What do we know about the big data researches? A systematic review from 2011 to 2017

Saeed Rouhani[a] , Sayna Rotbei[b] and Homa Hamidi[b]

[a]Department of it management, faculty of management, university of Tehran, Tehran, Iran;
[b]Department of information technology, Mehralborz University, Tehran, Iran

KEYWORDS Big data; systematic review; big data researches

ABSTRACT

Big data are defined as a new phenomenon that can be novel step for improving social life and business condition. Analysing the big data's researches to extract insights by systematic literature review is the main objective of this research. For synthesis systematically, data from 123 articles are extracted and kinds of studies that were usually done on big data area are investigated. The Systematic Review showed: the most studies were published in 2014, also the main journal that published big data's article was 'Big Data Research' and country with highest investigate about big data were 'United State and China'. Beside, most researches were done with analytic background. The main research method was experimental and major research type was case study. Our study proved that the majority of researches carried out around big data focused on data management, and most of them identify 'volume and variety' of as significant challenges of big data. Likewise, 'business analytics' was described in the major benefits.

1. Introduction

Over the recent years, along with progress in information and communication technology, volume and detail of captured data have been increased. These masses of available data that refer to huge amount of structured and unstructured form of data need real-time analysis for creating new opportunities. According to Bolón-Canedo, Sánchez-Maroño, and Alonso-Betanzos (2015), every 2 days scale of data produced are five Exabyte.

Dobre and Xhafa (2014) reported around 2.5 quintillion byte of data, produces in the world every day, 90% of which are unstructured, Furthermore, Sivarajah, Kamal, Irani, and Weerakkody (2017) stated that over 40 Zettabyes of data would be generated, copied and consumed until 2020. To extract value from a variety of immense data-sets, it is necessary that a new generation of technologies and

architectures, be designed economically to be able to acquire, explore and analyse the large amount of data from various sources (Phillips- Wren & Hoskisson, 2015). For this purpose, notice should be given to four major challenges, which include volume, variety, velocity and value. These concepts and developments have led to the emergence of a new concept called big data.

By using potential of big data, firms can utilise information originated from big data, for better understand the situation, obtain insight about the business and improve decision-making. Therefore, business world need developed and designed techniques for storing, processing and analysing to utilise big data's values to provide scalable, fault-tolerant and intelligent systems that can support business environment.

Recently, big data has great impact on business strategy and organisation, by creating new insights for solutions and offer the possibility and the chance for businesses that enable them to be big and getting bigger (Al Jabri, Al-Badi, & Ali, 2017; Kowalczyk & Buxmann, 2014). Thus, academic researchers need to define big data concept to study it meaningfully and communication results. Lately, the necessity to realise how, why, and when Big Data Analytics (BDA) applications can be a noteworthy resource for companies to achieve competitive privilege have been underlined by researchers and practitioners (Abbasi, Sarker, & Chiang, 2016; Agarwal & Dhar, 2014; Côrte-Real, Oliveira, & Ruivo, 2017; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). The BDA value chain subsists relatively undetected and should be better investigated in the future, though, BDA technologies have been identified as the 'next big thing for innovation'(Abbasi et al., 2016; Côrte-Real et al., 2017).

In the point of scientific view, three major layers are used in other articles for the scientific view of development and implementation of big data. Storage, processing and analysis of data are sections of big data over which the authors emphasised as layers of big data in most articles and they have consensus about these layers' importance and their roles on business success. Almost all articles about implementation of big data had concentrated and emphasised somehow on Storage, Process and Analysis. About *storage***,** it can be said that because of the

considerable increase in structured and unstructured data, business world demand to store and manage data more than past (Chen, Mao, & Liu, 2014). Furthermore, for providing on time data for real-time application, big data storage systems require appropriate and trustworthy interface (Kambatla, Kollias, Kumar, & Grama, 2014). In *process*, after data are collected and storage from various sources of data which generate different types of data (Bolón-Canedo et al., 2015), they are classified, integrated and eliminated noise and redundancy of them, for reducing redundant storage space and cater necessary data (Chen et al., 2014). For *analysing*, we require enterprise software and information systems (Wang & Hajli, 2017). After that for optimal use of data, they should be analysed. For this purpose, information technology tutors should be aided to prepare data analysts and scientists who have the skills of a database designer, software programmer and statistician. The job of a data scientist is described in the 2012 in more detail by Davenport and Patil (Davenport & Patil, 2012).

All researches in big data area, only referred to big data's implementation just in the own research area. Among all articles that concern big data area just some review the other articles, for example, Hashem et al. (2015), Kambatla et al. (2014) and Kshetri (2014), but none of the articles did not do their studies by systematic literature review (SLR). In our paper, we present a comprehensive survey of literature based on review 123 selected studies by SLR method, the aim of this study was to investigate and analyse the articles that focused on aspects of the storage, process and analysis of big data. This paper presents our results by SLR, the major contribution of our article includes identifying types of big data researches and trend of big data researches, recognise challenges and benefits of big data.

For deciding to place, papers in our review, all being published papers, were searched by subjects and topics and then abstracts of them that were potential candidates for inclusion in our review were studied more precisely. For data extractions, the following questions were developed with the aid of a pilot study:

Q1: What is big data and what kind of researches have been done about it?

Q2: How many categories do these researches divided to?

Q3: What are the major trends of big data researches?

Q4: What are the challenges of big data implementation?

Q5: What are the benefits of big data for data management, data processing and analytics?

After selecting researches, all parts of articles were studied. Then according to their concept and their construction, some information was extracted from them. For first research question, primarily, we extracted one comprehensive concept for big data and then elicited articles research method. For answering question number 2, kinds of research were extracted. For third question, after reading all parts of article, their main subjects and area were extracted. For two final questions, articles were read and then benefits and challenges of big data that they refer to, were identified.

The paper is organised as follows; Section 2 describes Background and Motivation for SLR. Section 3 illustrate the research method (Systematic Literature Review planning) in this section Publication quality assessment, Study selection criteria, Primary search strategy are described. Section 4 is reporting findings for research questions. In Section 5, outcomes of the research were presented as a conclusion.

2. Background

The big data term coined for referring to the magnitude of data generation that has grown rapidly and became of critical subject for investigation in academic literatures, business and industry. No global definition has been accepted for big data, and this term still has no comprehensive definition. It can be said from a macro perspective that big data can be considered as a link that connects and integrates precisely the physical world, human society and cyberspace. We knew the physical world has a reflection in cyberspace that covers big data through the Internet, the Internet of things and other information technologies, while the human community produces its big data based on mapping in cyberspace by using mechanisms such as human–computer interfaces, brain–machine

interfaces and mobile Internet (Cheng et al., 2014; Jin, Wah, Cheng, & Wang, 2015). That is, big data can essentially be divided into two categories, first, data from the physical world, which is usually captured through sensors, scientific experiments and observations (such as biological data, neural data, astronomical data and remote sensing data) and data from a human society often derived from resources or areas such as social networks, the Internet, health, finance, economics and transportation (Jin et al., 2015). One definition is provided, 'Big data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time' (Seddon & Currie, 2017, p. 2). Using valuable potential of big data to help business, industry and academic society is not possible unless using new advanced and unique technique, architectures and application for storage process and analysis for offering worthiness and improve modern society and business condition. For implementation, information systems must focus on three layers of big data chain including storage, process and analysis.

Extracting hidden value of big data need to collect data from all kinds of source and transform them to databases with specified form (Ward, Marsolo, & Froehle, 2014), and eventually they should be analysed for different targets.

Storage data, for extracting value of data sets for using big data's potential faced some challenges. Given available storage space for accessing and analysing data in traditional database systems, they cannot be fully capable of storing and indexing large data-sets (Kowalczyk & Buxmann, 2014; Murthy & Bowman, 2014). Generally, traditional databases have not been developed for retrieval and storage data from various distributive sources (Murthy & Bowman, 2014). In fact, data storage has principles, these principles should be based on policies of data, adaptation legislations and availability rein (Wang & Hajli, 2017). Additionally, for exploiting worthiness of big data we should do processing. This means that wide variety of data are extracted from various data sources and then processing engines begin to move and clean (Chen et al., 2014); integrate (Chen et al., 2014); split, translate and sort data (Wang & Hajli, 2017). Cleaning identifies inaccurate, incomplete or unreasonable data for removing data's noise, redundancy and store meaningless data (Chen et al., 2014); and integration involves to combine data from

different sources (Chen et al., 2014). In fact processing brings for data consistency, visibility and convenient access and provides for users uniform view of data and improve data quality for analysis in the future (Chen et al., 2014; Wang & Hajli, 2017).

