environment. Knowledge and Information Systems. 47(2). 449–465.
SLR_108	Liu. Z.,Li. J. Li. J, Jia. C bang. J., Yuan. K. (2014). SQL-based fuzzy query medianism over encrypted database. International Journal of Data Warehousing and Mining (1JDWM). 70(4). 71–87.
SLR_109	Lee. T. Lee. H. Rhee. K. H. Sc Shin. U. S. (2014). The efficient implementation of distributed indexing with Hadoop for digital investigations on Big Data. Computer Science and Information Systems. 17(3). 103 7–1054.
SLR_110	Lu, R., Zhu. H. Liu. X. Liiu, J. K. & Shao. J. (2014). Towards efficient and privacy-preserving computing in big data era. IEEE Network, 26(4). 46–50.
SLR_111	Takaishi. D. Xishivama. H. Kato. N., Sc Miura. R. (2014). Towards energy efficient big data gathering in densely distributed sensor networks. IEEE Transaction: on Emerging Topic: in Computing. 2(3). 388–397.
SLR_112	Zliobaite. I. Holmen. J. Koskmen. L. & Teittmen. J. (2015). Tow^{r}ards hardw are- driven design of low-energy algorithms for data analysis. ACM SIGMOD Record. 43(4). 15–20.
SLR_113	Watson. H. J. (2014). Tutorial: Big data anal’,lies: Concepts, technologies, and applications. Communications of the Association for Information Systems. 34(1). 1247–1268.
SLR_114	Zhang. F. Liu. M. Gui. F. Shen. W. Shami. A. & Ma. Y. (2015). A distributed frequent itemset mining algorithm using Spark for Big Data analytics. Cluster Computing. 76(4). 1493–1501.
SLR_115	Jun. S. Lee. S. J. &Rvu. J. B. (2015). A Divided Regression Analysis for Big Data. Imernationa! Journal ofSoftware Engineering and its. Applications. 9(5). 21–32
SLR_116	Xing. J., Sc Sieber. R. E. (2015). A land use land cover change gecspatial cyberinfrastmctnre to integrate big data and temporal topology. Internationa! Journal of Geographical Information Science. 39(3). 573–593.
SLR_117	Kim. 1. (2015). A Study on the Development of Next Generation Intelligent Integrated Security Management Model using Big Data Technology. International Journal ofSecurity and it: Applications, 9(6). 217–226.
SLR_118	Zhang. X. Hu. Y. Xie. K. Zhang. W„ Su. L. & Liu. M. (2015). An evolutionary trend reversion model for stock trading rule discovery. Knowledge-Based Systems. 79,27–35.
SLR_119	Juhic. N. Shaima. A. Nesterov, S. & Jukic. B. (2015). Augmenting Data Warehouses with Big Data. Information Systems Management. 32(3), 200–209.
SLR_120	Jara. A. J., Genoud,D., & Bocchi. Y. (2015). Big data for smart cities withKNIME a real experience in the SmartSantander testbed. Software: Practice and Experience. 4 5(S). 1145–1160.
SLR_121	Li, H., Lu, K, & Meng, S. (2015). Bigprovision: A provisioning framework for big data analytics. IEEE Network, 29(5), 50–56.
SLR_122	Zhang, S., Yin. D., Zhang, Y., & Zhou, W. (2015). Computing on Base Station Behaviour Using Erlang Measurement and Call Detail Record. IEEE Transactions on Emerging Topics in Computing, 3(3), 444–453.
SLR_123	Miller H J, & Goodebild, M. F. (2015). Data-driven geography. Geo Journal, 60(4), 449–461.
SLR_124	Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement audits. Accounting Horizons, 29(2), M423–429.
SLR_125	Wang. F. Hu, L., Zhou. D., Sun., R., Hu, J., & Zhao. K. (2015). Estimating online vacancies in real-time roadtraffic monitoring with traffic sensor data stream. AdHoc Networks, 35,3–13.
SLR_126	Farahat, A. K,Elgoharv, A., Gho dsi, A., & Kamel, M. S. (2015). Greedy column subset selection for large-scale data sets. Knowledge and Information Systems, 45(1), 1–34.
SLR_127	Shah. T., Rabhi, F. & Ray, P. (2015). Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. Cluster Computing, 76(1), 351–367.