The arrival and processing of data can be performed by three procedures, for some silent applications it can be accomplished in batch, while analytic applications need ongoing and real-time analyses, whereas, this is sometimes required to urgent and quick action on processing of incoming data streams (Wang et al., 2011). Overall, big data's velocity characteristic refers to speed of data generation but sometimes, also, referred to data stream that comes from various sources such as sensors, the Internet of Things (Maciejewski, 2017) and high- speed transactions because of intelligent systems and methods improvement (Phillips-Wren & Hoskisson, 2015).

Analytics,'include statistical models and other empirical methods that are amid creating empirical prediction … as well as methods for assessing the quality of those prediction in practice' (Shmueli & Koppius, 2011, p. 3).

Big data analytics illustrate methods, which help to improve analysing the unstructured and online data in a wide volume (Rekha & Parvathi, 2015).

However, the big data analytics has many unknown aspect that require further investigation for using maximum potential of big data for bringing business value and create relation among knowledge assets, organisational agility and performance (Côrte-Real et al., 2017). It has become clear that organisations which use big data and analysis in their innovation processes, had growth of income and operational efficiency 36% more than their competitors; In fact, more likely big data tools cause 23% and analysis tools 79% improvement in organisational performance (Marshall, Mueck, & Shockley, 2015). The research on big data analytics helps to identify hidden value of big data and use them to improve business performance (Wamba et al., 2017). overall big data analytics is a vital step in design business process and success in business world (Hagel, 2015).

For analysing structured invalid data from the numerous outlooks, data mining is one of the useful tools (Wang & Hajli, 2017). Data mining contains data analysis and finding algorithms to identify templates or models, therefore, for helping to reduce doubt, uncertainty and indecision data mining is one necessary technology

(Kowalczyk & Buxmann, 2014). Also raising the capacities of finding strong templates that can construct the foundation for predictive analytics, big data can be used (Dhar, 2013). Big data mining is a technology for data analysis and refers to the basic technologies to find out the value of big data, relocate data properties and use it effectively. In this way information are extracted from data, knowledge figure out from information, summarise intelligence from knowledge and aid to extract the value of big data (Wang & Yuan, 2014).

Finally, studied articles in big data area showed that the concept of big data has been raised in all three layers of big data; and current study covers all aspects about 'big data's concept', 'big data's architecture', 'big data's analytics' and 'implementation of big data' which studied in recent years.

3. Review plan

Systematic review is a method that enables the evaluation and interpretation of all accessible research relevant to a research question, subject matter or event of interest (Kitchenham, Mendes, & Travassos, 2007). Systematic reviews by applying reliable, precise and auditable procedure with purpose to represent an impartial assessment of a research subject. The aim of this study is to examine all the articles about big data. This is the reason, this research was planned, conducted and presented the review by following the SLR process suggested by Kitchenham and Charters (Wen, Li, Lin, Hu, & Huang, 2012). To this end, the formal protocol at the planning phase of the SLR is developed.

For validity of studies four different constructions were defined by Silva, de Souza, and de Souza (2017) which include: conclusion, internal, construction and external threats. The conclusion validity mentions the relationship between the treatment and the results. Inclusion and exclusion criteria are defined sufficiently accurate for an easy inclusion/exclusion of a paper which is determined for conclusion validity (Silva et al., 2017). Internal validity assesses the relationship between the treatment and the results which is assessed by internal validity (Silva et al., 2017).This relationship must not be a result of a factor on which we have no control or have not considered; In fact, it should be random (Silva et al., 2017). One

of our paper threats is related to works that are unavailable for downloading and unpublished. Therefore, all related studies in defined databases are included. The other validity is construct, it refers to the relationship between theory and observation (Silva et al., 2017). For this validity, inclusion and exclusion criteria exactly are employed to ensure that all related articles would be included and considered relevant databases for selecting articles. According to Silva et al. (2017) generalisation is referred by external validity. In our study, we selected only a few databases, as a result, it is impossible to determine whether the results would be different if other databases were selected.

Our SLR protocol mainly includes nine stages; at first, a context is prepared and then for well placing new research activities, research questions are defined. For comprehensive and perfect searching of early studies, online databases are identified. In forth stage concentrated review questions for search strategy design are formulated, after that, criteria are used for filtering the study, also quality assessment are used for quality evaluation of contained studies, in next phase data extracted by identifying needed data for answering the research question, after identifying needed data, classification of data are performed. Finally, by relevance assessment and data analysis outcomes were interpreted to specify their applicability. The nine stages of the SLR are demonstrated in Figure 1.

3.1. Identification research question

In definition of research question, we set a context of study about big data then detected gaps among big data's research and understood, in academic society there were not enough article which carried out a comprehensive review by SLR method, so based on the above needs and motivations that refer to them in the introduction, in this work, we systematically classify the body of knowledge so, our study focuses on answering the following five research questions:

Q1: What is big data and what kind of researches have been done about it?

Q2: What are the major trends of big data researches?

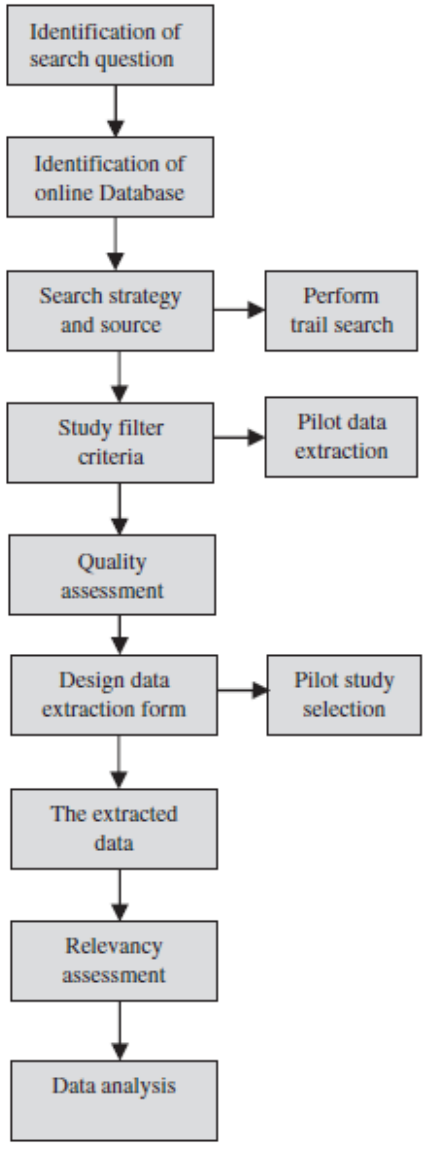

Figure 1. systematic literature review stages.

Q3: How many categories do these researches divided to? Q4: What are the challenges of big data implementation?

Q5: What are the benefits of big data for data management, data processing and analytics?

3.2. Identification online database

In this phase, to find relevant studies, we searched seven major online academic article (specifically conference proceedings, and journal papers) databases. The selected databases are known for including many of the studies related to big data, the name of digital libraries were: ACM, Emeralds, Springer, Taylor and Francis, IEEE, SAGE and IGI.

3.3. Primary search strategy

In Primary search, a wide review of early studies in a particular subject are done that target to identify what evidence is existing on the subject. The search strategy includes three phases of a systematic literature review: planning the review, managing the review and reporting the review. They provide a relatively high-level description.

Our initial search strategy included the following steps:

Step 1: Determining the research questions

Step 2: Identification of online Database

Step 3: Extracting main search terms from the research questions

Step 4: Perform pilot tests on major terms

Step 5: Using Boolean operators, obtained terms connected together to create a search string. With this method these five terms are identified to be used at research process: (1) Big Data, (2) Big data management, (3) Big Data Processing, (4) Big Data analytic and (5) Big data architecture. Briefly, our goal was to explore the relationship between big data implementation and management, big data and processing, big data and analytics and big data and architecture.

Step 6: key words verified in some articles in sixth stage.

Step 7: A wide spectrum of online databases, journal archives and conference proceedings were selected for searching. In summary, search strings customised for using at the online databases' interfaces.

Step 8: The seventh step was to retrieve quotes and summaries of results and manage them using the notes. By retaining compatibility logical order, strings customised for several online databases. In initial searches, a wide spectrum of resources selected to decrease search selection bias. They included ACM, Emeralds, Springer, Taylor and Francis, IEEE, SAGE and IGI. In the initial search, no restrictions were applied on the year of publication.

Step 9: In this stage, the extracted data are carefully classified.

Step 10: This stage included assessment of data irrelevant to the research subject.

Step 11: obtain primary papers: big data implementation.

3.4. Study selection criteria

By selection criteria, just those articles, which are related to the research questions, were selected. Only those papers considered that are relevant to big data. Inclusion and exclusion criteria are categorised in Table 1.

3.5. Publication quality assessment

Publications Selection includes steps that are shown in Figure 2. After final selection of publications, we followed Guinea, Nain, and Le Traon (2016) and Wen et al. (2012) way for quality assessment; So that, all articles were fully studied and the answers of 10 questions were searched in them, if there were adequate answers to all questions in the text, the article remained in the research process, otherwise it was removed from the research process. At the same time while extracting data, publications quality is also assessed.

The following questions are used for assessing quality:

- QA1: Were the research objectives clearly describe?
- QA2: Have determined research objectives been achieved?
- QA3: Were the assessment contents adequately described?

- QA4: Are evaluation procedures well described and investigated?
- QA5: Is there a justification for experimental design? Moreover, is experimental design appropriate?
- QA6: Is there sufficient data sets for doing experiment?
- QA7: Do the expressing and reporting the results support the findings of the research?
- QA8: Have the restrictions of the research been clearly analysed?
- QA9: Does the research add value to academia or industry community?
- QA10: Were the research findings obviously stated?

Each of the above factors will be marked as 'YES' or 'NO' or 'Partial' or 'N.A'.

Overall, all selected articles had clearly described about their objectives and findings, but nine papers were eliminated because, they did not determine research objectives. Fourteen

Table 1. inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Studies that describe challenge of big data implementation	Articles other than English language
Studies that describe benefits of big data for data management	Studies those are not relevant to the research questions
Studies that describe benefits of big data for data processing	Articles outside the domain of big data
Studies that describe benefits of big data for data analytics	Not using any experimental research procedure
Studies that describe architecture of big data implementation	Because relevant publications used in the PhD and Master theses were searched. PhD and Master Theses were eliminated from the study
Papers that did not focus on aspects of big data implementation	Books were excluded and Papers were published in conference journals but papers were published in conferences and proceedings which available in Scimago Journal & Country Rank (SJR) website remained in study
For duplication publication of the same study, only the most complete and newest Studies that describe challenge of big data implementation	Completely irrelevant articles that were obtained from inefficient search engines online

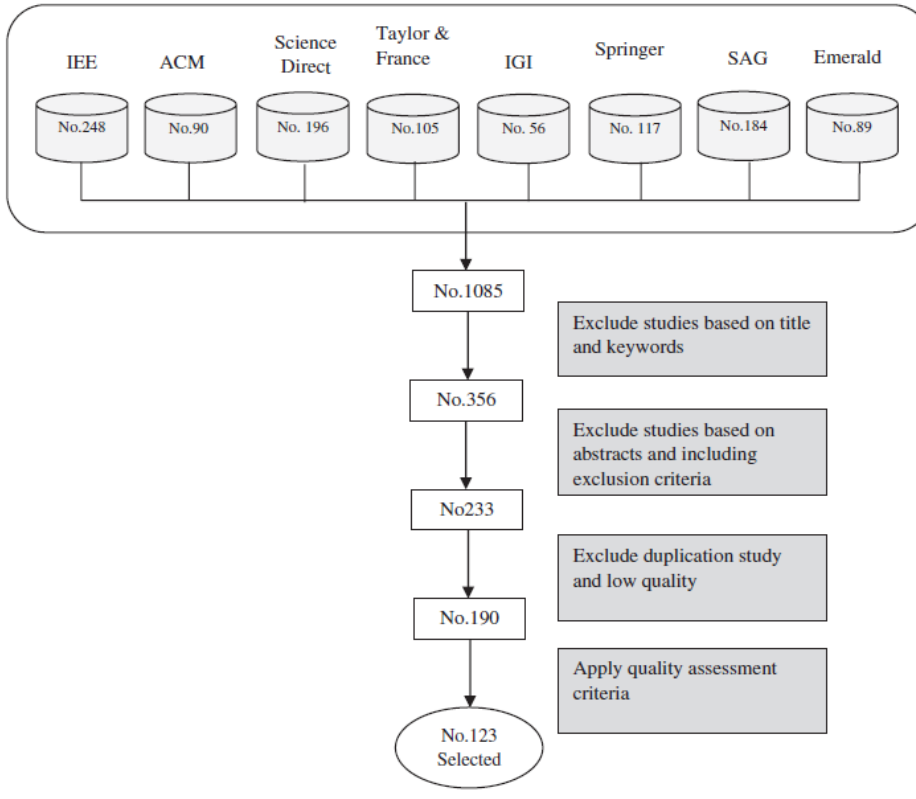

Figure 2. the selection process of publications.

articles go out from search process because they cannot answer questions number QA3, QA4. Five papers were pretermitted for QA5, and because of AQ6 and AQ7 exactly 27 researches were deleted from investigate on process. During the review, 12 papers were omitted because they cannot clearly analyse the restrictions of the research. Also, they cannot add value to academia or industry community.

3.6. Design extraction form

For executing data from 123 articles that are remained in survey process (Appendix 1), a spreadsheet with 11 columns in three main groups including publication details, context and finding is designed. Publication details group was designed for obtaining general knowledge about each paper, seven columns are defined including: research title, citation, authors, year, country and paper's keywords and Journal/Publication/Conference name. Publication details group was used for

catching information on when big data's articles were published, which journals published the most of studies about big data, which countries focus on big data more than others. The next group was context that helps us to know some information about article's background (storage, process and analytics), Research types (Concept, Implementation, Tools and Case), Research Field, Research method (Review, Survey, Case Study, Experiment, Field Study and Grounded) and research concentrate (Infrastructure, Data management, process and Analytics). Finally, in the last group four columns are located that in the two of them insight of each paper about big data system design and implementation were written and in the third column challenges of big data are extracted and in the last column big data's benefits are elicited.

3.7. Extracted data

The data extraction phases of an SLR purposes to identify the relevant information that should be extracted from each paper, in order to answer the research questions. In this investigate, all 123 articles have been studied and related data have been extracted from each article; and spreadsheet were filled by extracted data.

3.8. Relevancy assessment

For relevancy assessment, we scanned and reviewed all the references of included studies, and evaluated all spreadsheet's data which extract from its article. All eligible researches were applied for having related data according to column's concept and investigates targets.

3.9. Data analysis

One of the objectives of this research was to categorise data about big data's concept and collected them into a set of categories. We classified research papers to provide a general picture of the characteristics of big data research. This general picture served as a starting point for deeper investigation of findings that seemed, from the authors' perspectives, to suggest important shortcomings in estimation research and possibilities for improvement.

4. Results

In this section, we describe the characteristics of our 123 studies.

The Figure 3 shows the number of literature that are published around big data subject, 123 studies that started almost after 2011, had an increasingly trend until 2014. The most

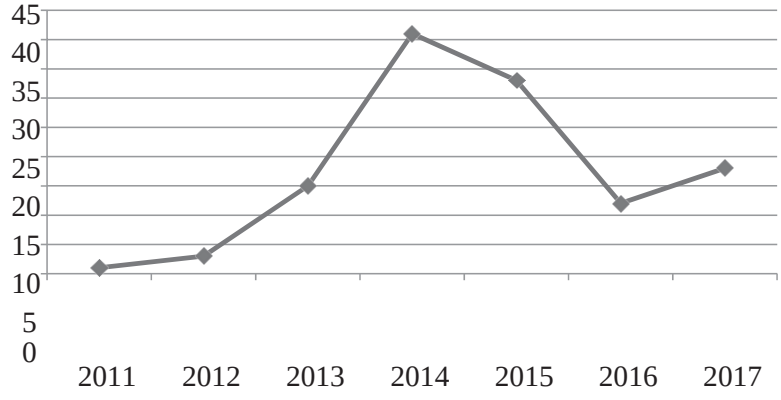

Figure 3. number of published literature about big data from 2011 until 2017.

of studies published in 2014 with 41 articles. For next two years, the number of articles had declined but in 2017, it had growth.

The Figure 4 shows 17 journals that published the most of studies about big data. The most of articles are published in 'Big Data Research' journal in the next level, respectively, 'Journal of Business Research', 'Big Data & Society' and 'Procedia Computer Science' are placed. Overall, by survey journal's H-index it was found that there was not any logical relationship between journal's H-index and the number of studies about big data they published. For example, 'Big Data Research' journal with

13 articles about big data had H-index = 3 and ‘Journal of Business Research’ with 7 studies had H-index = 114, while H-index ‘Environmental Modelling & Software’ with two articles was 86. The top 10 journals in results of SLR can be seen in Table 2.

The ranking shown in Figure 5 represents the number of studies conducted on big data in different countries. This ranking shows 11 countries that had most articles about big data, the United States and China with 24 articles about big data had most study about big data. In the next place India and the United Kingdom are placed. The ranking of Figure 5 that picked numbers of researches in different countries, almost is similar to the ranking of gross domestic product (GDP) countries, especially in the first levels. In this figure like the list of GDP, the United States placed in the first and China was next step.