SLR_128	Neish, P. (2015). Linked data: what is it and why should you care? The Australian Library Journal, 64(1), 3–10.
SLR_129	Ashraf, J. Hussain, O. K_, & Hussain, F. K_ (2015). Making sense from Big RDF Data: OUSAF for measuring ontology usage. Software: Practice and Experience, —75(8), 1051–1071.
SLR_130	La ebb e eke. C. &Picot,A. (2015). Reflections on societal andbusmessmodel transformation arising from digitization andbig data analytics: A research agenda. The Journal of Strategic Information Systems, 24(3), 149–157.
SLR_131	Otero. C. E. & Peter, A (2015). Research Directions for Engineering Big Data Analytics Software. IEEE Intelligent Systems, 36(1), 13–19.
SLR_132	Triguero,I.; del Rio. S., Lopez, V. Bacardit.J, Benitez, J. M.,& Herrera. F. (2015). ROSEFW-RF: the winner algorithm for the EC BDL’ 14 big data competition: an extremely imbalanced big data bioinformatics problem. Knowledge-Based Systems, 87,69–79.
SLR_133	Cicotti, G, Coppalino,L.,D'Antonio, S., & Romano, L. (2015). RuriimeModel Checking for SLACompliance Monitoring and QoS Prediction Journal of Wireless Mo bile Networks, Ubiquitous Computing, and Dependable Applications. 6(1), 4–20.
SLR_134	Benatallah. B. & Motahan-Nezha cL H. R. (2015). Scalable graph-based OLAP analytics over process execution data. Distributed and Parallel Databases. 34(3). 379–423.
SLR_135	Jiang. H. Chen, Y., Qiao, Z., Weng, T. H., & Li, K_ C (2015). Scalingup MapReduce-basedbig data processing onmulti-GPU systems. Cluster Computing, 76(1), 369-3S3.
SLR_136	Sandhu. R. & Sood, S. K. (2015). Scheduling ofbig data applications on distributed doudbased on QoS parameters. Cluster Computing, 76(2), S17—82S.
SLR_137	Xu, J. Huang, E. Chen, C. H.,& Lee, L. H. (2015). Simulation optimization: a review and exploration in the new era of cloud computing and big data. Asia-Pacific Journal of Operational Research, 32(3), 1–34.
SLR_138	Kitcbin. R., & Lauriault, T. P. (2015). Small data in the era ofbig data. GeoJournal, 60(4), 463–475.
SLR_139	Buhalis,D. & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. Journal of Destination Marketing di Management, 4(3), 151–161.
SLR_140	Kune, R., Konugurthi, P. K_, Agarwal, A, Chillarige, R. R. & Buyya, R. (2015). The anatomy ofbig data computing. Software: Practice and Experience, 46(1), 79–105.
SLR_141	Wang,Z., Chen, H., Fu, Y.,Liu, D., & Ban, Y. (2015). Workload balancing and adaptive resource management for the swift storage system on cloud. Future Generation Computer Systems, 51,120–131.
SLR_142	Hu, W., & Jia, C. (2015). A bootstrapping approach to entity linkage on the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web,34, 1–12.
SLR_143	Lin, W, Dou, W., Zhou, Z. & Liu, C. (2015). A cloud-based framework for Home-diagnosis service over big medical data. Journal of Systems and Software, 102, 192–206.
SLR_144	Chen, Z.Xu, G. Mahalingam. V., Ge.L. Nguyen. J. Yu, W. and Lu, C.; 2015. A Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures. Big Data Research.
SLR_145	Xu. G. Yu. W^{T}. Chen. Z. Zhang. H. Moulema. P., Fu.X. & Lu. C. (2015). Acloud computing b asedsv stanfor cyber se curitv management International Journal of Parallel, Emergent and Distributed Systems, 3 Of1). 29–45.
SLR_146	Smowton, C., BaHa. A, Antoruades.D.,Miller. C., Pallis, G., Dikaiakos,M.-D.,&Xmg. W. (2015). A cost-effective approachto improving performance ofbig genomic data analyses in clouds. Future Generation Computer Systems.
SLR_147	Simonet, A, Fedak, G., &Ripeanu, M. (2015). Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures. Future Generation Computer Systems, 53,25–42.