The Table 3 shows the five articles that have most been cited in Google Scholar website among other studies. In general, as indicated in the table, ‘Data mining with big data’ experience highest cited with 1194. In the next level ‘Big Data: A survey’ was placed by 936 cited and other articles are located in the next levels. Also, the lowest number of citations among all articles was 0, and the average citations were approximately 72. (based on Google scholar citation on date of 2017/10/15)

The Figure 6 shows 10 keywords that are used most among other keywords in the articles about big data. The most of them were ‘big data’ that was used in 94 studies. In the next level, key word ‘analytics’ was used more than other keywords. ‘Analytics’, ‘data analytics’,

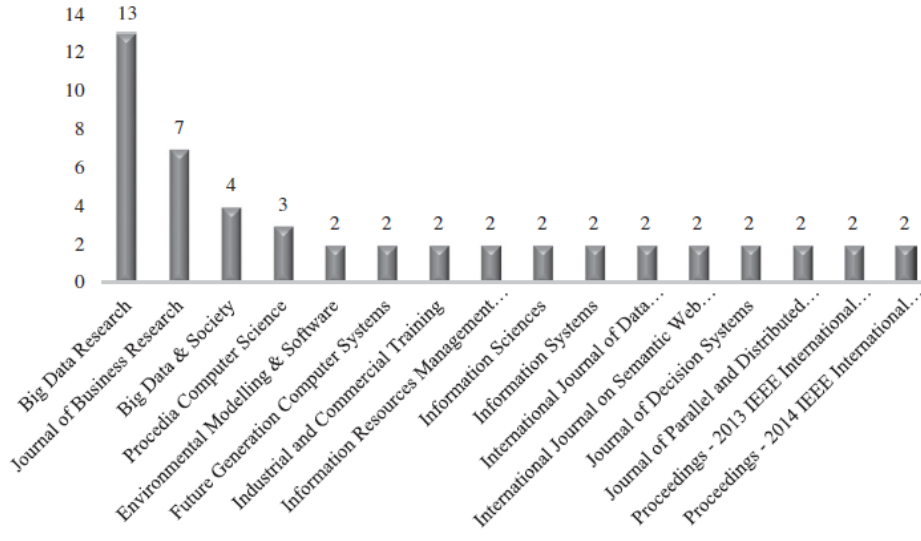

Figure 4. Journals that published the most of studies about big data.

Table 2. top ten journals that published the most of studies about big data.

	Top ten journals in results of SLR	Article number
1	Big Data Research	13
2	Journal of Business Research	7
3	Big Data & Society	4
4	Procedia Computer Science	3
5	Environmental Modelling & Software	2
6	Future Generation Computer Systems	2
7	Industrial and Commercial Training	2
8	Information Resources Management Journal	2
9	Information Sciences	2
10	Information Systems	2

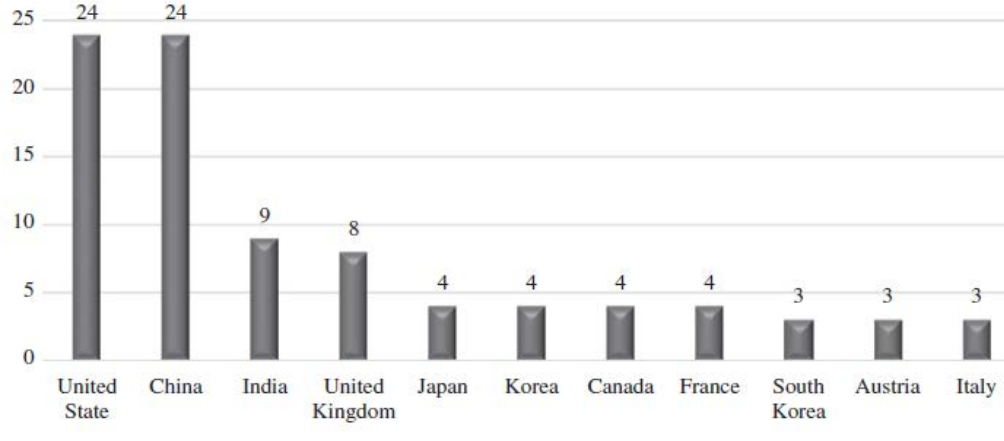

Figure 5. Countries with most articles about big data.

Table 3. Highly cited articles.

	Research Title	Citation count	Reference
1	Data Mining with Big Data	1194	(Wu et al., 2014)
2	Big Data: A Survey	936	(Chen et al., 2014)
3	Data-intensive applications, challenges, techniques and technologies: A survey on Big Data	845	(Chen & Zhang, 2014)
4	The rise of 'Big Data' on cloud computing: Review and open research issues	594	(Hashem et al., 2015)
5	Big Data, new epistemologies and paradigm shifts	427	(Kitchin, 2014)

'analysis', 'big data analytics' are ranked in the next place of this keywords' order, that showed wide part of big data's studies' concentration were 'big data analytics' researches. This shows that their main aim of investigation about big data is big data analysis and use its hidden value. For using big data's worthiness and harness, in academic and business community, they should manage big data's characteristics and for this propose in the next level after big data analysis, management is located. Though other keywords proved that implementation of big data applications in terms of scientists, engineers and academic and business world is important, but the main concern is how they can use big data's valence to improve their business and life condition.

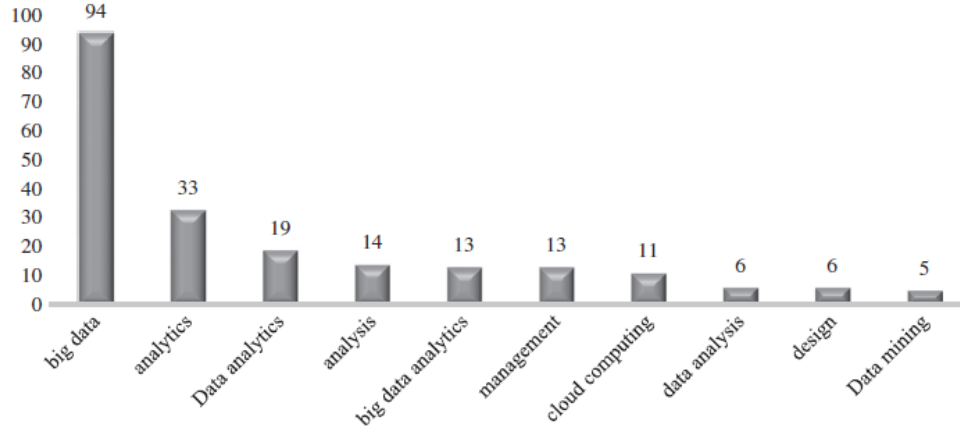

Figure 6. most used keywords among other key words.

The Figure 7 demonstrates seven researches' backgrounds on which articles focused. Moreover, it demonstrated percentage of researches that focus on different area of big data, which indicates the importance of that field.

It is clear that the 50% of them focus on analytics of big data that shows analytics of big data is an important aspect of big data research, while only a very small percentage, i.e. 2% of them studied on storage and analytics of big data at the same time. About 16% focus on three aspects of big data including analytics, storage and process. Furthermore, 5% of papers studied the process/ storage and 4% of them just focus on storage. Totally, chart illustrates analysis by assigning half of studies to itself is a vital field of big data. Because it can be extracted big data's value for improving business and social life condition by analysis. On the other hand, storage has a least important among other fields. It is possible, since now there are some technologies that can storage big data from all kind of resources.

5. Analysis of findings

Based on the data extraction from our set of 123 studies and systematic coding and analysis, we now present our findings to answer the research questions.

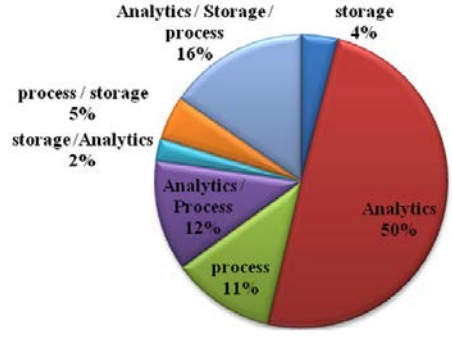

Figure 7. classification articles background.

5.1. Q1 – What is big data and what kind of researches have been done about it?

Nowadays with occurrence of technological revolutions such as social media, people are able to produce data faster and further than ever before (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). In recent years, we named these data as a big data which is defined as 'the amount of data just beyond technology's capability to store, manage and process efficiently' (Fan, Lau, & Zhao, 2015, p. 1).

Table 4 demonstrates that the highest number of research method is experimental. This kind of studies usually focus on experiment their findings and theories on real environment and used their findings in articles results and this shows that it is very important for scientists and experts to test their theories and know about their mount of accuracy and reliability. On the other side, the lowest of them are field study and grounded research. Table 5 presents types of research method and the number of them.

5.2. Q2 – How many categories do these researches divided to?

Several studies were done on a large variety of data, including: case, concept, tools and implementation study. Case studies refer to investigation on some special cases, for example, a company, application, industry and …, as an instance Xin, Wang, Qu, and Wang (2015) in their article they present a novel Elastic Extreme Learning Machine based on MapReduce framework, these kinds of papers are classified on case group. Concept term refers to the papers that try to describe big data's definition and basic concept. Usually, these kinds of articles review the other articles and try to classify basic concept in order to improve readers' and other experts' knowledge about big data. Usually, articles are named as tool study, peruse on one or more tools for doing some special task on big data, for example, Steed et al. (2013) work on Exploratory Data analysis Environment (EDEN), for improving critical under- standing of earth system processes, they worked on the kind of interactive visual analysis tools that are essential to transform data into insight. Implementation studies, refer to investigation that explain their experience outcomes about their opinions and thoughts on some cases. For instance, describing the acquisition purpose of big data analytics significantly from the theoretical outlooks

of data quality management and data usage experience is suggested by, Kwon, Lee, and Shin (2014); their empirical investigation discloses that com- pany's qualification in protecting the quality of corporate data can positively affect a company's purpose for big data analytics. In addition, company's desirable experience in employing external source data may encourage future acquisition of big data analytics. Astonishingly a company's desirable experience in employing external source data may prevent its adoption purpose of big data analytic. Figure 8 shows the number of research type.

Table 4. Frequency of research method

Research method	Number
Field study	2
Survey	9
Grounded	2
Case study	12
Experiment	56
Review	38

Table 5 researches focus area

Focus area	Number
Analytics	9
Process	1
Infrastructure	19
Data management	42
Data management/ Infrastructure	15
Data management/ Process	3
Data management/Analytics	24
Process/Analytics	3
Analytics/ Infrastructure	2
Data management/Process/Analytics	2
Infrastructure/ Process/Analytics	1
Infrastructure/Data management/Process/Analytics	2

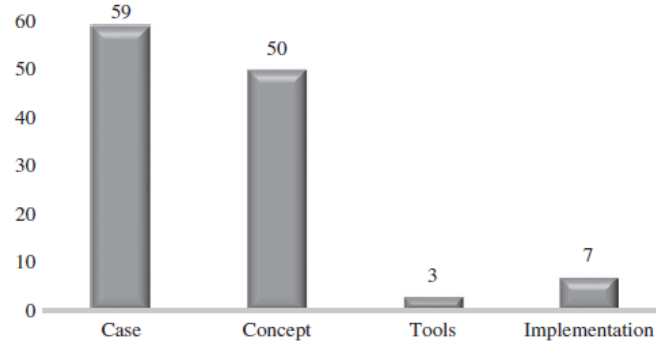

Figure 8 the number of research type.

As Figure 8 shows the largest number are articles that focus on one or several cases to survey big data by 59 articles and the lowest of them focus on big data by studying on tools.

5.3. Q3 – What are the major trends of big data researches?

Our survey showed that most researchers have focused on data management that proved big data management by the right way is one of the main concerns among experts, scientists and business and academic world. It illustrates according to information extracted from articles that if industry and business world cannot manage data definitely they cannot use big data's hidden value. Assunção, Calheiros, Bianchi, Netto, and Buyya (2015) stated that it is explicit that many of the big data challenges include data management, integration and processing. Hence, it is clear that the major articles focus on big data' management. In fact, it can be said that the integration and processing could be aspects of data management. On the other side, the least discussed issue was processing, it seems studies has paid less attention to it because it can be one aspect of data management, it means that if scientists can improve data management, surely, they do best processing on data. After data management, most investigators have focused simultaneously on other aspects of big data including data management / analysis, data management/ infrastructure and just infrastructure. At the same time the issues were considered less than others were infrastructure/data management/processing/analysis, Data management/Process/Analytics and Analytics/ Infrastructure. Focus area are shown in Table 5.

5.4. Q4 – What are the challenges of big data implementation?

According to investigation about big data's challenges, although each one studies in accordance with its subject and study's area refer to variety of challenges including: Data cleansing, acquisition, capture, submerged, different patterns, privacy, quality, safety mechanism, legal/ regulatory issues and governance or lack of clear big data strategies. Almost the most of them refer to

volume with other challenges, Assunção et al. (2015) stated that because in many organisations internally volume of data rise to a very great, or assembling other extensive amounts of data from the external, volume became the first and worst characteristic of big data. In addition, variety of data is one of the main features that distinguish big data from others and cause other big data's challenge. Abawajy (2015) noted that when they survey their organisations, over 50% say their biggest challenge is the variety of data. Besides that, Ian Bertram, Asia Pacific head of research at Gartner, has expressed 'The biggest challenge actually is variety'. From the perspective of Kambatla, a fundamental problem for analysing the big data is storing; So, due to rapidly increasing data production and improvement in hardware, storage platforms should be distributed, scalable, elastic and fault-tolerant (Kambatla et al., 2014). In some articles, also, researchers refer to velocity as a challenge of big data; Seddon and Currie (2017) explained a permanent reduction in time for the transmission, storage, processing and structured data into feed again. According to Fadiya, Saydam, and Zira (2014), data warehouse is traditional type, and its information is raw so apply and handle them according to share time frames for getting the best potential value from it and answer in the real time to a common requirement of advance organisations is inevitable (Fadiya et al., 2014). Hence, nowadays business need to acquire and storage data on time and at the same time analyse them. The ongoing HFT (high-frequency trading) companies try to remove performance impasse and it has conducted them to a technical arms race, with billions of dollars spent on trading infra- structure (Seddon & Currie, 2017). So it can prove that velocity is one of the important challenges. In addition, about 8% articles moreover volume, velocity and verity refer to veracity. Veracity characteristic measures the correctness of data and its potential use for analysis (Vasarhelyi, Kogan, & Tuttle, 2015). This is not just about data quality, in fact, it is more about understanding the data; this means that it should be known whether there are significant differences in the data collected (Sivarajah et al., 2017). The other challenge that papers refer to was value. To extract knowledge/value from too much structured and unstructured data without loss, for end users is demonstrated value; for example, internet users are left behind the

stream of clicks from which considerable values can be extracted that can become spine of the internet economy (Sivarajah et al., 2017). Some big data researchers consider that value is a necessary characteristic, as sometimes there can be high-valued data, although most of those data independently may seem insignificant (Sivarajah et al., 2017). Nowadays establishments are still encountered with challenges of storing, managing and extracting value from the data in a cost-effective procedure (Abawajy, 2015). They believe that it is not easy to earn confidence about veracity of data and extract their value.

Overall, volume and variety are about storage challenge, although variety can depend on process as a same time. Velocity can have a strong relation with storage, process and analysis because of reaction on time. Business need to store, process and analysis data fast. About veracity, there is a same condition with velocity because business needs to storage right data, process and analysis them by right way. However, about value it must be said that, business needs to analyse and process the data by correct way, so it has relationship with process and analysis for extracting the hidden value of data (see Table 6).

5.5. Q5 – What are the benefits of big data for data management, data processing and analytics?

When most articles, approximately 58%, discuss about big data's benefit they refer to acquire 'More precise knowledge exploration', this shows that using big data's hidden value is one important aspect of big data for academic and business world. It proves all efforts in big data field including processes, applications, tools and other attempts in big data follow one main aim that is extraction of big data's conceal potential to improve business condition and earn better position in their business. Following elicit big data's value, however, studies show that easy implementation of big data's system is in the next level of importance for academic world. After that, improvement analytics speed is located with approximately 9%. Also, it is clear that for reacting on time in changes of business world, companies want to implement big data's system by easy way and analyse storage data very fast. As one of big data's potentials Wu, Zhu, Wu, and Ding (2014) refer to big data technologies, they stated that they hope to be able to represent most related and most accurate social evaluation feedback to better

understand our society at real-time. In addition, they hope, they can better motivate the public audiences to take part in the data production circle for societal and economical events. On the other hand, decrease of needed capacity is the least benefit of big data that refer to studies. It can be possible because of new technology, for example, cloud computing. Table 7 includes the per cent of big data benefits.

7. Conclusion

In recent years, digital transformation technologies cause new outlook for humankind in constructing a smart firm, society or city which has led to the development of modern life. Big Following elicit big data's value, phenomenon that was recognised as new technology, led to new opportunities for firms to extract insight of immense volume and variety of data in real time from multiple structured and unstructured data types.

Table 6 the percent of big data challenges.

Challenges	Number	Percent (%)
Volume of data	20	16.2
Analyse of Big Data	5	4
Volume and variety of data	24	19.5
Velocity and volume of data	2	1.6
Variety, velocity and volume of data	9	7.3
Variety, volume and veracity of data	2	1.6
Variety, velocity, volume and veracity of data	4	3.2
Variety, velocity, volume, value and veracity of data	10	8.1
Variety, velocity, volume, value, visualisation, variability and veracity of data	2	1.6

Table 7. the percent of big data benefits.