Appendix A (*continued*)

Paper code	Citation
SLR_148	Zhang, F, Cao, J., Hwang, K, Li. K. & Khan, S. U. (2015). Adaptive Workflow Scheduling on Cloud Computing Platforms with Iterative Ordinal Optimization. IEEE Transactions on Cloud Computing, 3(2), 156–168
SLR_149	Merino, J., Caballero, I., Rivas, B., Serrano, M., &. Piattini, M. (2015). A Data Quality in Use model for Big Data. Future Generation Computer Systems.
SLR_150	□ie-Zudor,E.Ekart, A.,Kemenv, Z. Buckingham. C. Welch. P. & Monostori. L. (2015). Advancedpiedictive-analvsis-based decision support for collaborativelogistics networks.Supply Chain Management: An International Journal, 26(4), 369–388.
SLR_151	Wang, Y., & Ma. X. (2015). A General scalable and elastic content-based publish[1] sub scribe service. IEEE Transactions on Parallel and Distributed Systems, 26(8), 2100–2113.
SLR_152	Bhattadiaijee, S., Rahim, L. B. A, & Aziz, I. B. A. (2015). ALossless CompressionTeehniquetc Increase Robustnessin Big Data TransmissionSvstem. International Journal of Advances in Soft Computing & its Applications, 7(3), 126–145.
SLR_153	Phillip s-Wren. G., & Hoskisson. A. (2015). An analytical joumev towards big data. Journal of Decision Systems. 24(1). 87–102.
SLR_154	Rexit,R.,Tsui, F. R., Espino, J., Cbiysar!his,P·K_, Wesaratdialdt, S.,& Ye, Y. (2015). An analytics appliance for identifying (near) optimal over-the-counter medicine products as health indicators for influenza surveillance. Information Systems, 48.151–163.
SLR_155	Kolomvatsos. K_, Ana gno stop oulos, C. & Hadjiefthvmiades, S. (2015). An Efficient Time Optimized Scheme for Progressive Analytics in Big Data. Big Data Research, 2(4), 155–165.
SLR_156	*A*1 Nuaimt- E. *A*1 Neyadi. H. Mohamed. N., & Al-Jaroodt. J. (2015). Applications of big data to smart cities. Journal of Internet Services and Applications. 6(1). 1–15.
SLR_157	Xardulli. P. F. Althaus. S. L. & Haves. M. (2015). A Progressive Supervised-leaming Approach, to Generating Rich Civil Strife Data. SociologicalMethodology: 43(1). 14S—1S3.
SLR_158	Dens. Z. Hu. 1. Zhn. M. Huang. X. Sc Du. B. (2015). A scalable and fast OPTICS for clustering trajectory bie data. Cluster Computing. 13(2). 549–562.
SLR_159	Baek. J. Vu. QH Ltu. J. K. Huang. X. & Xiang. Y. (2015). A secure cloud computing based framework for big data information management of smart grid. IEEE Transactions on Cloud Computing. 3(2), 233–244.
SLR_160	Tang. Z. Jiang, L. Zhou. J. Li. K. & Li. K. (2015). A self-adaptive scheduling algorithm for reduce start time. Future Generation Computer Systems. 43. 51–60.
SLR_161	Zhang. Q.T. Liu. Y. Zhou. W. Sc Yang. Z.W. (2015). A Sequential P. egress ion Model for Big Data with Attributive Explanatory Variables. Journal of the Operations Research Society of China. 3(4). 475-4SS.
SLR_162	Tans. C. Liu. C. Zhang. X. Nepal. S. & Chen. J. (2015). A time efficient approach for detecting errors in big sensor data ondoud. IEEE Transactions on Parallel and Distributed Systems. 26(2). 329–339.
SLR_163	Hong. S. Kan. H. Kim T. & Chang. J. (2015). A User Access Control Scheme for Reducing Authentication Kevs in Cloud Systems. International Journal of Security and its Applications. 9(4). 217–228.
SLR_164	Crosas. M. King. G. Honaker. J. Sc Sweeney. L. (2015). Automating Open Science for Big Data. The .4XN.4LS of the American Academy of Political and Social Science. 659(1). 260–273.
SLR_165	Chandler. D. (2015). A world without causation: big data and die coming of age of posthumanism. Millennium-Journal of Internationa! Studies. 43(3). S33-S51.
SLR_166	Brown-Liburd. H. Issa. H. & Lombardi. D. (2015). Behavioral implications of Big Data's impact on audit judgment and decisionmaking and future research directions. Accounting Horizons. 29(2). 451–468.
SLR_167	Zoomers. A. Getter. A. & Schafer. M. T. (2015). Between two hypes: Will “big data[1] help unravel blmd spots in understanding the “global landrush?” Geoforum. 69. 147–159.
SLR_168	Daniel. B. (2015). Big Data and analvtics in higher education: Opportunities and challenges. British Journal of Educational Technology. 46(5). 904–920.
SLR_169	Vasarhelvi. M. A. Kogan. A. Sc Tutde. B. M. (2015). Big data in accounting: An overview. Accounting Horizons. 29(2). 381–396.
SLR_170	Chan. A. (2015). Bie data interfaces and the problem of indusion. Media. Culture & Society. 37(7). 107S—10S3.
SLR_171	Zuboff. S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology, 30(1), 7 5–89
SLR_172	Hindman. M. (2015). Building better models prediction, replication, and machine Learning in die social sciences. The ANNALS of the American Academy of Political and Social Science. 659(1). 4S-62.
SLR_173	Kumar. M. Sc Radi. S. K (20i5). Classification of microarrav using MapReduce based proximal support vector machine classifier. Knowledge-Based Systems. 39,584–602.
SLR_174	Yue. X. Cat. H. Yan. H. Zou. C. Sc Zhou. K. (2015). Cloud-assisted industrial cvber-physical systems: An insight Microprocessors and Microsystems. 39(S). 1262–1270.
SLR_175	Kumar. X. Misra. S. Rodrigues. J. X. 3c Obaidat. M. S. (2015). Coalition games for spatio-temporal big data in internet of vehicles environment: a comparative analysis. IEEE Internet of Thing: Journal. 2(4). 310–320.
SLR_176	Abawajv. J. (2015). Comprehensive analysis ofbig data variety landscape. International Journal of Parallel Emergent andDktributedSystems. 30(1), 5–14.
SLR_177	Silva. A. 3c Antunes. C. (2015). Constrained pattern mining in the new era Knowledge and Information Systems. 47. 489–516.
SLR_178	Tans. H. & Fong. S. (2015). Countering die concept-drift problems in bie data bv an incrementally optimized stream mining model. Journal of Systems and Software. 102.15S-i66.
SLR_179	Earley. C. E. (2015). Data analvtics in auditing: Opportunities and challenges. Business Horizons. 53(5),493–500.
SLR_180	Ramirez-Gallego. S. Garda. S., Mourmo-Talm. H., Martmez-Rego. D., Bolon-Canedo. V., Alonso-Betanzos. A. Benitez, J. M., Sc Herrera. F. (2015). Data discretization: taxonomy and big data challenge. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Dkcoverv. 6(1). 5–21.
SLR_181	Strauti. S. (20,150- Datafication and the Seductive Power of Uncertainty - A Critical Exploration of Big Data Enthusiasm. Information. 6(4). S36-S47.
SLR_182	Zhang, S. Z. Qu, X. K. & Sim. J. B. (2015). Data Integration and Minins based on Web Big Data. International Journal of Multimedia and Ubiquitous Engineering, 10(6). 123–130.
SLR_183	Abdullah. X. Ismail. S. A. Sophiavati. S. and Sam. S.M. 2015. Data Quality in Big Data: A Review. International Journal of Advances in Soft Computing & Its Applications. 7(3).
SLR_184	Easton-Calabria. E. Si Allen. W. L. (2015). Developing ethical approaches to data and civil society: from availability to acces sibditv. Innovation: The European Journal of Social Science Research. 23(1). 52–62.
SLR_185	Edwards. R. Sc Fenwick. T. (2015). Digital analytics in professional work and learning. Studies in Continuing Education. 1–15.
SLR_186	O'Brien. D T. Sampson. R. J. & Winship. C. (2015). Ecometrics m the Age of Big Data Measuring and Assessing “Broken Windows” Using Large-scale Administrative Records. Sociological Methodology 43(1). 101–147.
SLR_187	Ma. M, Wang. P. Chu. C. H. & Liu. L. (2015). Efficient Multipattem Event Processing Over High-Speed Train Data Streams. IEEE Internet of Thing: Journal. 2(4). 295–309.
SLR_188	Jiang. D. Chen. G. Ooi. B. C. Tan. K. L. Si Wu. S. (2014). epiC: an extensible and scalable svstem for processing big data. Proceedings of the VLDB Endowment. 7(7), 541–552.
SLR_ 89	Liu. C. Yane. C. Zhang. X. & Chen. J. (2015). External integrity verification for outsourced big data in cloud and IoT: A big picture. Future Generation Computer Systems. 42, 58–67.
SLR_190	Hesse. B. W, Moser. P. P. & Rilev. W. T. (2015). From Big Data to Knowledge in die Social Sciences. The ANNALS of the. Imerkan Academy of Political and Social Science. 659(1). 16–32.
SLR_191	Chow-White. P. A. MacAulay. M. Charters, A., Chow. P. (2015). From die bench to the bedside in die big data age: ethics and practices of consent and privacy for clinical genomics and personalized medicine. Ethics and Information Technology. 17(3). 189–200.
SLR_192	Zhao. L. Chen. L. Ranjan. R. Choo. K K R. & He. J. (2015). Geographical information svstem parallehzation for spatial big data processing: A Review. Cluster Computing. 1–14.
SLR_193	Lee. J. G. & Kang. M. (2015). Geospatial Big Data: Challenges and Opportunities. Big Data Research. 2(2), 74—S1.
SLR_194	Song. J. Guo. C. Wang. Z. Zhang, Y. Yu, G. & Pierson. J. M. (2015). HaoLap: a Hadoop based OLAP svstem for big data. Journal of Systems and Software. 102.167–181
SLR_195	Qian. J.Lv.P. Yue. X. Liu. C. 3c Jing. Z. (2015). Hierarchical attribute reduction algonduns for big data using MzpRsiv.ee. Knowledge-Based Systems. 73. 18–31.
SLR_196	Dou. W. Zhang. X. Liu. J. &Chen. J. (2015). HireSome-EI: Towards privacy-aware cross-cloud service composition for big data applications. IEEE Transactions on Parallel and Distributed Systems. 26(2). 455–466.