Benefits	Number	Per cent (%)
More precise knowledge exploration	71	57.7
Ease of implementation; improvement of processing speed	11	8.9
Improvement of processing speed	9	7.3
Ease of implementation	7	5.6
Improvement of processing speed; more precise knowledge exploration; ease of implementation	4	3.2
Decrease of needed capacity; ease of implementation	3	2.4

This paper addressed a systematic literature review on big data area to identify big data's system benefits and its challenges. Also, this paper, presented a

comprehensive landscape about academic concern in big data, however, this research only examines scientific resources and does not include executive projects and professional white paper. For doing that task eight online databases were searched with different keywords, important of which was big data. Moreover, research strategy was defined, that included three phases: planning review, managing review and reporting review and after searching databases 1085 articles were found for doing investigation on them. In the next step, study selection criteria were defined and just articles remained in review process that included inclusion criteria. Other articles that included exclusion criteria were eliminated from research process; 190 studies remained in search process. Finally, quality assessment criteria were described and papers, which did not have appropriate quality were deleted again, 67 articles were deleted. Eventually, 123 researches that remain in the investigation process were studied and reviewed to extract their information to answer research questions.

Current investigate demonstrated the common way for doing researches on big data was 'experimental'. Also, the answer of question number 2 brought out, the type huge part of articles which studied on big data area were 'Case'. Moreover, 'data management' was the most popular field in the zone of big data. Finally, the last two question revealed that the main challenge of big data was 'volume and variety of data' and the best advantage of it was 'More precise knowledge exploration'.

This research like all researches had restriction. The main restriction of this article was articles in other languages except English. In addition, books, PhD and master theses were eliminated from study. Generally, for future research we suggest: doing research to present comprehensive landscape for big data's system architecture; doing research to review how academic and business world can implement big data's application easier; doing research to find affecting factors and challenges of big data in implementation of big data's systems; doing research about big data's system benefits for human life and business world and doing research to find what kind of survey and review to do on big data's area.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

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

Abbasi, A., Sarker, S., & Chiang, R.H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17*(2), 3.

Agarwal, R., & Dhar, V. (2014). Editorial – Big data, data science, and analytics: The opportunity and challenge for IS research: INFORMS.

Al Jabri, H.A., Al-Badi, A.H., & Ali, O. (2017). Exploring the usage of big data analytical tools in telecommunication industry in Oman. *Information Resources Management Journal (IRMJ), 30*(1), 1–14.

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.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems, 86*, 33–45.

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

Cheng, X.Q., Jin, X.L., Wang, Y.Z., Guo, J., Zhang, T., & Li, G. (2014). Survey on big data system and analytic technology. *Journal of software, 25*(9), 1889–1908.

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

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in European firms. *Journal of Business Research, 70*, 379–390.

Davenport, T.H., & Patil, D. (2012). Data scientist. *Harvard business review, 90*(5), 70–76. Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64–73.

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

Fadiya, S.O., Saydam, S., & Zira, V.V. (2014). Advancing big data for humanitarian needs. *Procedia Engineering, 78*, 88–95.

Fan, S., Lau, R.Y., & Zhao, J.L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research, 2*(1), 28–32.

Guinea, A.S., Nain, G., & Le Traon, Y. (2016). A systematic review on the engineering of software for ubiquitous systems. *Journal of Systems and Software, 118*, 251–276.

Hagel, J. (2015). Bringing analytics to life. *Journal of Accountancy, 219*(2), 24.

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.
Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing, 74*(7), 2561–2573.
Kitchenham, B.A., Mendes, E., & Travassos, G.H. (2007). Cross versus within-company cost estimation studies: A systematic review. *IEEE Transactions on Software Engineering, 33*(5), 316–329.
Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 2053951714528481.
Kowalczyk, D.-W.-I.M., & Buxmann, P. (2014). Big data and information processing in organizational decision processes. *Business & Information Systems Engineering, 6*(5), 267–278.
Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134–1145.
Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387–394.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review, 52*(2), 21.
Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences, 83*(1_suppl), 120–135.
Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. *Strategy & Leadership, 43*(5), 32–39.
McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., & Barton, D. (2012). Big data. The management revolution. *Harvard Business Review, 90*(10), 61–67.
Murthy, D., & Bowman, S.A. (2014). Big data solutions on a small scale: Evaluating accessible high- performance computing for social research. *Big Data & Society, 1*(2), 2053951714559105.
Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. *Journal of Decision Systems, 24*(1), 87–102.
Rekha, J., & Parvathi, R. (2015). Survey on software project risks and big data analytics. *Procedia Computer Science, 50,* 295–300.
Seddon, J.J., & Currie, W.L. (2017). A model for unpacking big data analytics in high-frequency trading.
Journal of Business Research, 70, 300–307.
Shmueli, G., & Koppius, O.R. (2011). Predictive analytics in information systems research. *MIS Quarterly,*
553–572.
Silva, R.A., de Souza, S.R.S., & de Souza, P.S.L. (2017). A systematic review on search based mutation testing. *Information and Software Technology, 81,* 19–35.
Sivarajah, U., Kamal, M.M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70,*

263–286.
Steed, C.A., Ricciuto, D.M., Shipman, G., Smith, B., Thornton, P.E., Wang, D., & Williams, D.N. (2013). Big data visual analytics for exploratory earth system simulation analysis. *Computers & Geosciences, 61*, 71–82.
Vasarhelyi, M.A., Kogan, A., & Tuttle, B.M. (2015). Big data in accounting: An overview. *Accounting Horizons, 29*(2), 381–396.
Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.-F., Dubey, R., & Childe, S.J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365. Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research, 70*, 287–299.
Wang, S., & Yuan, H. (2014). Spatial data mining: A perspective of big data. *International Journal of Data Warehousing and Mining (IJDWM), 10*(4), 50–70.
Wang, C., Schwan, K., Talwar, V., Eisenhauer, G., Hu, L., & Wolf, M. (2011). *A flexible architecture integrating monitoring and analytics for managing large-scale data centers*. Paper presented at the Proceedings of the 8th ACM international conference on Autonomic computing.
Ward, M.J., Marsolo, K.A., & Froehle, C.M. (2014). Applications of business analytics in healthcare. *Business Horizons, 57*(5), 571–582.
Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology, 54*(1), 41–59.
Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering, 26*(1), 97–107.
Xin, J., Wang, Z., Qu, L., & Wang, G. (2015). Elastic extreme learning machine for big data classification.
Neurocomputing, 149, 464–471.

Appendix 1. Included studies on systematic review

(1) 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.
(2) Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. Journal of Parallel and Distributed Computing, 74(7), 2561–2573.
(3) Hofstee, H. P., Chen, G. C., Gebara, F. H., Hall, K., Herring, J., Jamsek, D. … & Wong, P. W. Y. (2013). Understanding system design for Big Data workloads. IBM Journal of Research and Development, 57(3/4), 3–1.
(4) Abawajy, J. (2015). Comprehensive analysis of big data variety landscape. International Journal of Parallel, Emergent and Distributed Systems, 30(1), 5–14.
(5) Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. Knowledge-Based Systems, 86, 33–45.
(6) Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134–1145.

(7) Otero, C. E., & Peter, A. (2015). Research directions for engineering big data analytics software. IEEE Intelligent Systems, 30(1), 13–19.
(8) Lee, J., Lapira, E., Bagheri, B., & Kao, H. A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters, 1(1), 38–41.
(9) Power, D. J. (2014). Using 'Big Data'for analytics and decision support. Journal of Decision Systems, 23(2), 222–228.
(10) Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. ieee transactions on knowledge and data engineering, 26(1), 97–107.
(11) Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data comput- ing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, 79, 3–15.
(12) Xin, J., Wang, Z., Qu, L., & Wang, G. (2015). Elastic extreme learning machine for big data classification. Neurocomputing, 149, 464–471.
(13) Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C. J., & Buytaert, W. (2015). Web technologies for environmental Big Data. Environmental Modelling & Software, 63, 185–198.
(14) Fadiya, S. O., Saydam, S., & Zira, V. V. (2014). Advancing big data for humanitarian needs. Procedia Engineering, 78, 88–95.
(15) Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.
(16) Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314–347.
(17) Zhong, T., Doshi, K. A., Tang, X., Lou, T., Lu, Z., & Li, H. (2013, October). On mixing high-speed updates and in-memory queries: A big-data architecture for real-time analytics. In Big Data, 2013 IEEE International Conference on (pp. 102–109). IEEE.
(18) Zhai, Y., Guo, Y., Chen, Q., Yang, K., & Mbarushimana, E. (2013, November). Design and Optimization of a Big Data Computing Framework Based on CPU/GPU Cluster. In High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), 2013 IEEE 10th International Conference on (pp. 1039–1046). IEEE.
(19) Truong, H. L., & Dustdar, S. (2014, March). Principles of software-defined elastic systems for big data analytics. In Cloud Engineering (IC2E), 2014 IEEE International Conference on (pp. 562–567). IEEE.
(20) Li, H., & Lü, X. (2014, November). Challenges and trends of big data analytics. In P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2014 Ninth International Conference on (pp. 566–567). IEEE.
(21) Huai, Y., Lee, R., Zhang, S., Xia, C. H., & Zhang, X. (2011, October). DOT: a matrix model for analyzing, optimizing and deploying software for big data analytics in distributed systems. In Proceedings of the 2nd ACM Symposium on Cloud Computing (p. 4). ACM.
(22) Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. International Journal of Information Management, 34(3), 387–394.
(23) Song, J., Guo, C., Wang, Z., Zhang, Y., Yu, G., & Pierson, J. M. (2015).