(continued on next page)

Appendix A (*continued*)

Paper code	Citation
SLR_197	Tavlor. L. & Schroeder. R. (2015). Is bigger better? Hie emergence of big data as a tool cor international development policy. Geo Journal. 30(4). 503–518.
SLR_198	Mohebt. A. Aghabozorgi. S. Ying Wah. T. Herawan. T., Sc Yahvapour. R. (2015). Iterative big data clustering algorithms: A Review. Software: Practice and Experience. 46(1). 107–129.
SLR_199	Wu, X. Fan. W, Peng. J. Zhang. K. Sc Yu. Y. (2015). Iterative sampling based frequent items et mining for big data International Journal of Machine Learning and Cybernetics. 6(6). S75-SS2.
SLR_200	Kolilert. M. Sc Kbnig. A. (2015). Large, high-dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization m polymer film industry. Neural Computing and Applications. 26(3). 581–588.
SLR_201	Berrar. D. (2015). Learning from automatically labeled data: case studv on dick fraud prediction. Knowledge and Information Systems. 46(2). 477–490.
SLR_202	Hsn. C. H. Slagter. K D. Sc Chung. Y. C. (2015). Locality and Loading aware virtual machine mapping techniques for optimizing communications in MapReduce applications. Future Generation Computer Systems. 33.43–54.
SLR_203	Liao. Z., Yin. Q. Huang. Y., Sc Sheng. L. (2015). Management and application of mobile big data. International Journal of Embedded Systems. 7(1). pp.63–70.
SLR_204	Chen. Y. Li. F. Sc Fan. J. (2015). Mining association rules in big data with NGEP. Cluster Computing. 13(2). 577-5S5.
SLR_205	Nita, M. C. Pop. F. Voicu. C. Dobre. C., & Xhafa. F. (2015). MOMTH: multi-objective scheduling algorithm of manv tasks m Hadoop. Cluster Computing. 13(3). 1011–1024.
SLR_206	F eniminella. M. Xunzi. E. Reali. G. Sc Valocchi. D. (2015). Networking issues related to delivering and processing genomic big data. Internationa! Journal of Parallel. Emergent and Dktributed Systems. 30(1). 46–64.
SLR_207	Constantiou. I. D, & Kallmikos. J. (2015). New games, new rules: big data and the changing context of strategy. Journal of Information Technology. 30(1). 44–57
SLR_208	Li. F. He. J. Hnang. G. Zhang. Y. Shi. Y. Sc Zhou. R. (2015). Node-coupling dustering approaches for link prediction. Knowledge-Based Systems. 39.669-6S0.
SLR_209	W:alkowiak. K. W:oiniak. M. Khnkowski. M. Sc Kmiecik. W. (2015). Optical networks for cost-effident and scalable provisioning ofbig data traffic. Inter. Journal of Parallel, Emergent & Distributed Systems. 30(1). 15-2S.
SLR_210	Ludwig. N. FeuemegeL. S. & Neumann D. (2015) Putting Big Data analvtics to work: Feature selection for forecasting electricrtv prices using the LAS SO and random forests. Journal of Decision Systems. 24(1). 19–36
SLR_211	Bolon-Canedo. V. Sanchez-Marono. X. Sc Alonso-Betanzos. A. (2015). Recent advances and emerging challenges of feature selection in die context ofbig data. Knowledge-Based Systems. 36. 33–45.
SLR_212	Qtu. J. L. (2015). Reflections on Big Data: 'Just because it is accessible does not make it ethical'. Media. Culture & Society: 37(7), 1089–1094.
SLR_213	Ma. Y. Wu. H. Wans. L. Hums. B. Ranjan. R. Zomava. A. & Jie. W. (2015). Remote sensing big data computing: Challenges and opportunities. Future Generation Computer Systems. 51.47–60.
SLR_214	V as lie. M. A. Pop. F. Tumesnu. R. I. Cristea. V. Sc Kolodztej. J. (2015). Resource-aware hvbnd scheduling algcnthm m heterogeneous distributed computing. Future Generation Computer Systems, 51. 61–71.
SLR_215	Barkhordari. M. Sc Xiamanesh. M. (2015). ScaDiPaSi:. An effective scalable and distributable MapReduce-based method to fmdpatient similarity on huge healthcare networks. BigData Research. 2(1). 19–27.
SLR_216	Sun. M. Zhuang. H. Li. C. Lu. K. & Zhou. X. (2015). Scheduling algorithm based on prefetching in MapReduce clusters. Applied Soft Computing. 33.1109-111S.
SLR_217	Jin. X. Wah. B.W. Cheng. X. & Wang, Y. (2015). Significance and challenges ofbig data research. Big Data Research. 2(2). 59–64.
SLR_218	Zezula. P. (2015). Similarity Searching for die Big Data. Mobile Networks and Applications. 20(4). 4S7–496.
SLR_219	Anshari. M. and Alas. Y. 2015. Smartphones habit3, necessities, and big data challenges. The Journal of High Technology Management Research. 26(2). 177–185.
SLR_220	Wans. J. Sc Li. X (2015). Task scheduling for MapReduce in heterogeneous networks. Cluster Computing. 1–14.
SLR_221	Skeggs. B. Sc YuiL S. (2015). The methodology of a multi-model project examining how Facebook infrastructures social relations. Information. Communication & Society. 1–17.
SLR_222	Hashem. I. A. T. Yaqoob. 1. Anuar. N. B. Mokhtar. S. Gani. A. Sc Khan. S. U. (2015). The rise of "big data" on doud computing: Review and open research issues. Information Systems. 47.98–115.
SLR_223	Martmez-Prieto. M. A. Cuesta. C. E. Arias. M. & Fernandez. J. D. (2015). The solid architecture for real-time management ofbig semantic data. Future Generation Computer Systems. 47. 62–79.
SLR_224	Romero. O. Herrero. V. Abeho. A. & Ferrarons. J. (2015). Tuning small analytics on Big Data: Data partitioning and secondary indexes in die Hadoop ecosystem. Information Systems. 54.336–356.
SLR_225	Xiang. Z. Schwartz. Z. Gerdes. J. H. Sc UysaL M. (2015). What can big data and text analvtics tell us about hotel guest experience and satisfaction? International Journal ofHospitalin Management. 44.120–130.
SLR_226	Li. Y. & Guo. Y. (2016). W'iki-Health: from quantified self to self-understanding. Future Generation Computer Systems. 56. 333–359.
SLR_227	Ahmad. N. B. Ishak. M. K. Alias. U. F. & Mohamad. N. (2015). An Approach for *E*-Leaming Data Analytics using SOM Clustering. International Journal of Advances in Soft Computing & its Applications. 7(3), 94–112.

References

Abawajy, J. (2015). Comprehensive analysis of big data variety landscape. *International Journal of Parallel, Emergent and Distributed Systems*, *30*(1), 5–14.

Abawajy, J. H., Kelarev, A., & Chowdhury, M. (2014). Large iterative multitier ensemble classifiers for security of big data. *IEEE Transactions on Emerging Topics in Computing*, *2*(3), 352–363.