HaoLap: a Hadoop based OLAP system for big data. Journal of Systems and Software, 102, 167–181.

(24) Yim, K. S. (2014, November). Norming to performing: Failure analysis and deployment auto- mation of big data software developed by highly iterative models. In Software Reliability Engineering (ISSRE), 2014 IEEE 25th International Symposium on (pp. 144–155). IEEE.

(25) Jutla, D. N., Bodorik, P., & Ali, S. (2013, June). Engineering Privacy for Big Data Apps with the Unified Modeling Language. In Big Data (BigData Congress), 2013 IEEE International Congress on (pp. 38–45). IEEE.

(26) Girtelschmid, S., Steinbauer, M., Kumar, V., Fensel, A., & Kotsis, G. (2014). On the application of Big Data in future large-scale intelligent Smart City installations. International Journal of Pervasive Computing and Communications, 10(2), 168–182.

(27) Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. Industrial and Commercial Training, 47(4), 174–181.

(28) Anshari, M., Alas, Y., Sabtu, N. I., & Yunus, N. (2016). A Survey Study of Smartphones Behavior in Brunei: A Proposal of Modelling Big Data Strategies. International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), 6(1), 60–72.

(29) Calvard, T. S. (2016). Big data, organizational learning, and sensemaking: Theorizing interpre- tive challenges under conditions of dynamic complexity. Management learning, 47(1), 65–82.

(30) Moon, H., Cho, H. S., Jeong, S. H., & Park, J. (2014, June). Policy design based on risk at big data era: case study of privacy invasion in South Korea. In Big Data (BigData Congress), 2014 IEEE International Congress on (pp. 756–759). IEEE.

(31) Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., … & Keim, D. (2012, October). Visual analytics for the big data era – A comparative review of state-of-the-art commercial systems. In Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on (pp. 173–182). IEEE.

(32) Dai, L., Gao, X., Guo, Y., Xiao, J., & Zhang, Z. (2012). Bioinformatics clouds for big data manip- ulation. Biology direct, 7(1), 43.

(33) Mohammed, E. A., Far, B. H., & Naugler, C. (2014). Applications of the MapReduce program- ming framework to clinical big data analysis: current landscape and future trends. BioData mining, 7(1), 22.

(34) Di Tria, F., Lefons, E., & Tangorra, F. (2014, October). Design process for Big Data warehouses. In Data Science and Advanced Analytics (DSAA), 2014 International Conference on (pp. 512–518). IEEE.

(35) Qin, X., & Zhou, X. (2013, July). Designing a big data processing platform for algorithm trading strategy evaluation. In Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference on (pp. 1005–1009). IEEE.

(36) Rajbhoj, A., Kulkarni, V., & Bellarykar, N. (2014, December). Early experience with model-driven development of mapreduce based big data application. In Software Engineering Conference (APSEC), 2014 21st Asia-

Pacific (Vol. 1, pp. 94–97). IEEE.
(37) Zin, T. T., Tin, P., Toriu, T., & Hama, H. (2013, October). A big data application framework for consumer behavior analysis. In Consumer electronics (GCCE), 2013 IEEE 2nd global confer- ence on (pp. 245–246). IEEE.
(38) Yu, Y., Tang, W., Zou, H., & Liu, L. (2014, October). A CCG virtual system for big data application communication costs analysis. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 54–60). IEEE.
(39) Suto, K., Nishiyama, H., Kato, N., Nakachi, T., Sakano, T., & Takahara, A. (2015, May). A failure-tol- erant and spectrum-efficient wireless data center network design for improving performance of Big Data mining. In Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st (pp. 1–5). IEEE.
(40) Hu, B., Ma, Y., Zhang, L. J., Shi, J., & Zhong, J. (2014, June). A Key-Value Based Application Platform for Enterprise Big Data. In Big Data (BigData Congress), 2014 IEEE International Congress on (pp. 446–453). IEEE.
(41) Shang, W., Jiang, Z. M., Hemmati, H., Adams, B., Hassan, A. E., & Martin, P. (2013, May). Assisting developers of big data analytics applications when deploying on hadoop clouds. In Proceedings of the 2013 International Conference on Software Engineering (pp. 402–411). IEEE Press.
(42) 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.
(43) Muhtaroglu, F. C. P., Demir, S., Obali, M., & Girgin, C. (2013, October). Business model canvas perspective on big data applications. In Big Data, 2013 IEEE International Conference on (pp. 32–37). IEEE.
(44) Neves, M. V., De Rose, C. A., Katrinis, K., & Franke, H. (2014, May). Pythia: Faster big data in motion through predictive software-defined network optimization at runtime. In Parallel and Distributed Processing Symposium, 2014 IEEE 28th International (pp. 82–90). IEEE.
(45) Zhao, Y., Wu, J., & Liu, C. (2014). Dache: A data aware caching for big-data applications using the MapReduce framework. Tsinghua science and technology, 19(1), 39–50.
(46) Friedrich, J., Le, H., Starke, W., Stuechli, J., Sinharoy, B., Fluhr, E. J., … & Hogenmiller, D. (2014, May). The POWER8 TM processor: Designed for big data, analytics, and cloud environments. In IC Design & Technology (ICICDT), 2014 IEEE International Conference on (pp. 1–4). IEEE.
(47) Sowe, S. K., & Zettsu, K. (2013, October). The architecture and design of a community-based cloud platform for curating big data. In Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2013 International Conference on (pp. 171–178). IEEE.
(48) Lim, J. H., Kim, I. K., Bae, S., & Lee, S. H. (2014, January). System proposal and CRS model design applying personal Information protection for BIG DATA analysis. In Big Data and Smart Computing (BIGCOMP), 2014 International Conference on (pp. 231–234). IEEE.
(49) Liu, Z. (2014, July). Research of performance test technology for big data applications. In Information and Automation (ICIA), 2014 IEEE International Conference on (pp. 53–58). IEEE.
(50) Rathore, M. M. U., Paul, A., Ahmad, A., Chen, B. W., Huang, B., & Ji, W.

(2015). Real-time big data analytical architecture for remote sensing application. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(10), 4610–4621.

(51) Dai, D., Chen, Y., Kimpe, D., & Ross, R. (2014, October). Provenance-based object storage prediction scheme for scientific big data applications. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 271–280). IEEE.

(52) Pandey, S., & Tokekar, V. (2014, April). Prominence of mapreduce in big data processing. In Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on (pp. 555–560). IEEE.

(53) Bu, Y., Borkar, V., Xu, G., & Carey, M. J. (2013, June). A bloat-aware design for big data applica- tions. In ACM SIGPLAN Notices (Vol. 48, No. 11, pp. 119–130). ACM.

(54) Ke, H., Li, P., Guo, S., & Guo, M. (2016). On traffic-aware partition and aggregation in mapreduce for big data applications. IEEE Transactions on Parallel and Distributed Systems, 27(3), 818–828.

(55) Chang, V. (2015). Towards a Big Data system disaster recovery in a Private Cloud. Ad Hoc Networks, 35, 65–82.

(56) Mocanu, E. M., & Tapus, N. (2014, September). Non-linear prediction over time-based big data application. In Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on (pp. 405–409). IEEE.

(57) Eridaputra, H., Hendradjaya, B., & Sunindyo, W. D. (2014, November). Modeling the require- ments for big data application using goal oriented approach. In Data and Software Engineering (ICODSE), 2014 International Conference on (pp. 1–6). IEEE.

(58) Lin, M. S., Chiu, C. Y., Lee, Y. J., & Pao, H. K. (2013, October). Malicious URL filtering – A big data application. In big data, 2013 IEEE international conference on (pp. 589–596). IEEE.

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

(60) Zou, H., Yu, Y., Tang, W., & Chen, H. W. M. (2014). FlexAnalytics: a flexible data analytics frame- work for big data applications with I/O performance improvement. Big Data Research, 1, 4–13.

(61) Brunswicker, S., Bertino, E., & Matei, S. (2015). Big Data for Open Digital Innovation–A Research Roadmap. Big Data Research, 2(2), 53–58.