Abdellatif, T. M., Capretz, L. F., & Ho, D. (2015). Software analytics to software practice: a systematic literature review. *Proceedings of the 1st International Workshop on BIG Data Software Engineering – IEEE Press* (pp. 30–36).

Agarwal, R., & Dhar, V. (2014). Editorial – big data, data science, and analytics: the opportunity and challenge for is research. *Information Systems Research*, *25*(3), 443–448.

Akerkar, R. (2014). *Big data computing.* Florida, USA: CRC Press, Taylor & Francis Group.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 1–15.

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.

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: hyped up aspirations or true potential. *Vikalpa. The Journal for Decision Makers*, *38*(4), 1–11.

Barbierato, E., Gribaudo, M., & Iacono, M. (2014). Performance evaluation of NoSQL big-data applications using multi-formalism models. *Future Generation Computer Systems*, *37*, 345–353.

Barnaghi, P., Sheth, A., & Henson, C. (2013). From data to actionable knowledge: big data challenges in the web of things. *IEEE Intelligent Systems*, *28*(6), 6–11.

Berners-Lee, T., & Shadbolt, N. (2011). There's gold to be mined from all our data. The Times, London 1:1–2. Online Available at: http://www.thetimes.co.uk/tto/opinion/columnists/article3272618.ece [Accessed on 21st April 2016].

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big Data, open government and e-government: issues, policies and recommendations. *Information Polity*, *19*(1, 2), 5–16.

Bhimani, A., & Willcocks, L. (2014). Digitisation, Big Data and the transformation of accounting information. *Accounting and Business Research*, *44*(4), 469–490.

Bihani, P., & Patil, S. T. (2014). A comparative study of data analysis techniques. *International Journal of Emerging Trends & Technology in Computer Science*, *3*(2), 95–101.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662–679.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of Big Data? *The McKinsey Quarterly*, *4*, 24–35.

Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big Data analytics for security. *IEEE Security and Privacy*, *6*, 74–76.

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E., Jr., & Mitchell, T. (2010). Toward an architecture for never-ending language learning. *Proceedings of the Conference on Association for the Advancement of Artificial Intelligence* (pp. 1306–1313).

Chen, C. L. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Information Sciences*, *275*, 314–347.

Chen, G., Chen, K., Jiang, D., Ooi, B. C., Shi, L., Vo, H. T., & Wu, S. (2012b). E3: an elastic execution engine for scalable data processing. *Journal of Information Processing*, *20*(1), 65–76.

Chen, H., Chiang, R. H., & Storey, V. C. (2012a). Business intelligence and analytics: From Big Data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, *7*(2), 157–164.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: a survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Crawford, K. (1 April, 2013). The hidden biases of big data. Harvard Business Review Blog. Available at: http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/ (accessed 5 January 2016)

Cukier, K. (2010). The economist, data, data everywhere: A special report on managing information. Online Available at http://www.economist.com/node/15557443 (Accessed on 20th April 2016).

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. International Institute for Analytics. Available Online at http://www.demonish.com/cracker/1431316877_1217a9641e/bigdata-bigcompanies-106461.pdf (Accessed 5th January 2016).

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning.* Harvard Business Press.

David, R. J., & Han, S. K. (2004). A systematic assessment of the empirical support for transaction cost economics. *Strategic Management Journal*, *25*(1), 39–58.

Delbufalo, E. (2012). Outcomes of inter-organizational trust in supply chain relationships: a systematic literature review and a meta-analysis of the empirical evidence. *Supply Chain Management: An International Journal*, *17*(4), 377–402.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. *IEEE international conference on collaboration technologies and systems (CTS)* (pp. 48–55).

Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, *37*, 267–281.

Dwivedi, Y. K., & Mustafee, N. (2010). Profiling research published in the Journal of Enterprise Information Management. *Journal of Enterprise Information Management*, *23*(1), 8–26.

Dwivedi, Y. K., Kiang, M., Lal, B., & Williams, M. D. (2008). Profiling research published in the Journal of Electronic Commerce Research. *Journal of Electronic Commerce Research*, *9*(2), 77–91.

Edwards, R., & Fenwick, T. (2015). Digital analytics in professional work and learning. *Studies in Continuing Education* (pp. 1–15).

Eembi, N. B. C., Ishak, I. B., Sidi, F., Affendey, L. S., & Mamat, A. (2015). A systematic review on the profiling of digital news portal for Big Data veracity. *Procedia Computer Science*, *72*, 390–397.

Frehe, V., Kleinschmidt, T., & Teuteberg, F. (2014). Big data in logistics-identifying potentials through literature, case study and expert interview analyzes. *In GI-Jahrestagung*, 173–186.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Gantz, J., & Reinsel, D. (2012). The Digital Universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC – EMC Corporation. Online Available at http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf (Accessed 16th January 2016).

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and management. *Academy of Management Journal*, *57*(2), 321–326.

Gu, L., Zeng, D., Li, P., & Guo, S. (2015). Cost minimization for big data processing in geo-distributed data centers. *In Cloud Networking for Big Data* (pp. 59–78). Springer International Publishing.