(62) Lee, C. W., Hsieh, K. Y., Hsieh, S. Y., & Hsiao, H. C. (2014). A dynamic data placement strategy for hadoop in heterogeneous environments. Big Data Research, 1, 14–22.

(63) Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In Contemporary Computing (IC3), 2013 Sixth International Conference on (pp. 404–409). IEEE.

(64) Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the Big Data Ecosystem. In Collaboration Technologies and Systems (CTS), 2014 International Conference on (pp. 104–112). IEEE.

(65) Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. Big Data

Research, 2(1), 28–32.

(66) Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. Industrial and Commercial Training, 47(4), 174–181.

(67) Guo, T., Papaioannou, T. G., & Aberer, K. (2014). Efficient indexing and query processing of model-view sensor data in the cloud. Big Data Research, 1, 52–65.

(68) Li, X., Wang, Y., Li, X., Wang, X., & Yu, J. (2014). GDPS: an efficient approach for skyline queries over distributed uncertain data. Big Data Research, 1, 23–36.

(69) Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: research challenges and future directions. The Journal of Supercomputing, 72(4), 1494–1516.

(70) Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. Big Data Research, 2(4), 166–186.

(71) Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How'big data'can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics, 165, 234–246.

(72) Chen, G., Wu, S., & Wang, Y. (2015). The evolvement of big data systems: from the perspective of an information security application.

(73) Olsson, N. O., & Bull-Berg, H. (2015). Use of big data in project evaluations. International Journal of Managing Projects in Business, 8(3), 491–512.

(74) Olsson, N. O., & Bull-Berg, H. (2015). Use of big data in project evaluations. International Journal of Managing Projects in Business, 8(3), 491–512.

(75) Wang, C., Li, X., Zhou, X., Wang, A., & Nedjah, N. (2016). Soft computing in big data intelligent transportation systems. Applied Soft Computing, 38, 1099–1108.

(76) Tannahill, B. K., & Jamshidi, M. (2014). System of Systems and Big Data analytics–Bridging the gap. Computers & Electrical Engineering, 40(1), 2–15.

(77) Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems, 55(1), 412–421.

(78) Rekha, J. H., & Parvathi, R. (2015). Survey on Software Project Risks and Big Data Analytics. Procedia Computer Science, 50, 295–300.

(79) Nirmala, M. B. (2014, February). Wan optimization tools, techniques and research issues for cloud-based big data analytics. In Computing and Communication Technologies (WCCCT), 2014 World Congress on (pp. 280–285). IEEE.

(80) Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big data chal- lenges in building the global earth observation system of systems. Environmental Modelling & Software, 68, 1–26.

(81) Steed, C. A., Ricciuto, D. M., Shipman, G., Smith, B., Thornton, P. E., Wang,

D. … & Williams, D. N. (2013). Big data visual analytics for exploratory earth system simulation analysis. Computers & Geosciences, 61, 71–82.
(82) Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. Big Data Research, 2(3), 87–93.
(83) Srivastava, U., & Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. Procedia Computer Science, 50, 643–652.
(84) Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of big data computing in health sciences. Big Data Research, 2(1), 2–11.
(85) Chen, K., Li, X., & Wang, H. (2015). On the model design of integrated intelligent big data analytics systems. Industrial Management & Data Systems, 115(9), 1666–1682.
(86) Qin, Y., Yalamanchili, H. K., Qin, J., Yan, B., & Wang, J. (2015). The current status and challenges in computational analysis of genomic big data. Big data research, 2(1), 12–18.
(87) Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. Strategy & Leadership, 43(5), 32–39.
(88) H. Davenport, T. (2014). How strategists use“big data”to support internal business decisions, discovery and production. Strategy & Leadership, 42(4), 45–50.
(89) Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. Journal of Decision Systems, 24(1), 87–102.
(90) Wu, C., Chen, Y., & Li, F. (2016). Decision model of knowledge transfer in big data environment. China Communications, 13(7), 100–107.
(91) Kim, W., Jeong, O. R., & Kim, C. (2016). A holistic view of big data. Big Data: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, 73.
(92) Wilson, L., Goh, T. T., & Wang, W. Y. C. (2012). Big Data Management Challenges in a Meteorological Organisation. International Journal of E-Adoption (IJEA), 4(2), 1–14.
(93) Batini, C., Rula, A., Scannapieco, M., & Viscusi, G. (2016). From data quality to big data quality. In Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1934–1956). IGI Global.
(94) Hagos, D. H. (2016). Software-Defined Networking for Scalable Cloud-based Services to Improve System Performance of Hadoop-based Big Data Applications. International Journal of Grid and High Performance Computing (IJGHPC), 8(2), 1–22.
(95) Wang, Y., & Wiebe, V. J. (2016). Big Data Analytics on the characteristic equilibrium of collective opinions in social networks. In Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1403–1420). IGI Global.
(96) Wang, S., & Yuan, H. (2014). Spatial data mining: a perspective of big data. International Journal of Data Warehousing and Mining (IJDWM), 10(4), 50–70.
(97) Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics

success in healthcare. Journal of Business Research, 70, 287–299.
(98) Maciejewski, M. (2016). To do more, better, faster and more cheaply: using big data in public administration. International Review of Administrative Sciences, 0020852316640058.
(99) Kowalczyk, D. W. I. M., & Buxmann, P. (2014). Big Data and information processing in organ- izational decision processes. Business & Information Systems Engineering, 6(5), 267–278.
(100) Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 2,053,951,714,528,481.
(101) Murthy, D., & Bowman, S. A. (2014). Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research. Big Data & Society, 1(2), 2,053,951,714,559,105.
(102) Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 2,053,951,716,631,130.
(103) Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society, 3(1), 2,053,951,716,650,211.
(104) Jagadish, H. V. (2015). Big data and science: myths and reality. Big Data Research, 2(2), 49–52.
(105) Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263–286.
(106) Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high fre- quency trading. Journal of Business Research, 70, 300–307.
(107) Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365.
(108) Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. Journal of Business Research, 70, 308–317.
(109) Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. Journal of Business Research, 70, 366–378.
(110) Li, C. S., Franke, H., Parris, C., Abali, B., Kesavan, M., & Chang, V. (2017). Composable architec- ture for rack scale big data computing. Future Generation Computer Systems, 67, 180–193.
(111) Corbellini, A., Mateos, C., Zunino, A., Godoy, D., & Schiaffino, S. (2017). Persisting big-data: The NoSQL landscape. Information Systems, 63, 1–23.
(112) Pahins, C. A., Stephens, S. A., Scheidegger, C., & Comba, J. L. (2017). Hashedcubes: Simple, low memory, real-time visual exploration of big data. IEEE Transactions on Visualization and Computer Graphics, 23(1), 671–680.

(113) Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. Journal of Business Research, 70, 379–390.
(114) Lytras, M. D., Raghavan, V., & Damiani, E. (2017). Big Data and Data Analytics Research: From Metaphors to Value Space for Collective Wisdom in Human Decision Making and Smart Machines. International Journal on Semantic Web and Information Systems (IJSWIS), 13(1), 1–10.
(115) Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. Tourism Management, 59, 554–563.
(116) Önder, I. (2017). Classifying multi-destination trips in Austria with big data. Tourism Management Perspectives, 21, 54–58.
(117) Li, J., & Yu, H. (2017). Large time behavior of solutions to a bipolar hydrodynamic model with big data and vacuum. Nonlinear Analysis: Real World Applications, 34, 446–458.
(118) Chen, Y., Crespi, N., Ortiz, A. M., & Shu, L. (2017). Reality mining: A prediction algorithm for disease dynamics based on mobile big data. Information Sciences, 379, 82–93.
(119) Cui, B., Shi, P., Qi, W., & Li, M. (2017). Uploading multiply deferrable big data to the cloud platform using cost-effective online algorithms. Future Generation Computer Systems, 67, 276–285.
(120) Rathore, M. M., Paul, A., Ahmad, A., & Jeon, G. (2017). IoT-Based Big Data: From Smart City towards Next Generation Super City Planning. International Journal on Semantic Web and Information Systems (IJSWIS), 13(1), 28–47.
(121) Halaweh, M., & El Massry, A. (2017). A Synergetic Model for Implementing Big Data in Organizations: An Empirical Study. Information Resources Management Journal (IRMJ), 30(1), 48–64.
(122) Al Jabri, H. A., Al-Badi, A. H., & Ali, O. (2017). Exploring the Usage of Big Data Analytical Tools in Telecommunication Industry in Oman. Information Resources Management Journal (IRMJ), 30(1), 1–14.
(123) Fang, Y., Chen, X., Song, Z., Wang, T., & Cao, Y. (2017). Modelling Propagation of Public Opinions on Microblogging Big Data Using Sentiment Analysis and Compartmental Models. International Journal on Semantic Web and Information Systems (IJSWIS), 13(1), 11–27.