Halevy, A., Rajaraman, A., & Ordille, J. (2006). Data integration: The teenage years. *Proceedings of the 32nd International Conference on Very Large Data Bases* (pp. 9–16).

Hargittai, E. (2015). Is bigger always better? *Potential biases of big data derived from social network sites, The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76.

Hasan, S., Shamsuddin, S. M., & Lopes, N. (2014). Machine learning big data framework and analytics for big data problems. *International Journal of Advance Soft Computing Application*, *6*(2), 1–14.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115.

Intel IT Center (2012). Big Data Analytics: Intel's IT Manager Survey on How Organizations Are Using Big Data. Available at: http://www.intel.co.za/content/www/za/en/big-data/data-insights-peer-research-report.html [Accessed 5 Jan. 2016]

Irani, Z. (2010). Investment evaluation within project management: an information systems perspective. *Journal of the Operational Research Society*, *61*(6), 917–928.

Irani, Z., Ghoneim, A., & Love, P. E. (2006). Evaluating cost taxonomies for information systems management. *European Journal of Operational Research*, *173*(3), 1103–1122.

Irani, Z., Sharif, A., Kamal, M. M., & Love, P. E. (2014). Visualising a knowledge mapping of information systems investment evaluation. *Expert Systems with Applications*, *41*(1), 105–125.

Jiang, H., Chen, Y., Qiao, Z., Weng, T. H., & Li, K. C. (2015). Scaling up MapReduce-based big data processing on multi-GPU systems. *Cluster Computing*, *18*(1), 369–383.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, *2*(2), 59–64.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, *15*(6), 43–48.

Jukić, N., Sharma, A., Nestorov, S., & Jukić, B. (2015). Augmenting data warehouses with Big Data. *Information Systems Management*, *32*(3), 200–209.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. *46th Hawaii International Conference on System Sciences (HICSS)* (pp. 995–1004).

Kamal, M. M., & Irani, Z. (2014). Analysing supply chain integration through systematic literature review: a normative perspective. *Supply Chain Management: An International Journal*, *19*(5/6), 523–557.

Karacapilidis, N., Tzagarakis, M., & Christodoulou, S. (2013). On a meaningful exploitation of machine and human reasoning to tackle data-intensive decision making. *Intelligent Decision Technologies*, *7*(3), 225–236.

Khan, M. A., Uddin, M. F., & Gupta, N. (2014). Seven Vs of Big Data understanding Big Data and extract value. *Proceedings of 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1) – IEEE* (pp. 1–5).

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78–85.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic review process research in software engineering. Online Available at http://www.citeulike.org/group/14013/article/7874938 (Accessed on 19th December 2015).

Krishnamurthy, R., & Desouza, K. C. (2014). Big data analytics: the case of the social security administration. *Information Polity*, *19*(3/4), 165–178.

Kumar, A., Niu, F., & Ré, C. (2013). Hazy: making it easier to build and maintain big-data analytics. *Communications of the ACM*, *56*(3), 40–49.

Kune, R., Konugurthi, P. K., Agarwal, A., Chillarige, R. R., & Buyya, R. (2016). The anatomy of big data computing. *Software: Practice and Experience*, *46*(1), 79–105.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032–2033.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Baraba´si, A., Brewer, D., ... Van Alstyne, M. (2009). 'Computational social science'. *Science*, *vol. 323*(no. 5915), 721–723.

Lebdaoui, I., Orhanou, G., & Elhajji, S. (2014). An integration adaptation for real-time Datawarehousing. *International Journal of Software Engineering and its Applications*, *8*(11), 115–128.

Lettieri, E., Masella, C., & Radaelli, G. (2009). Disaster management: findings from a systematic review. *Disaster Prevention and Management: An International Journal*, *18*(2), 117–136.

Liao, Z., Yin, Q., Huang, Y., & Sheng, L. (2014). Management and application of mobile big data. *International Journal of Embedded Systems*, *7*(1), 63–70.

Lu, R., Zhu, H., Liu, X., Liu, J. K., & Shao, J. (2014). Toward efficient and privacy-preserving computing in big data era. *IEEE Network*, *28*(4), 46–50.

Machanavajjhala, A., & Reiter, J. P. (2012). Big privacy: protecting confidentiality in big data. XRDS: Crossroads. *The ACM Magazine for Students*, *19*(1), 20–23.

du Mars, R. (2012). Mission impossible? Data governance process takes on big data. Online Available at http://searchdatamanagement.techtarget.com/feature/Mission-impossible-Data-governance-process-takes-on-big-data (Accessed on 9th January 2016).

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Boston, MA: Eamon Dolan/Houghton Mifflin Harcourt.

Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2016). Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*. http://dx.doi.org/10.1007/s10479-016-2236-y.

MIT Technology Review (2013). The Big Data Conundrum: How to define it? Available Online at https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/ (Accessed 19th May 2016).

Office of Science and Technology Policy (OSTP), Executive Office of the President (2012O). Big data press release final 2. Available http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf (Accessed on 7th October 2015).

Otto, B. (2011). Organizing data governance: findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1), 45–66.

Paris, J., Donnal, J. S., & Leeb, S. B. (2014). NilmDB: the non-intrusive load monitor database. *Smart Grid, IEEE Transactions on*, *5*(5), 2459–2467.

Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. *Journal of Decision Systems*, *24*(1), 87–102.

Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: a systematic review of the evidence. *International Journal of Management Reviews*, *5*(3-4), 137–168.

Polato, I., Ré, R., Goldman, A., & Kon, F. (2014). A comprehensive view of Hadoop research – a systematic literature review. *Journal of Network and Computer Applications*, *46*, 1–25.

Raghavendra, R., Ranganathan, P., Talwar, V., Wang, Z., & Zhu, X. (2008). No power struggles: coordinated multi-level power management for the data center. *In ACM SIGARCH Computer Architecture News*, *36*(1), 48–59.

Rehman, M. H., Chang, V., Batool, A., & Teh, Y. W. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management* (Accepted).

Russom, P. (2013). Managing Big Data. Available Online at: *The Data Warehousing Institute*. [Accessed 5th January 2016] https://tdwi.org/articles/2013/10/01/executive-summary-managing-big-data.aspx

Sandhu, R., & Sood, S. K. (2014). Scheduling of big data applications on distributed cloud based on QoS parameters. *Cluster Computing*, *18*, 1–12.

Savitz, E. (2012a). Gartner: Top 10 strategic technology trends for 2013. Online Available at http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technology-rends-for-2013/ (Accessed on 3rd March 2016).

Savitz, E. (2012b). Gartner: 10 critical tech trends for the next five years. Online Available at http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/ (Accessed on 3rd March 2016)

Shah, T., Rabhi, F., & Ray, P. (2015). Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. *Cluster Computing*, *18*(1), 351–367.

Simonet, A., Fedak, G., & Ripeanu, M. (2015). Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures. *Future Generation Computer Systems*, *53*, 25–42.

Sivarajah, U., Irani, Z., & Jones, S. (2014). Application of Web 2.0 technologies in E-Government: A United Kingdom case study. *2014 47th Hawaii International Conference on System Sciences* (pp. 2221–2230).

Sivarajah, U., Irani, Z., & Weerakkody, V. (2015). Evaluating the use and impact of Web 2.0 technologies in local government. *Government Information Quarterly*, *32*(4), 473–487.

Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., & Philippart, L. (2014). Using big data to improve customer experience and business performance. *Bell Labs Technical Journal*, *18*(4), 3–17.

Su, K., Li, J., & Fu, H. (2011). Smart city and the applications. *IEEE International Conference on Electronics, Communications and Control (ICECC)* (pp. 1028–1031).

Sun, N., Morris, J. G., Xu, J., Zhu, X., & Xie, M. (2014). iCARE: A framework for big data-based banking customer analytics. *IBM Journal of Research and Development*, *58*(5/6), 4-1.

Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media. *6th IEEE international conference on digital ecosystems technologies (DEST)* (pp. 1–6).

Taheri, J., Zomaya, A. Y., Siegel, H. J., & Tari, Z. (2014). Pareto frontier for job execution and data transfer time in hybrid clouds. *Future Generation Computer Systems*, *37*, 321–334.

Ting, K. M., Washio, T., Wells, J. R., Liu, F. T., & Aryal, S. (2013). DEMass: a new density estimator for big data. *Knowledge and Information Systems*, *35*(3), 493–524.

Tole, A. A. (2013). Big data challenges. *Database Systems Journal*, *4*(3), 31–40.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: an overview. *Accounting Horizons*, *29*(2), 381–396.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77–84.

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98–110.

Wang, Y., & Wiebe, V. J. (2014). Big Data Analytics on the characteristic equilibrium of collective opinions in social networks. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *8*(3), 29–44.

Watson, H. J. (2014). Tutorial: big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*(1), 1247–1268.

Web, G. (2007). SensorMap for wide-area sensor webs. Embedded computing. Online Available at http://www.fengzhao.com/pubs/embcomp.pdf (Accessed on 13th March 2016)

Weill, P., & Ross, J. W. (2009). *IT savvy: What top executives must know to go from pain to gain*. Harvard Business Press.

Xu, J. S., Zhang, E., Huang, C. -H., Chen, L. H. L., & Celik, N. (2014). Efficient multi-fidelity simulation optimization. *Proceedings of 2014 winter simulation conference*. GA: Savanna.

Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. *IEEE Network*, *28*(4), 5–13.

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a service and big data. *International Conference on Advances in Cloud Computing (ACC-2012), Bangalore, India* (pp. 21–29).

Zhang, F., Liu, M., Gui, F., Shen, W., Shami, A., & Ma, Y. (2015a). A distributed frequent itemset mining algorithm using Spark for Big Data analytics. *Cluster Computing*, *18*(4), 1493–1501.

Zhang, X., Hu, Y., Xie, K., Zhang, W., Su, L., & Liu, M. (2015b). An evolutionary trend reversion model for stock trading rule discovery. *Knowledge-Based Systems*, *79*, 27–35.

Zhao, Z., Zhang, R., Cox, J., Duling, D., & Sarle, W. (2013). Massively parallel feature selection: an approach based on variance preservation. *Machine Learning*, *92*(1), 195–220.

Zicari, R. V. (2014). Big Data: Challenges and Opportunities. (2014) In R. (Ed.), *Big data computing* (pp. 103–128). Florida, USA: CRC Press, Taylor & Francis Group.